



DOCUMENTO DE TRABAJO

METODOLOGÍA Y RESULTADOS DE LA VALIDACIÓN DEL MAPA DE COBERTURA DE BOSQUE NATIVO DE URUGUAY PARA EL AÑO 2016 CON IMÁGENES SENTINEL

Proyecto REDD+ Uruguay

Coordinación Técnica: Diego Martino

Autora principal: María Eugenia Riaño

Autores colaboradores: Lucia Bernardi, Alejandra Boccoardo, Carlos Miguel, Juan Olivera,
Cecilia Penengo, Gonzalo Rama

El proyecto REDD+ es ejecutado en el marco de un acuerdo interministerial entre el Ministerio de Ganadería, Agricultura y Pesca y el Ministerio de Vivienda, Ordenamiento Territorial y Medio Ambiente. Como parte de este acuerdo participan directamente en su implementación la Dirección General Forestal, la Oficina de Programación y Políticas Agropecuarias, la Dirección Nacional de Medio Ambiente y la División de Cambio Climático. El proyecto REDD+ cuenta con apoyo financiero del Fondo Cooperativo para el Carbono de los Bosques (FCPF).

Este documento debe citarse como:

Proyecto REDD+ Uruguay (2019). Metodología y resultados de la validación del mapa de cobertura de bosque nativo de Uruguay para el año 2016 con imágenes Sentinel 2. Riaño, M.E., Bernardi, L., Boccardo, A., Miguel, C., Olivera, J., Penengo, C. y Rama, G. Ministerio de Ganadería, Agricultura y Pesca y Ministerio de Vivienda, Ordenamiento Territorial y Medio Ambiente. Montevideo.

En este documento se emplea el masculino gramatical, como término inclusivo para aludir a colectivos mixtos, o en contextos genéricos o inespecíficos.

CONTENIDO:

1.	Introducción	4
2.	Descripción de la metodología aplicada	4
2.1.	Diseño muestral para la validación	4
2.1.1.	Estratificación	4
2.1.2.	Tamaño de muestra y asignación entre estratos	5
2.1.3.	Probabilidades de inclusión	6
2.1.4.	Calibración	7
2.1.5.	Estimadores	7
2.2.	Interpretación visual de los segmentos incluidos en la muestra de validación	9
3.	Resultados de la validación del mapa de cobertura de bosque nativo de Uruguay para el año 2016 con Sentinel 2	11
3.1.	Matriz de confusión	11
3.2.	Exactitud Global, del Usuario y del Productor	11
	Bibliografía	14

1. Introducción

En el presente documento se detalla la metodología aplicada para la validación del mapa de cobertura de bosque nativo de Uruguay para el año 2016 a partir de imágenes Sentinel 2.

Dicha metodología incluye: a) los criterios utilizados para la selección de la muestra para la validación (diseño muestral); b) los pasos seguidos para la interpretación de la muestra de validación; c) la estimación de las superficies de Bosque Nativo y No Bosque; d) la estimación de la exactitud global de la clasificación de los segmentos, la exactitud del usuario y la exactitud del productor para cada una de las clases del mapa, con sus respectivos intervalos de confianza y la precisión asociada.

2. Descripción de la metodología aplicada

2.1. Diseño muestral para la validación

De acuerdo a las recomendaciones de buenas prácticas de Oloffson et al. [2], el diseño muestral debe ser probabilístico, de forma que proporcione una base rigurosa para las estimaciones, basadas en el diseño. Se sugiere estratificar por clase y aumentar el tamaño de muestra en las clases “raras” (de menor tamaño) para aumentar la precisión de la estimación.

El tamaño de muestra final depende de los objetivos planteados en términos de precisión para la evaluación de los errores de comisión y omisión. Stehman [4] plantea una amplia variedad de diseños muestrales que cumplen con las recomendaciones de buenas prácticas (Diseños Estratificados, Simples, Sistemáticos, por Conglomerados y combinaciones de éstos). Oloffson et al. [2] concluyen que si bien el diseño estratificado simple por clase es el más sencillo, puede cumplir con los objetivos propuestos para la precisión y para la estimación de áreas y, en general, es recomendable por su fácil implementación práctica.

El diseño implementado para la validación del mapa de cobertura de bosque nativo de Uruguay para el año 2016 con Sentinel 2 fue aleatorio, estratificado simple por clase y se calcularon los tamaños de muestra por estrato de forma que las estimaciones de los errores de comisión para las diferentes clases tuviera una alta precisión.

