

Análisis de regresión para el cálculo de biomasa en el departamento de Nariño (Colombia) utilizando imágenes satelitales Landsat

GIEE
Universidad de Nariño
San Juan de Pasto, Colombia
Email:

GIIWW
Universidad de Nariño
San Juan de Pasto, Colombia
Email:

Abstract—En este artículo se describen los componentes básicos de la investigación aplicada que tiene como objetivo construir un primer mapa del potencial de biomasa en el departamento de Nariño (Colombia) a partir de imágenes satelitales Landsat de libre acceso. Se analizan las diferentes bandas de las imágenes satelitales disponibles y su relación con bases de datos previas de biomasa aplicando diferentes técnicas de regresión y obteniendo un modelo para la generación de mapas actualizados en el departamento de Nariño.

Keywords—biomass, regression models

I. INTRODUCCIÓN

Según estudios del Ministerio de Minas y Energía de Colombia, en el departamento de Nariño hay 15 municipios con cobertura eléctrica inferior al 80% [1]. Como nueva estrategia para enfrentar esta problemática se ha planteado la medición y estimación de potenciales energéticos en las zonas más viables de la región. Uno de los componentes a analizar es el potencial de biomasa para la generación eléctrica. Sin embargo, uno de los problemas que se plantea para la ubicación de lugares propicios es la ausencia de bases de datos actualizadas en el área de estudio que permitan su respectivo análisis.

Diversas investigaciones han demostrado la utilidad del uso de imágenes satelitales para la generación de modelos que permitan calcular la cantidad de biomasa presente en un determinado lugar. Desde hace más de 30 años, se cuenta con libre acceso al repositorio de imágenes satelitales Landsat [2] que con el debido tratamiento pueden ser usadas para calcular valores nominales de biomasa. Sin embargo, dichos modelos requieren la ejecución de trabajo de campo en la zona para inferir fórmulas iniciales a partir de la medición tradicional de una muestra. Dadas las dificultades para realizar dicho trabajo de campo se utilizó imágenes provistas por investigaciones anteriores [3], [4] donde se proporciona niveles de biomasa a nivel pan-tropical. El acceso a imágenes para cada uno de los países tomados en cuenta para el estudio están disponibles en [5].

Esta investigación está orientada a cumplir con los requerimientos necesarios para la generación de un modelo de predicción de biomasa y su extrapolación al resto del área de estudio.

II. TRABAJOS RELACIONADOS

El estudio de índices de biomasa ha sido ampliamente registrado por diferentes estudios. Muchos de ellos demuestran la utilidad del uso de imágenes satelitales a diferentes resoluciones. Por lo general, estos estudios parten de un trabajo de campo donde se calcula el valor nominal de biomasa a diferentes muestras haciendo uso de técnicas tradicionales en laboratorio. Posteriormente, se utilizan estos resultados y las diferentes bandas proveídas por las imágenes de satélite para inferir un modelo utilizando alguna técnica de regresión que después es extrapolada al resto del área de estudio.

Por ejemplo, [6] usa esta metodología para detectar cambios en los niveles de biomasa en diferentes zonas costeras de los Estados Unidos utilizando imágenes LIDAR y regresión lineal. De forma análoga, [3] usa imágenes MODIS y árboles de decisión (en adición a las técnicas tradicionales de regresión) para estimar el índice AGB (Above-Ground Biomass) en una extensa área del África tropical. Similar a este trabajo, [7] utilizan imágenes de radar para predecir AGB en cuatro reservas y parques nacionales africanos clasificando diferentes tipos de corteza terrestre. [8] hacen también uso de imágenes MODIS en conjunto con imágenes ASTER para estimar biomasa con el fin de levantar un inventario de captura de carbono. Un aporte importante de esta publicación es que comparten la metodología utilizada durante el proyecto como lo muestra la figura 1. [9] introducen el uso de nuevas técnicas de regresión (reduced major axis regression, gradient nearest neighbor imputation y random forest regression trees) para la generación de modelos de biomasa esta vez analizando imágenes Landsat. En [10] se introduce bioSTRUCT, un método para generar correlaciones entre los valores continuos medidos por las bandas de las imágenes satelitales y el AGB medido previamente usando técnicas de laboratorio. El artículo ilustra la metodología con un caso de estudio en Alberta (Canadá) e imágenes Landsat ETM+ de libre acceso. Como resultado se obtienen fórmulas de regresión a partir de un número limitado de muestras que pueden extrapolarse al resto del área de estudio.

