

Propuesta metodológica para la cartografía periódica de la cobertura del suelo en Latinoamérica y el Caribe: estado de situación y avances

P. Blanco¹, G. López Saldaña², R. Colditz², L. Hardtke¹, N. Mari³,
A. Fischer³, C. Caride³, P. Aceñolaza⁴, H. del Valle¹, S. Opazo⁵, W. Sione⁴,
M. Lillo-Saavedra⁶, P. Zamboni⁴, I. Cruz López², J. Anaya⁷,
F. Morelli⁸ y S. de Jesús⁸

¹ Centro Nacional Patagónico, CONICET, Puerto Madryn, Chubut, Argentina

² Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), México DF, México

³ Instituto de Clima y Agua, INTA, Castelar, Buenos Aires, Argentina

⁴ Centro Regional de Geomática (CEREGEO-FCyT-UADER), Paraná, Entre Ríos, Argentina

⁵ Escuela de Ciencia y Tecnologías Agropecuarias, Universidad de Magallanes, Punta Arenas, Chile

⁶ Departamento de Mecanización y Energía, Facultad de Ing. Agrícola,
Universidad de Concepción, Chile

⁷ Facultad de Ingeniería Ambiental, Universidad de Medellín, Medellín, Colombia

⁸ Centro de Previsão de Tempo e estudos Climáticos, INPE, Sao José dos Campos, Brasil

Resumen

Este trabajo fue desarrollado en el marco de la Red Latinoamericana de Seguimiento y Estudio de los Recursos Naturales (SERENA) en respuesta a la gran necesidad existente de contar con información de cobertura del suelo, con una resolución espacial media, en la región de Latinoamérica y el Caribe (LAC). En el presente artículo se detalla la metodología y se presentan los resultados preliminares obtenidos para la cartografía de la cubierta superficial de LAC para el año 2008.

PALABRAS CLAVE: cobertura del suelo, Latinoamérica y Caribe, árboles de decisión, MODIS, SERENA.

Abstract

A new methodology for systematic mapping of land covers in Latin America and Caribbean

This work was developed in the framework of the Latin American Network of Monitoring and Study of Natural Resources (SERENA) in response to the need for accurate and reliable information on landcover, with a medium spatial resolution in the region of Latin America and the Caribbean (LAC). This article describes the methodology and presents the results obtained for the year 2008.

Key words: landcover, Latin America and Caribe, decision trees, MODIS, SERENA.

Introducción

Durante los últimos 40 años, Latinoamérica ha experimentado un proceso de cambios ambientales sin precedentes en la historia. La deforestación, los incendios, la expansión de la

agricultura, el cambio climático y la globalización de la economía mundial son algunas de las principales causas. Para el estudio del cambio global, se requiere de información actualizada y accesible para el seguimiento del estado de los ecosistemas, el cual se viene de-

* Autor para la correspondencia: blanco@cenpat.edu.ar
Recibido: 03-04-12; Aceptado: 30-11-12.

sarrollando en distintos países a través de diferentes herramientas y metodologías. Este trabajo se enmarca en el contexto de la Red Latinoamericana de Seguimiento y Estudio de los Recursos Naturales (SERENA), que reúne a 18 institutos de investigación, con 52 investigadores participantes, de 11 países. El principal objetivo del proyecto SERENA es el seguimiento, el estudio y la difusión de información asociada a la quema de biomasa y de los cambios producidos en el uso y cobertura del suelo sobre Latinoamérica y el Caribe (LAC). En este contexto, una adecuada caracterización de la superficie, obtenida en forma periódica y con una resolución espacial media es esencial para implementar un sistema de monitoreo del cambio de la superficie terrestre que pueda ser actualizado anualmente y sea una fuente de información sostenible y crítica para los usuarios de esta región.

Los principales productos de cobertura del suelo en la región de LAC incluyen cartografías elaboradas a través de tres iniciativas internacionales a nivel global, como son el producto Global Land Cover (GCL2000) producido a partir de imágenes SPOT (Bartholome y Belward 2005), el GlobeCover elaborado con imágenes MERIS (Arino et al. 2008), y más recientemente, el producto Land Cover obtenido a partir de imágenes MODIS colección 5 (Friedl et al. 2010). Estos productos presentan diferencias básicas a nivel instrumental (sensores utilizados), técnico (algoritmo de clasificación) y conceptual (sistema de clasificación). Asimismo, estudios recientes (Pérez-Hoyos y García-Haro 2009, Cabral et al. 2010) han mostrado importantes diferencias e inconsistencias entre los mismos.