2.1.1. Estratificación

Los estratos se construyeron a partir de las dos clases definidas en el mapa de cobertura y de las tres zonas de trabajo en las que dividió el país para su elaboración.

Las clases fueron el “Bosque Nativo” y el “No Bosque”, esta última conformada por el resto de las clases del mapa agrupadas en una única categoría (ver Tabla 1).

Las zonas de trabajo fueron: la Zona Norte, al norte del Río Negro limitado al este por la ruta 6; la Zona Sur, al sur del Río Negro limitado al este por las rutas 11, 7, 14 y 6 y la Zona Este, al este de las ruta 6 y 7 (ver Figura 1).

La unidad de análisis fueron los segmentos que conformaron las coberturas de Bosque Nativo y de No Bosque.

Bosque Nativo	Bosque Nativo
No Bosque	Plantaciones Forestales
	No Bosque
	Agua
	Bajos Inundables

Tabla 1: Clasificación de las categorías de segmentos

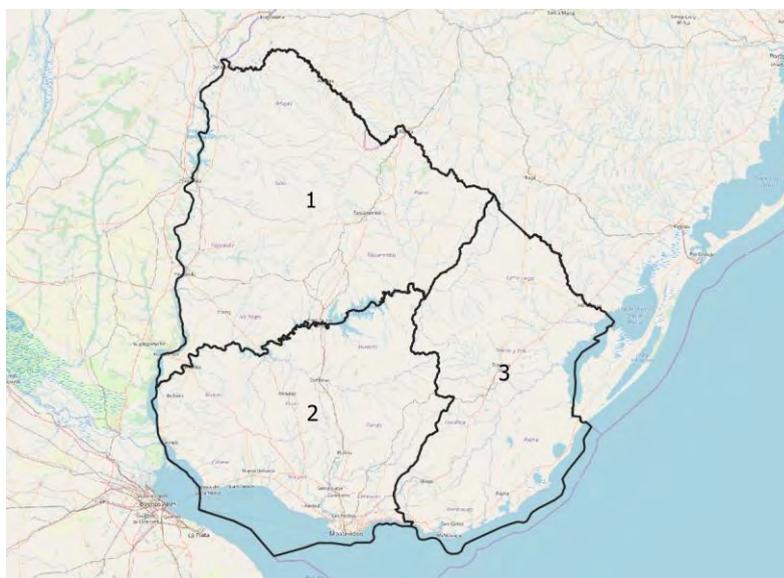


Figura 1: Zonas de trabajo

Se consideraron así seis estratos para la selección de la muestra de validación:

- Bosque Nativo – Zona Norte
- Bosque Nativo – Zona Sur
- Bosque Nativo – Zona Este
- No Bosque – Zona Norte
- No Bosque – Zona Sur
- No Bosque – Zona Este

2.1.2. Tamaño de muestra y asignación entre estratos

El tamaño de muestra se calculó de forma de obtener una precisión de ± 0.02 en los errores de comisión de las capas de cobertura de Bosque Nativo y de No Bosque, ambos con una confianza del 95%.

Para el cálculo del tamaño de muestra fue necesario realizar un supuesto sobre la magnitud del error de clasificación. Se asumió que este último era como máximo de un 20% para cada una de las clases consideradas (Bosque Nativo y No Bosque).

La fórmula utilizada para el cálculo del tamaño de muestra fue la siguiente¹:

$$n = \frac{Nz_{1-\alpha}^2 p(1-p)}{N\varepsilon^2 + z_{1-\alpha}^2 p(1-p)}$$

Donde N es la cantidad de segmentos totales, $z_{1-\alpha}$ es el valor de la Normal Estándar que corresponde a una confianza del $1-\alpha\%$ (en este caso es 1.96 para una confianza del 95%), ε es la precisión, $p(1-p)$ es la varianza de una variable aleatoria Bernoulli, que corresponde al caso en donde la variable de interés es dicotómica (en este caso $p = 0.20$).

Los resultados se presentan en la siguiente tabla:

Clase	N	ε	$z_{1-\alpha}$	$p(1-p)$	n
Bosque Nativo	914.952	0,02	1,96	0,1875	1.793
No Bosque	14.479.284	0,016	1,96	0,16	2.330

Tabla 2: Tamaños de muestra por clase

La asignación de la muestra entre las zonas de trabajo se realizó en forma proporcional a la cantidad de segmentos de cada área.