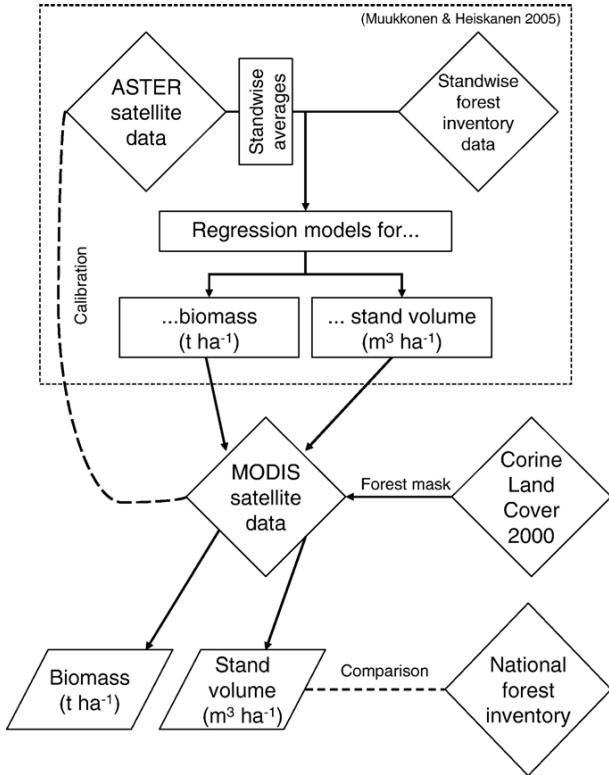


Figura 1. Metodología propuesta por [8]

III. METODOLOGÍA

A. Obtención de datos

El proceso de obtención de datos se realizó tomando imágenes satelitales que provee el satélite Landsat 7. En este proceso se descargaron 1362 imágenes satelitales desde el año 1999 hasta el año 2015, que cubren el departamento de Nariño. Para cubrir todo el departamento fue necesario descargar las imágenes satelitales con los siguientes paths y rows: (009,059), (009,060), (010,058), (010,059), (011,059)

En la obtención de datos también se utilizó el mapa de biomasa construido por [3] para el año 2000 a 2003.

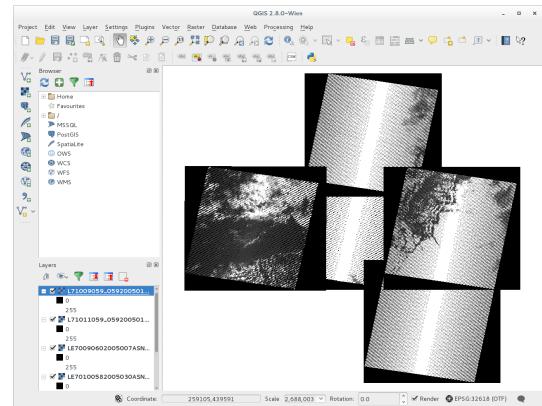
B. Preprocesamiento

En esta etapa de preprocesamiento se reproyectó las imágenes obtenidas, debido a que las cinco imágenes que cubren el departamento de Nariño, están en distintos sistemas de coordenadas (EPSG:32618 y EPSG:32617) y se lo reproyectó al sistema EPSG:3857. Así como también se recortó las imágenes con el fin de únicamente tener el área que cubre el departamento de Nariño, como lo muestra la figura 2

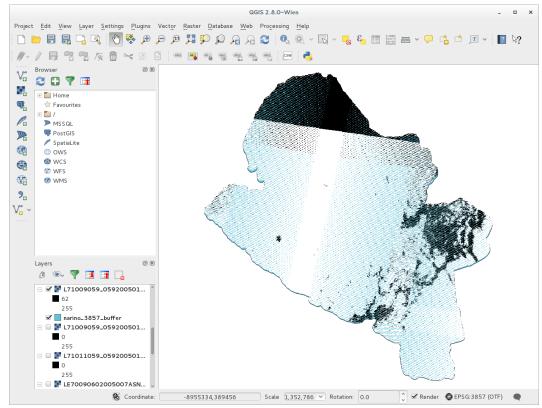
De igual manera este proceso se lo realizó para el mapa de biomasa, como se muestra en la figura 3

C. Procesamiento y limpieza de datos

Se diseñó una base de datos para capturar los datos, como lo muestra la figura 4, la cual tiene 4 tablas.