El objetivo de este trabajo es aprovechar el alto potencial espectral y temporal que ofrece el sensor MODIS para derivar información precisa sobre la cobertura de suelo en LAC, utilizando algoritmos de clasificación supervisada, valiéndose del conocimiento de los diferentes tipos de cubiertas, a nivel local, de los expertos que son miembros de la red SERENA. Asimismo, se pretende elaborar un producto que sea compatible, tanto en la leyenda utilizada como a nivel técnico-instrumental, con la cartografía de cobertura y uso del suelo generado por el proyecto North

American Land Cover Change Monitoring System (NALCMS) para Canadá, EEUU y México, de manera de unificar ambas cartografías a nivel continental.

Metodología

Las fases metodológicas realizadas para la generación del producto LandCover SERENA para la región de LAC, tomando como base el año 2008, fueron las siguientes:

1. Definición de la leyenda.
2. Adquisición y pre-procesamiento de la información satelital base derivada del sensor MODIS y datos auxiliares (datos climáticos, topográficos).
3. Recopilación de datos de entrenamiento a partir del muestreo aleatorio de cartografía de detalle de la región.
4. Clasificación con árboles de decisión.
5. Aplicación de máscaras.
6. Validación de resultados.

Definición de la leyenda SERENA

Se utilizó un sistema de clasificación jerárquico que fuera compatible con el Land Cover Classification System propuesto por el Fondo de las Naciones Unidas para la Agricultura y la Alimentación (FAO). Esta elección permite compatibilizar la cartografía de coberturas de Latinoamérica y Caribe con la generada por el proyecto NALCMS. La Tabla 1 muestra la leyenda empleada.

Adquisición y pre-procesamiento de los datos MODIS y la información auxiliar

El Sensor MODIS (Moderate Resolutions Imaging Spectroradiometer), a bordo de las plataformas satelitales Terra (desde 1999) y Aqua (desde 2002), posee una anchura de barrido de 2.300 km., proporcionando una visión completa de la tierra cada 1 ó 2 días. Para este trabajo se generaron compuestos mensuales, para el año 2008, con 500 m de resolución espacial de las primeras 7 bandas del sensor a

Tabla 1. Leyenda SERENA para la clasificación de coberturas y uso del suelo de LAC

Clase	
1	Bosque de latifoliadas perennifolio tropical
2	Bosque de latifoliadas caducifolio tropical
3	Bosque de latifoliadas perennifolio subtropical
4	Bosque de latifoliadas caducifolio subtropical
5	Bosque de latifoliadas perennifolio templado
6	Bosque de latifoliadas caducifolio templado
7	Bosque de coníferas subtropical
8	Bosque de coníferas subtropical
9	Bosque mixto
10	Arbustal tropical
11	Arbustal subtropical
12	Arbustal templado
13	Pastizal tropical
14	Pastizal subtropical
15	Pastizal templado
16	Cuerpo de agua
17	Áreas urbanizadas
18	Hielo y nieve permanente
19	Suelo desnudo
20	Cultivo
21	Humedal
22	Salinas
23	Datos nulos

partir de imágenes diarias de reflectividad (producto original MOD09GA). Asimismo, el estimado de calidad del píxel fue utilizado para reemplazar aquellos píxeles marcados con baja calidad empleando un algoritmo de interpolación lineal basado en los píxeles vecinos en tiempo y espacio.

Para cada mes del año 2008 se han utilizado 8 variables [7 bandas espectrales más el Índice Diferencial de Vegetación Normalizada (NDVI)]; siendo evaluadas con 5 estadísticos: media, desvío estándar, mínimo, máximo, y rango entre ambos, obteniendo un total de 40 métricas derivadas de los datos MODIS. Esto fue calculado para año redondo, 6, 4 y 3 meses. Las variables auxiliares topográficas fueron derivadas del SRTM (elevación, pendiente, aspecto) y las variables climáticas de la base de datos WORDCLIM (se utilizaron las siguientes variables: bio 1, 2, 4-7, y 12-15).