Para los estratos considerados, los tamaños de muestra fueron los siguientes:

Estrato	N_h	n_h
Bosque – Zona Norte	410.957	817
Bosque – Zona Sur	183.894	357
Bosque – Zona Este	320.101	619
Total	914.952	1.793
No Bosque – Zona Norte	5.820.338	1.091
No Bosque – Zona Sur	4.453.617	628
No Bosque – Zona Este	4.205.329	611
Total	14.479.284	2.330

Tabla 3: Tamaños de muestra por estrato

2.1.3. Probabilidades de inclusión

Las probabilidades de inclusión se calcularon, en el caso del Bosque Nativo, como:

$$\pi_{k_B} = \frac{n_B}{N_B}$$

¹ En el documento de buenas prácticas de Olofsson et al. [2] la fórmula para el cálculo del tamaño de muestra asume un muestreo aleatorio simple con reposición. En este caso la fórmula corresponde a un diseño Simple sin reposición. Dados los tamaños de los estratos los resultados serían similares con ambas fórmulas.

Donde n_B es el tamaño de muestra para la clase Bosque Nativo y N_B es la cantidad total de segmentos de la capa de Bosque Nativo.

En el caso de No Bosque como:

$$\pi_{k_{NB}} = \frac{n_{NB}}{N_{NB}}$$

Donde n_{NB} es el tamaño de muestra para la clase No Bosque y N_{NB} es la cantidad total de segmentos de la capa de No Bosque.

Las probabilidades de inclusión de los segmentos de la capa de Bosque Nativo eran diferentes a las de la capa de No Bosque. La metodología que proponen Olofsson et al. [2] para el cálculo de las proporciones de área bien clasificada se basa en un diseño con probabilidades constantes. En este caso, al ser diferentes entre las clases, se utilizaron estimadores más generales (tipo Hurwitz – Thompson y de razón), como los propuestos por Strahler et al. [5].

2.1.4. Calibración

Los pesos muestrales se obtuvieron como el inverso de la probabilidad de inclusión. Una vez calculados, se ajustaron de forma que se estime sin error el área total de Bosque Nativo y de No Bosque por zona de trabajo.

Sea d_{k_B} el peso muestral para el Bosque Nativo y $d_{k_{NB}}$ el peso correspondiente al No Bosque, para cada zona de trabajo h el peso muestral calibrado se calculó como:

$$w_{k_B} = d_{k_B} \frac{t_{h_B}}{\hat{t}_{h_B}}$$

Donde \hat{t}_{h_B} es el área estimada y t_{h_B} el área total calculada de la clase Bosque Nativo a partir del mapa en la zona de trabajo h , con $h = 1, 2, 3$.

Para el caso de No Bosque, los pesos calibrados se calcularon de forma análoga, utilizando las áreas estimadas y las áreas calculadas de la clase No Bosque a partir del mapa, por zona de trabajo.

2.1.5. Estimadores

De acuerdo a Oloffson et al. [2] y a Stehman [4], los errores deben calcularse en términos de proporciones de superficie bien clasificada. Los datos a estimar implican el uso de dos tipos de estimadores: estimadores calibrados [3], utilizados para el cálculo de las estimaciones de áreas totales y estimadores de razón [6], utilizados para el cálculo del error global y de los errores de comisión y omisión.

Mapa	Referencia ²		
	Bosque	No Bosque	Total
Bosque Nativo	$\hat{t}_{B,B}$	$\hat{t}_{B,NB}$	t_B
No Bosque	$\hat{t}_{NB,B}$	$\hat{t}_{NB,NB}$	t_{NB}
Total	\hat{t}_B	\hat{t}_{NB}	t

Tabla 4: Estimadores de la superficie según la clasificación de la matriz de confusión

² Los datos de Referencia son los que surgen de la revisión de los segmentos con imágenes de alta resolución.

El estimador calibrado utilizado para la estimación del total en las celdas de la matriz de confusión (Tabla 4) se presenta a continuación:

$$\hat{t}_i = \sum_{s_i} w_k x_k$$

Donde s_i es el subconjunto de la muestra correspondiente a cada celda de la matriz de confusión ($i = (B, B), (NB, NB), (NB, B), (B, NB)$), w_k es el peso calibrado del segmento k y x_k su correspondiente área en hectáreas. Los totales de las columnas se obtienen como la suma de las estimaciones de las celdas. El total de las filas coincide con el área calculada a partir de las capas de Bosque Nativo y de No Bosque del mapa de cobertura.