(a) Imágenes Satelitales de Nariño



(b) Imágenes recortadas de Nariño

Figura 2. Preprocesamiento

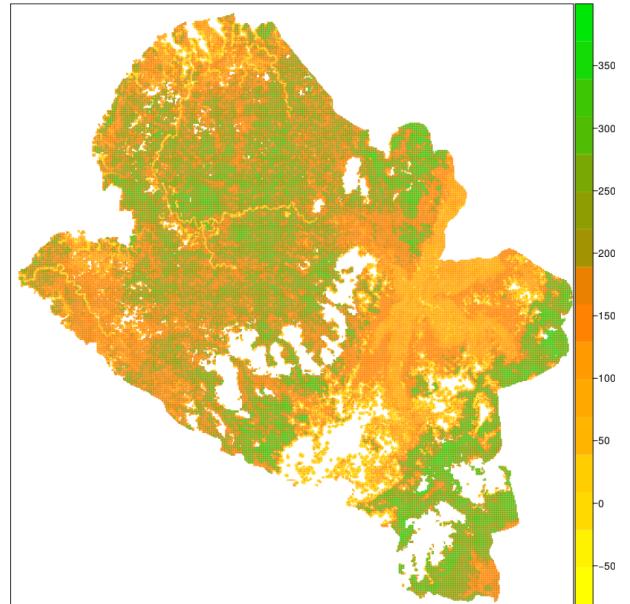


Figura 3. Mapa de biomasa en Nariño de 2000-2003 [3]

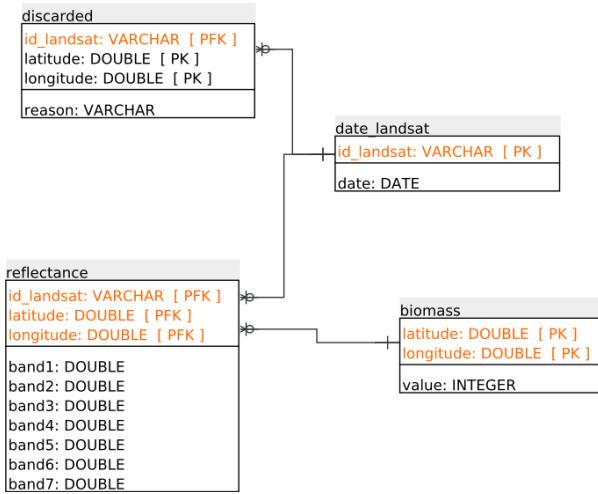


Figura 4. Modelo entidad-relacion Landsat

Tabla I. DATOS OBTENIDOS EN EL PROCESO DE PROCESAMIENTO Y LIMPIEZA DE DATOS

Nombre	Valor	Detalle
Imágenes landsat procesadas	1321	Imágenes de Nariño de 2000 a 2014
Datos Totales	51.076.512	Registros Totales desde año 2000 a 2014
Datos biomasa	81.993	Registros de biomasa para año 2000 a 2003 de [3]
Nube caliente	3.731.768	Registros de 2000 a 2014
Nube Fria	27.827.009	Registros de 2000 a 2014
No vegetacion	3.459.210	Registros de 2000 a 2014
Ambiguo	11.987.340	Registros de 2000 a 2014
Datos Validos Reflectance	4.071.185	Registros de 2000 a 2014

Tabla date_landsat: en la cual se almacenan las fechas de las imágenes satelitales.

Tabla reflectance: en la cual se almacenan los datos capturados y convertidos en reflectance, de las bandas landsat (1 - 5,7) y la temperatura en grados kelvin de la banda 6.

