Regression analysis to calculate biomass of the department of Nariño (Colombia) using Landsat satellite imagery

GIEE
University of Nariño
San Juan de Pasto, Colombia
Email:

GIIWW
University of Nariño
San Juan de Pasto, Colombia
Email:

***Abstract*—This paper describes the basic components of the applied research which aims to build the first biomass potential map of the department of Nariño (Colombia) using Landsat open source satellite imagery. The different band are analyzed from available satellite imagery and their relation with previous biomass data bases applying different regression techniques; and generating a model to creation of updated maps at the department of Nariño.**

Keywords*—*biomass, regression models

I. Introduction

According with studies of Ministry of Mines and Energy of Colombia, in the department of Nariño 15 municipalities have electric coverage under 80 % [1]. As new strategy to address this situation was defined the mensuration and estimation of energy potential at more viable region zones. One component to analyze is biomass potential to electricity generation. However, one problem stated to locate propitious places is the lack of updated databases in this study area which allow their analysis.

Several research have shown how advantageous is use satellite imagery to generate models which allow to calculate how much biomass exists in an specific place. For more than 30 years, the access to Landsat satellite imagery repository is granted [2]; this images, with appropriated treatment, can be used to calculate nominal values of biomass. however, this models require fieldwork in selected zones to infer initial formulas from a sample measurement. Due the difficulties to execute this fieldwork were used images provided by previous researches [3], [4] where are provided biomass levels at pantropical level. The access to images to each country used for this research are available at [5].

This research is oriented to fulfill the requirements to generate a model for biomass prediction and its extrapolation to the full study area.

II. Related work

The biomass indexes has been widely explored by several researches. Most of them show how useful is use satellite imagery with different resolutions. In general, this researches have as start point a fieldwork where is calculated the biomass nominal value to different samples using traditional laboratory techniques. Subsequently, this results are used and the different bands given by the satellite imagery to infer a model using some regression technique to finally extrapolate to the study area.

As an example, [6] uses this methodology to detect changes in biomass levels at different coast zones at United States using LIDAR imagery and lineal regression. Similarly, [3] use MODIS imagery and decision trees (in addition to traditional regression techniques) to estimate AGB (Above-Ground Biomass) index at an wide tropical African area. Comparable to this study, [7] use radar imagery to predict AGB at four reserve and African national parks classifying different types of earth crust. [8] also use of MODIS imagery in conjunction with ASTER imagery to estimate biomass pursuing an carbon sequestration inventory. An important contribution of this paper is the shared methodology used through the project as is showed the figure 1. [9] introduce the use of new regression techniques (reduced major axis regression, gradient nearest neighbor imputation y ramdom forest regression trees) to generate biomass models this time using Landsat imagery. In [10] is introduced bioSTRUCT, a method to generate correlations between continued values measured by the satellite imagery bands and the AGB measured previously using laboratory techniques. The paper shows the methology as a study case at Alberta (Canada) and open source Landsat imagery ETM+. As a result are obtained regression formulas from the limited number of samples than can be extrapolated to the full study area.

III. Metodology

A. Data acquiring

The process of acquiring data was performed using satellite imagery provided by Landsat 7 satellite. In this process were downloaded 1362 satellite images, covering the department of Nariño, from years 1999 to 2015. To cover the whole department was necessary download satellite imagery with the next paths and rows: (009,059), (009,060), (010,058), (010,059), (011,059)

Also, in the data acquiring was used the biomass map build by [3] for years 2000 to 2003.

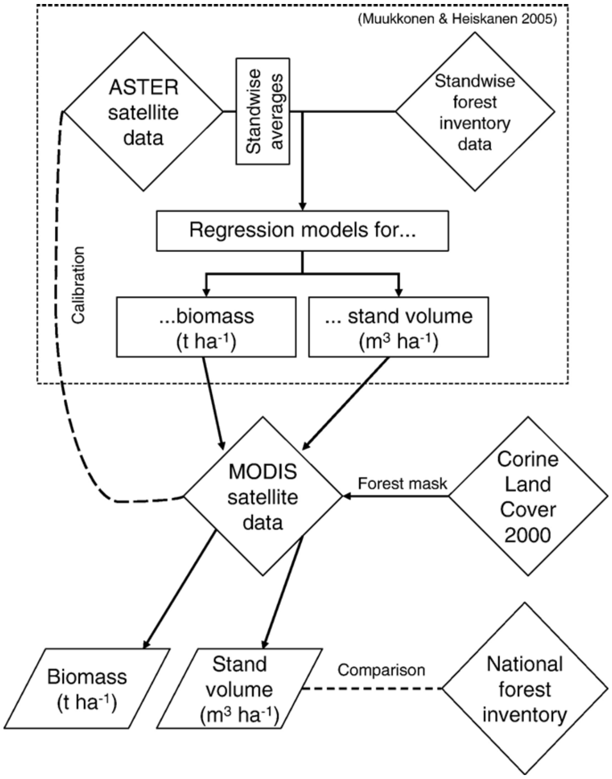


Figura 1. Methodology proposed by [8]

B. Preprocessing

In this step, the imagery acquired was reprojected, because five images are in different coordinate systems (EPSG:32618 and EPSG:32617) and were reprojected to the EPSG:3857 system. As well as the images were cut to cover the department of Nariño, as is shown in the figure 2

Similarly, this process was applied to biomass map as is shown in the figure 3

C. Processing and data cleaning

This research designs a data bse to capture data, as is shown in the figure 4, la cual tiene 4 tablas.