Los estimadores de Razón son el cociente entre dos estimadores. El cálculo de la exactitud global, O se calculó como el cociente entre las áreas bien clasificadas (diagonal de la matriz de confusión) y la superficie total:

$$O = \frac{\hat{t}_{B,B} + \hat{t}_{NB,NB}}{t}$$

La exactitud del usuario para la clase Bosque Nativo U se calculó como el cociente entre el área bien clasificada como Bosque Nativo y el área de Bosque Nativo obtenida a partir de la capa de Bosque Nativo del mapa de cobertura (filas de la matriz de confusión):

$$U = \frac{\hat{t}_{B,B}}{\hat{t}_{B.}}$$

La exactitud del usuario para la clase No Bosque se obtuvo de forma análoga considerando el área bien clasificada de No Bosque ($\hat{t}_{NB,B}$) y el área total de No Bosque obtenida a partir de la capa de No Bosque del mapa de cobertura.

La exactitud del productor, P se calculó como el cociente entre el área bien clasificada como Bosque Nativo y el área estimada total de Bosque Nativo a partir de los datos de referencia (columnas de la matriz de confusión):

$$P = \frac{\hat{t}_{B,B}}{\hat{t}_{.B}}$$

La exactitud del productor para la clase No Bosque se obtuvo de forma análoga considerando el área bien clasificada de No Bosque ($\hat{t}_{NB,B}$) y el área total estimada de No Bosque a partir de los datos de referencia.

La estimación del error global se obtuvo como $1 - \hat{Q}$ la del error de comisión (o del usuario) como $1 - \hat{U}$ la del error de omisión (o del productor) es $1 - \hat{P}$

Para el cálculo de la varianza de los estimadores se utilizó la librería *survey* [1] del paquete estadístico R. En el caso del estimador de Razón con diseño simple o estratificado simple la librería utiliza las fórmulas presentadas por Särndal et al. [3] en la Sección 5.6, para estimadores de razón.

2.2. Interpretación visual de los segmentos incluidos en la muestra de validación

Una vez definidos los segmentos a evaluar para cada clase del mapa de cobertura de bosque nativo para el año 2016 con imágenes Sentinel 2 (ver Tabla 3, página 6 de este documento), se llevaron a cabo los siguientes pasos para generar la información necesaria para construir la matriz de confusión y estimar la exactitud temática del mapa:

a) Definición de los datos a utilizar como “verdad de campo” para la validación

Para la interpretación visual de los segmentos incluidos en la muestra de validación se utilizaron, como “verdad de campo”, imágenes de alta resolución de Google Earth y de Esri. Como criterio general, siempre se trató de elegir aquella imagen cuya fecha fuera lo más cercana posible a la fecha de las imágenes de Sentinel 2 utilizadas para la elaboración del mapa (ver documento “Metodología aplicada para la elaboración del mapa de cobertura de bosque nativo para el año 2016 con el sensor Sentinel 2”, Proyecto REDD+ Uruguay, 2019).

b) Determinación del equipo responsable por la interpretación visual de los segmentos de la muestra de validación

Para la interpretación visual de las clases correspondientes a cada uno de los segmentos de la muestra de validación, se decidió conformar un grupo de expertos en interpretación visual de imágenes integrado por miembros externos al equipo que elaboró el mapa y miembros del equipo que estuvo a cargo de la elaboración del mapa de cobertura de bosque nativo 2016 con Sentinel 2. Esta decisión se basó en las recomendaciones brindadas al equipo del Proyecto REDD+ por parte de los expertos en sistemas de Medición, Reporte y Verificación del Banco Mundial y expertos en Validación de Productos Cartográficos de la Universidad de Boston, EEUU.

Es así que la muestra total de segmentos a evaluar se distribuyó aleatoriamente entre los integrantes de este equipo de validación, correspondiendo evaluar un 50% de la muestra a los técnicos externos y el restante 50% a los técnicos internos del equipo técnico del Proyecto REDD+. Los técnicos externos fueron representantes de la empresa ICA³ de Uruguay (Ingenieros Consultores Asociados).

c) Elaboración de la ficha a completar por cada intérprete durante el proceso de interpretación visual

A los efectos de registrar los resultados de la interpretación visual de cada uno de los segmentos de la muestra de validación, se confeccionó una ficha específica en la que cada uno de los operadores debía completar los siguientes campos:

Campo 1_ (Nombre: Cat_bosque): % de cobertura de bosque nativo en el segmento evaluado.