Tabla discarded: en la cual se almacenan datos que fueron descartados, por varias razones, son nubes calientes, nubes frias, datos ambiguos o no son vegetación.

Tabla biomass: en la cual se almacenan los datos de biomassa del mapa de [3].

Para procesar las imágenes y llenar la base de datos se realizó un Script, el cual captura el Digital Number de las imágenes satelitales y lo transforma en valor en reflectance. En este procesamiento de imágenes, se adiciono al Script unos filtros para para detección de nubes calientes, nubes, frias, datos ambiguos como lo muestra el algoritmo propuesto por [11], además se aplico un filtro adicional, el NVDI(normalized difference vegetation index) para trabajar únicamente con datos de vegetación.

La tabla I muestra la relación de los datos obtenidos en este proceso.

D. Análisis de regresión

El análisis de regresión se realizó tomando los valores de las bandas landsat obtenidas año 2000 y 2003 y el valor de biomasa obtenido en [3], para poder obtener un mejor modelo se agrupó y se saco un promedio con valores de las bandas landsat en cada punto, se fue iterando con valores que superaban al menos N número de muestras,

Tabla II. MÉTRICAS DE MODELOS ANALIZADOS CON 35 MUESTRAS Y 1009 DATOS

	SAE	MAE	RAE	RMSE	COR	R2
crtree	10406.58225	30.88007	65.04650	40.02893	0.69401	0.48165
rpart	10197.95826	30.26100	63.74249	39.37592	0.70520	0.49730
kknn	9147.51425	27.14396	57.17667	36.86581	0.74955	0.56182
mlp	9179.79310	27.23974	57.37843	34.70711	0.78122	0.61031
mlpe	8746.27740	25.95335	54.66874	34.57953	0.78309	0.61323
ksvm	8462.61487	25.11162	52.89570	34.67742	0.79830	0.63729
randomForest	8807.76477	26.13580	55.05306	34.70615	0.78239	0.61214
mr	10410.13919	30.89062	65.06873	38.61068	0.72000	0.51840
mars	8842.91866	26.24011	55.27279	33.96852	0.79161	0.62665
cubist	9012.54150	26.74345	56.33302	35.70576	0.77611	0.60235
pqr	10337.63121	30.67546	64.61552	38.59290	0.72023	0.51873
plsR	10337.63121	30.67546	64.61552	38.59290	0.72023	0.51873
cppls	10337.63121	30.67546	64.61552	38.59290	0.72023	0.51873

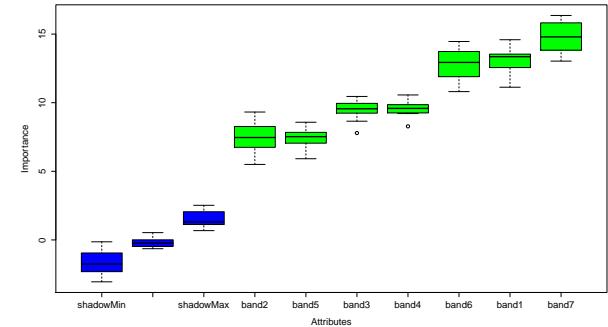


Figura 5. Relevancia de bandas landsat en el análisis de regresión

siendo N desde 1 hasta 45 muestras, el mejor modelo obtenido fué cuando el número de muestras en cada punto superaba al menos las 35 muestras, este conjunto de datos obtenido tenía 1009 registros. El comportamiento en las demás iteraciones muestra que con menos muestras hay más registros y eso hace que no se encuentre un buen modelo, pero cuando hay mas muestras los registros son menores y esto también hace que el resultado del modelo tampoco sea bueno.

En la tabla II se muestra las métricas de los modelos analizadas con 35 muestras y 1009 datos, el cual es el mejor modelo, esta tabla se la realizó usando la biblioteca de código abierto rminer presentada por [12] para la herramienta R.