Obtención de datos de entrenamiento

Se convocó a expertos miembros de la red SERENA para generar puntos de entrenamiento según el protocolo que se detalla a continuación:

1. Recopilación de cartografía de referencia detallada y actualizada de los distintos países/regiones que conforman LAC.
2. Traducción de las distintas leyendas de la cartografía de referencia a la leyenda SERENA.
3. Generación de puntos de muestreo aleatorios a partir de la cartografía de referencia, distanciados 5 km y construcción de zonas cuadradas buffer de 2×2 km a partir de los mismos.
4. Aplicación de un primer filtro: cruzamiento de la capa buffer con los polígonos de la capa de referencia, con la restricción que los

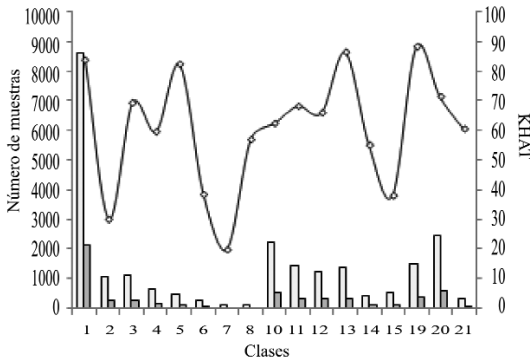


Figura 1. Distribución de puntos de entrenamiento (barras blancas) y validación (barras grises) sobre eje izquierdo, y coeficiente Kappa (KHAT) (línea) sobre eje derecho, para cada clase considerada en la clasificación. Ver Tabla 1 para los códigos de las clases.

elementos buffer que se superpongan con más de un polígono sean eliminados. Este filtro permite excluir puntos que podrían confundir más de una clase.

5. Aplicación de un segundo filtro: los elementos buffer son exportados como archivos KML y visualizados en Google Earth®. A partir del conocimiento experto son eliminados los objetos que no se correspondan con la clase de interés o que no cumplan con el requisito de pertenecer a clases homogéneas. A partir de la aplicación de estos dos filtros se genera la capa de puntos de entrenamiento (Figura 1).

Clasificación con árboles de decisión

Los árboles de decisión ofrecen varias ventajas sobre los algoritmos de clasificación tradicionalmente utilizados para el mapeo de la cobertura del suelo. Además de su naturaleza no-paramétrica (lo cual permite incorporar datos discretos a la clasificación, independientemente de la distribución y la correlación que exista entre ellos), posibilitan una fácil inclusión de distintos tipos de datos.

En este trabajo fue utilizado el algoritmo C5.0 para la generación de árboles de decisión, implementado en el software See5. Este algoritmo trabaja a partir de los puntos de entrenamiento determinando las características apropiadas de la variable de respuesta al fraccionar recurrentemente los datos explicativos

en grupos cada vez más homogéneos con el resultado de producir un árbol jerárquico integrado por reglas. De manera de mejorar el desarrollo del algoritmo de aprendizaje se utilizó la técnica de *boosting* con diez iteraciones, que construye clasificadores mediante la asignación de pesos a los ejemplos de forma adaptativa. En cada iteración se construye un clasificador que intenta compensar los errores cometidos previamente por otros clasificadores. Por último, el conjunto de reglas y los clasificadores así obtenidos son utilizados para clasificar la región del LAC. Para esto se utilizó un programa desarrollado en Interactive Data Language (IDL).

Aplicación de mascarar

Las clases Cuerpos de Agua, Hielo y nieve permanente, y Areas Urbanizadas no son clasificadas, sino que son aplicadas en forma de mascarar a posteriori de la clasificación.

En particular, para la obtención de la máscara de cuerpos de agua, se utilizó el producto Global Raster Water Mask (GRWM-MOD44W). Para la generación de máscaras de nieve y hielo permanente, se calibró el índice NDSI6, que corresponde a la diferencia normalizada entre las bandas 4 y 6 de MODIS. La máscara de áreas urbanizadas se obtuvo a partir de imágenes DMSP-OLS (Nighttime Lights Time Series).

Validación de resultados

El conjunto de datos fue dividido en entrenamiento (80%) y validación (20%). Se elaboraron matrices de confusión, las cuales permiten confrontar la información de los sitios de verificación y la imagen clasificada. Se calcularon los valores de la fiabilidad del usuario (proporción de sitios de verificación de la categoría considerada clasificados como tal) y del productor (proporción de sitios clasificados en la categoría considerada en el mapa que están correctamente clasificados). Estos dos índices de fiabilidad están relacionados con los errores de omisión y de comisión respectivamente. Asimismo, fueron cal-



Figura 2. Cartografía preliminar de cobertura y uso del suelo para LAC con su respectiva confianza. Ver Tabla 1 para los códigos de la leyenda.

culados los coeficientes Kappa (KHAT) global y por clase derivados a partir de la matriz de confusión.