Clase 1_ 0%

Clase 2_ >0%-25%

Clase 3_ >25-50%

Clase 4_ >50-75%

Clase 5_ >75-<100%.

Clase 6_ 100%

³ <http://www.ica.com.uy/>

Campo 2_(Nombre: **Uso_altern**): % de cobertura de otro uso diferente al bosque nativo en el segmento evaluado (en el caso de que haya más de un uso diferente al bosque nativo en el segmento evaluado seleccionar únicamente el mayoritario).

Clase 1_Pastizal

Clase 2_Forestación

Clase 3_No bosque en general, otros

Clase 4_Agua

Clase 5_Bajos y/o Pajonales

Clase 6_Cultivo

Campo 3_(Nombre: **Fecha_img**): fecha de la imagen de alta resolución utilizada para la interpretación visual.

Campo 4_(Nombre: **Observ**): cualquier observación que entienda necesario agregar para facilitar la evaluación.

A continuación se presenta como ejemplo una de las fichas utilizadas a estos efectos:

id	Cat_bosq	Uso_altern	Fecha_img	Observ
Z1B1005264	1	0	22/03/2014	
Z1B1006079	1	0	11/04/2017	
Z1B1007372	1	0	22/03/2014	
Z1B1008058	1	0	19/01/2015	
Z1B1010205	1	0	22/03/2014	
Z1B1013773	1	0	22/03/2014	
Z1B1014788	1	0	11/04/2017	
Z1B101926	1	0	11/04/2017	
Z1B1021564	1	0	19/01/2015	
Z1B1021586	1	0	11/04/2017	
Z1B1024352	2	5	11/04/2017	
Z1B1028893	2	1	22/03/2014	
Z1B1029652	1	0	13/02/2015	

Tabla 5: Ejemplo de una porción de una ficha completada por parte de los responsables de la interpretación visual de los segmentos de la muestra de validación

d) *Uso de la información relevada para evaluar la exactitud temática del mapa*

Una vez finalizado el trabajo del equipo a cargo de la validación, se procedió a cotejar el resultado de la interpretación de cada uno de los segmentos de la muestra de validación por parte del equipo a partir de las fichas de registro con el resultado de la clasificación, es decir con la clase correspondiente a cada segmento en el mapa de cobertura. De esta manera se obtuvo la información necesaria para construir la matriz de confusión del mapa de cobertura y para estimar la exactitud global, del usuario y del productor, aplicando las fórmulas presentadas en los puntos 2.1.4. y 2.1.5. de este documento. Los resultados obtenidos se presentan en el punto siguiente.

3. Resultados de la validación del mapa de cobertura de bosque nativo de Uruguay para el año 2016 con Sentinel 2

3.1. Matriz de confusión

La matriz de confusión estimada para el mapa de cobertura de bosque nativo de Uruguay para el año 2016 con Sentinel 2, expresada en términos de superficie, es la siguiente:

Mapa	Referencia		Total
	Bosque Nativo	No Bosque	
Bosque Nativo	797.232	38.119	835.351
No Bosque	234.716	16.666.364	16.901.081
Total	1.031.949	16.704.483	17.736.432

Tabla 6: Totales estimados de superficie (en hectáreas) de acuerdo a la clasificación de la matriz de confusión

Los intervalos de confianza para cada una de las celdas de la matriz de confusión se presentan a continuación:

	Inf. 95% ⁴	Sup. 95% ⁵
B- B	755.317	839.148
B- NB	29.503	46.734
NB -B	160.120	309.312
NB - NB	15.336.342	17.996.386

Tabla 7: Intervalos de Confianza para las superficies estimadas (en hectáreas) de acuerdo a la clasificación de la matriz de confusión

La superficie de Bosque Nativo bien clasificada se encuentra entre 755.317 y 839.148 hectáreas, con una confianza del 95%.

Para la superficie de No Bosque bien clasificada el intervalo es de entre 15.336.342 y 17.996.386 hectáreas, con una confianza del 95%.

Las superficies mal clasificadas se encuentran entre 29.503 y 46.734 hectáreas en el caso de la superficie de No Bosque clasificada como Bosque Nativo, y entre 160.120 y 309.312 hectáreas en el caso de la superficie de Bosque Nativo clasificada como No Bosque, ambos intervalos con una confianza del 95%.

3.2. Exactitud Global, del Usuario y del Productor

La exactitud global del mapa es de un 98.46% (con un intervalo de confianza al 95% de entre el 98.03% y el 98.9%). Este indicador es útil como medida de la performance global del mapa, pero no es atribuible directamente a una clase en particular (Bosque Nativo o No Bosque).