En este proceso tambien se usó el paquete R Boruta [13], el cual es un nuevo algoritmo de selección de características para encontrar todas las variables relevantes. El algoritmo está diseñado como un recubrimiento alrededor del algoritmo de clasificación random forest. Esto para saber si todas las bandas de landsat utilizadas eran relevantes para encontrar biomasa, en la figura 5 se puede observar la relevancia de las bandas landsat para encontrar biomasa en el mejor modelo.

E. Construcción de mapas

Para la construcción de mapas de biomasa se utilizó el método Kriging que provee una solución al problema de la estimación basada en un modelo continuo de variación espacial estocástica, el objetivo de Kriging es el de estimar el valor de una variable aleatoria, Z, en uno o más puntos no muestreados o sobre grandes bloques.

El método Kriging recibe como entrada datos de la muestra, y una malla dependiendo de la resolución que se

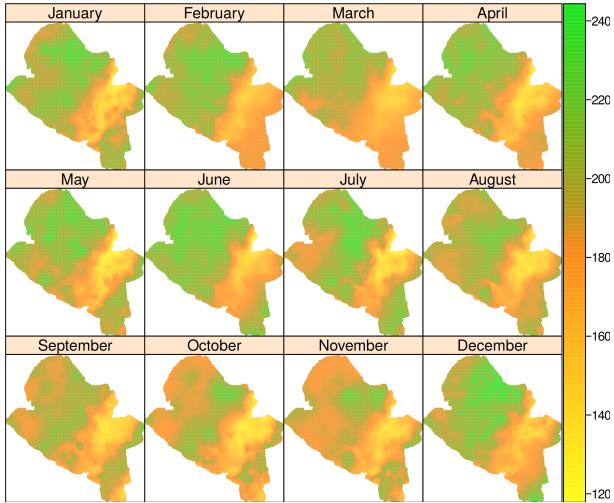


Figura 6. Mapas biomasa por meses

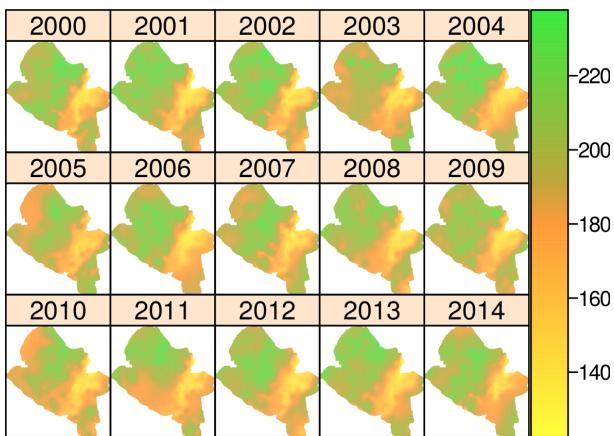


Figura 7. Mapas biomasa por años

quiera obtener, por ello los datos de muestra se obtuvieron aplicando el modelo obtenido en el análisis de regresión a datos agrupados en cada punto por mes, año y uno general entre el año 2000 a 2014; y la malla se construyó con puntos regulares espaciados cada 450 metros.

En la figura 6, figura 7, figura 8 se muestra los mapas obtenidos por meses, años y general entre el año 2000 a 2014 respectivamente.

IV. CONCLUSIONES Y TRABAJOS FUTUROS

AGRADECIMIENTOS

Esta investigación se hizo posible gracias a los recursos otorgados por el Sistemas General de Regalías en el marco de proyecto “Análisis de Oportunidades Energéticas con Fuentes Alternativas en el Departamento de Nariño” ejecutado por el programa de Ingeniería Electrónica de la Universidad de Nariño.