Resultados preliminares

La validación de los resultados de la clasificación preliminar mostró una fiabilidad satisfactoria (Kappa global=71.7%), con un rango amplio de valores de precisión por clase comprendidos entre el 20 y 88.3%, siendo la clase Bosque de Coníferas Subtropical la que resultó con mayores errores de confusión con el resto de las clases. En el futuro, se prevé mejorar estos resultados mediante el balanceo de las muestras de entrenamiento en función del área ocupada por cada clase. En la figura 2 se muestra el resultado preliminar de la clasificación discreta. Se determinaron 20 tipos de coberturas, las cuales se discriminan adecuadamente teniendo en cuenta los tipos de ecosistemas y los climas. Los bosques tropicales y subtropicales poseen una importante distribución espacial en comparación a los del tipo templado. Las áreas de suelo desnudo o con baja cobertura se encuentran asociadas, regionalmente, a las zonas desérticas del Caribe y del centro oeste de Latinoamérica y localmente, a playas y actividades antrópicas de denudación. Los humedales presentan una amplia distribución espacial, posiblemente asociada a la presencia de un evento ENSO durante el año 2008. Puede observarse también que existe una

buena discriminación para los bosques subtropicales estacionales y no estacionales.

Conclusiones

En este trabajo se presentó un método para la generación periódica de mapas de cobertura y uso del suelo para la región de LAC a partir de imágenes MODIS. Para ello se propuso utilizar conocimiento experto para entrenar algoritmos de árboles de clasificación. El análisis de los resultados preliminares permite concluir que la utilización de la dinámica anual de la respuesta espectral permitió discriminar adecuadamente la heterogeneidad de la distribución de las coberturas, ajustándose las mismas al concepto de tipos funcionales de ecosistemas.

La técnica de *boosting* utilizada para entrenar los árboles de clasificación, a partir de los datos de entrenamiento considerando las dimensiones espectrales y temporales, permite generar reglas de decisión que pueden ser generalizables a cualquier otro de año de la base de datos MODIS (2000-2012) para la misma zona de estudio.

Los próximos productos que se esperan generar en el marco del proyecto SERENA incluirán mapas actualizados de cobertura del suelo a intervalos regulares y productos para la detección del cambio. Estos productos servirán, entre otras aplicaciones, como indicadores diagnósticos del cambio en la cobertura del suelo en LAC y contribuirán a una mejor comprensión de la dinámica de la cobertura vegetal en la región.

Agradecimientos

Los autores agradecen la financiación recibida a través del proyecto SERENA, así como el apoyo de su coordinador Carlos Di Bella y la invaluable asistencia de Julieta Straschnoy para la realización de este trabajo.

Referencias

ARINO, O., BICHERON, P., ACHARD, F., LATHAM, J., WITT, R., & WEBER, J. 2008. GLOB-

- COVER-The most detailed portrait of Earth. *ESA Bulletin-European Space Agency*, 24:31.
- BARTHOLOME, E., & BELWARD, A. 2005. GLC2000: A new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, 26, 1959-1977.
- FRIEDL, M. A., SULLA-MENASHE, D., TAN, B., SCHNEIDER, A., RAMANKUTTY, N., SIBLEY, N., & HUANG, X. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* 114:168-182.
- CABRAL, A., VASCONCELOS, M., & OOM, D. 2010. Comparing information derived from global land cover datasets with Landsat imagery for the Huambo Province and Guinea-Bissau. En: Wagner W, Székely B (eds.): ISPRS TC VII Symposium-100 Years ISPRS, Vienna, Austria, 5-7 Julio, IAPRS, Vol. XXXVIII, Part 7B. pp. 107-112.
- PÉREZ-HOYOS, A., & GARCÍA-HARO, F. 2009. Análisis comparativo de las principales cartografías de coberturas del suelo en la Península Ibérica. En: Montesinos Aranda S y Fernández Fornos L (eds.): .XIII Congreso de la Asociación Española de Teledetección. Calatayud, 23-26 Septiembre, pp. 209-212.