Las exactitudes del usuario y del productor (definidas a continuación), son medidas que sí pueden atribuirse directamente a una clase del mapa y, a partir de ellas, pueden obtenerse estimaciones de los errores de comisión (sobrecobertura) y omisión (subcobertura) de cada clase en particular.

⁴ Inf. 95% se refiere al extremo inferior del intervalo de confianza obtenido al 95%.

⁵ Sup. 95% se refiere al extremo superior del intervalo de confianza obtenido al 95%.

La exactitud del usuario es la proporción de superficie bien clasificada de cada clase (Bosque Nativo; No Bosque), considerando como superficie total de cada clase la superficie de dicha clase resultante del proceso de clasificación (filas de la matriz de confusión). A partir de la exactitud del usuario se obtienen los errores de comisión.

La exactitud del productor es la proporción de superficie bien clasificada considerando como superficie total de Bosque Nativo y de No Bosque las estimadas a partir de los datos de referencia⁶ (columnas de la matriz de confusión). A partir de la exactitud del productor se obtienen los errores de omisión.

	Exactitud del Usuario	Exactitud del Productor
Bosque Nativo	95.44%	77.25%
No Bosque	98.61%	99.77%

Tabla 8: Estimación de la exactitud del usuario y del productor

Los intervalos de confianza para la exactitud del usuario son:

	Inf. 95%	Sup. 95%	Precisión
Bosque Nativo	94.41%	96.46%	1.03%
No Bosque	98.16%	99.06%	0.45%

Tabla 9: Intervalos de confianza y precisión para la estimación de la exactitud del usuario

La proporción⁷ de superficie bien clasificada de la capa de Bosque Nativo es de un 95.44%, y se encuentra entre el 94.41% y el 96.46% con una confianza del 95%. El error en este caso es de ± 1.03 puntos porcentuales, encontrándose dentro de los objetivos planteados al calcular el tamaño de muestra.

Para la capa de No Bosque la proporción de superficie bien clasificada es de un 98.61%, y se encuentra entre un 98.16% y un 99.06% con una confianza del 95%. El error de estimación es de ± 0.45 puntos porcentuales, encontrándose también dentro de los objetivos planteados inicialmente.

Los intervalos de confianza para la exactitud del productor son:

	Inf. 95%	Sup. 95%	Precisión
Bosque Nativo	71.54%	82.97%	5.71%
No Bosque	99.72%	99.83%	0.06%

Tabla 10: Intervalos de confianza y precisión para la estimación de la exactitud del productor

La proporción⁸ de superficie bien clasificada de la capa de Bosque Nativo es de un 77.25%, y se encuentra entre el 71.54% y el 82.97% con una confianza del 95%. El error en este caso es de ± 5.71 puntos porcentuales.

⁶ Los datos de referencia se obtuvieron mediante la interpretación visual con imágenes de alta resolución de los segmentos incluidos en la muestra de validación.

⁷ Se toma como total a la superficie calculada a partir de las capas de coberturas obtenidas para Bosque Nativo y No Bosque (835.351 y 16.901.081 hectáreas respectivamente).

⁸ Se toma como total a la superficie estimada a partir de los datos de referencia para Bosque Nativo y No Bosque (1.031.949 y 16.704.483 hectáreas respectivamente).

Para la capa de No Bosque la proporción de superficie bien clasificada es de un 99.77%, y se encuentra entre un 99.72% y un 99.83% con una confianza del 95%. El error de estimación es de ± 0.06 puntos porcentuales.

Bibliografía

- [1] Lumley, T. (2019). "Survey: analysis of complex survey samples". R Package version 3.35-1.
- [2] Olofsson P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., Wulder, M. A. (2014). *Good practices for estimating area and assessing accuracy of land change*. Remote Sensing of Environment 148 42–57.
- [3] Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Springer series in statistics. Model assisted survey sampling*. New York, NY, US: Springer-Verlag Publishing.
- [4] Stehman, S. V. (2009). *Sampling design for accuracy assessment of land cover*. International Journal of Remote Sensing, 30:20,5243 — 5272.
- [5] Strahler, A. H., Boschetti, L., Foody, G. M., Friedl, M.A., Hansen, M. C., Herold, M., et al. (2006). *Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps*. EUR 22156 EN — DG. Luxembourg: Office for Official Publications of the European Communities (48 pp.).