REFERENCIAS

- [1] M. de Minas y Energia, “Plan Indicativo de Expansión de la Cobertura del Servicio de Energía Eléctrica 2006 - 2010.” Unidad de Planeación Minero Energética - Ministerio de Minas, Colombia, Tech. Rep., 2008. [Online]. Available: http://www.siel.gov.co/portals/0/PIEC_REV.pdf
- [2] U. G. Survey, “Landsat missions,” <http://landsat.usgs.gov/>.
- [3] A. Baccini, N. Laporte, S. J. Goetz, M. Sun, and H. Dong, “A first map of tropical africa’s above-ground biomass derived from satellite imagery,” *Environmental Research Letters*, vol. 3, no. 4, p. 045011, oct 2008. [Online]. Available: <http://iopscience.iop.org/1748-9326/3/4/045011>
- [4] A. Baccini, S. J. Goetz, W. S. Walker, N. T. Laporte, M. Sun, D. Sulla-Menashe, J. Hackler, P. S. A. Beck, R. Dubayah, M. A. Friedl, S. Samanta, and R. A. Houghton, “Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps,” *Nature Climate Change*, vol. 2, no. 3, p. 182–185, Jan. 2012.
- [5] W. H. R. Center, “Pantropical national level carbon stock dataset,” http://www.whrc.org/mapping/pantropical/carbon_dataset.html.
- [6] V. Klemas, “Remote sensing of coastal wetland biomass: An overview,” *Journal of Coastal Research*, vol. 290, pp. 1016–1028, sep 2013. [Online]. Available: <http://jcronline.org/doi/abs/10.2112/JCOASTRES-D-12-00237.1>
- [7] E. T. A. Mitchard, S. S. Saatchi, I. H. Woodhouse, G. Nangendo, N. S. Ribeiro, M. Williams, C. M. Ryan, S. L. Lewis, T. R. Feldpausch, and P. Meir, “Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different african landscapes,” *Geophysical Research Letters*, vol. 36, no. 23, 2009. [Online]. Available: <http://onlinelibrary.wiley.com/doi/10.1029/2009GL040692/abstract>
- [8] P. Muukkonen and J. Heiskanen, “Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories,” *Remote Sensing of Environment*, vol. 107, no. 4, pp. 617–624, apr 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S003442570600407X>
- [9] S. L. Powell, W. B. Cohen, S. P. Healey, R. E. Kennedy, G. G. Moisen, K. B. Pierce, and J. L. Ohmann, “Quantification of live aboveground forest biomass dynamics with landsat time-series and field inventory data: A comparison of empirical modeling approaches,” *Remote Sensing of Environment*, vol. 114, no. 5, pp. 1053–1068, may 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0034425709003745>
- [10] R. Hall, R. Skakun, E. Arsenault, and B. Case, “Modeling

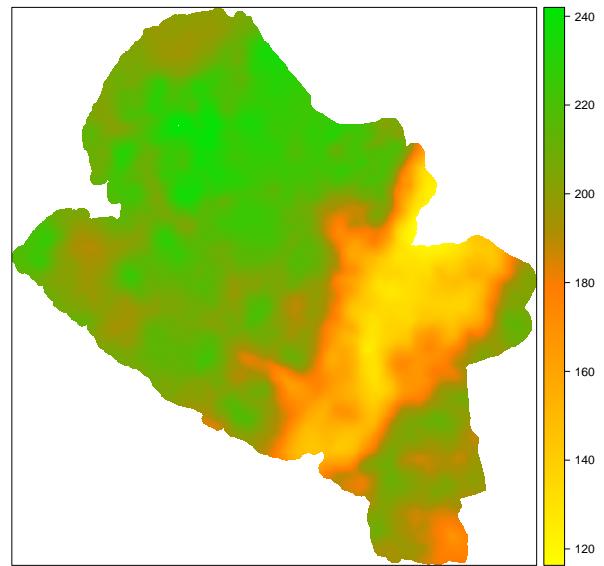


Figura 8. Mapas biomasa general años 2000-2014

- forest stand structure attributes using landsat ETM+ data: Application to mapping of aboveground biomass and stand volume,” *Forest Ecology and Management*, vol. 225, no. 1-3, pp. 378–390, apr 2006. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0378112706000235>
- [11] R. R. Irish, “Landsat 7 automatic cloud cover assessment,” in *AeroSense 2000*. International Society for Optics and Photonics, 2000, pp. 348–355.
 - [12] P. Cortez, “Data mining with neural networks and support vector machines using the r/rminer tool,” in *Advances in data mining. Applications and theoretical aspects*. Springer, 2010, pp. 572–583.
 - [13] M. B. Kursa, W. R. Rudnicki *et al.*, “Feature selection with the boruta package,” 2010.