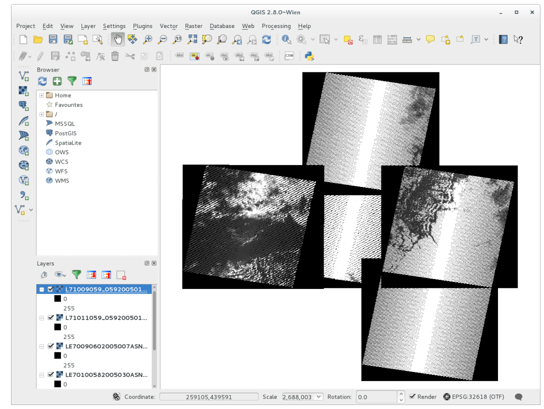
Table date_landsat: This table stores dates of each satellite image.

Table reflectance: This table stores data captured and converted in reflentance of the Landsat band (1 - 5,7) and temperature in kelvin scale of band number 6.

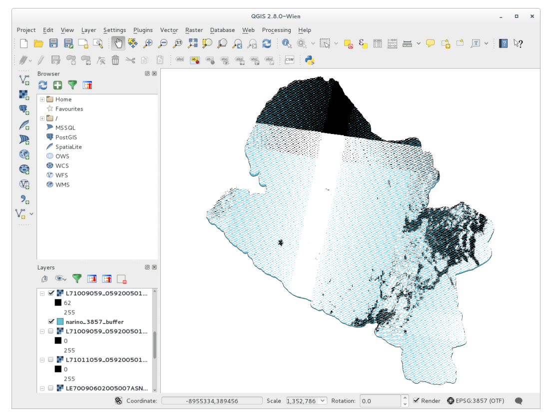
Table discarded: This table stores discarded data, by the next but not limited to: warm clouds, cold clouds, ambiguous data or not vegetation.

Table biomass: This table stores biomass map data from [3].

To process the imagery and load the data base was written an script, this script capture the Digital Number of each satellite image and transform it in reflectance value. In this image processing, the script was complemented adding filters to detect warm clouds, cold clouds, ambiguous data as is show by the algorithm proposed by [11],



(a) Department of Narino satellite imagery



(b) Department of Narino cut images

Figura 2. Preprocessing

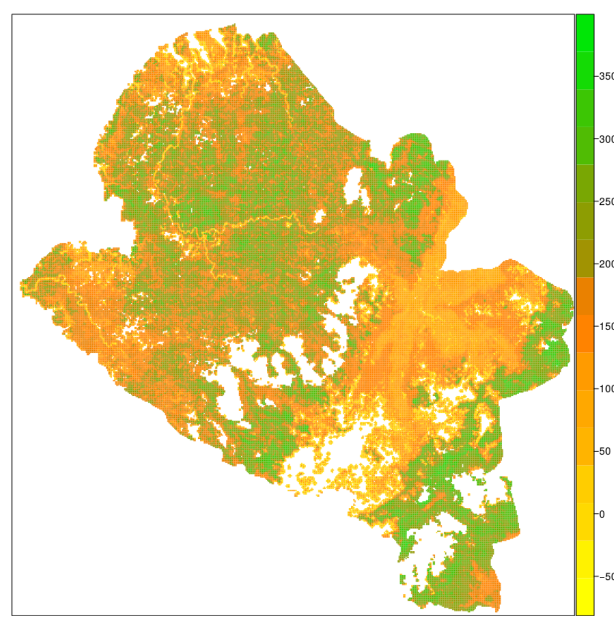


Figura 3. Department of Nariño biomass map years 2000-2003 [3]

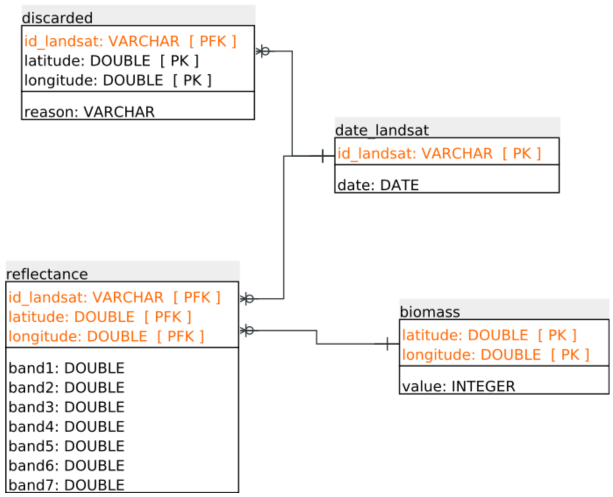


Figura 4. Landsat entity–relationship model

Tabla I. Data generated in process and data cleaning

Name	Value	Detail
Processed Landsat Images	1321	Department of Nariño images from years 2000 to 2014
Total data	51.076.512	Total records form years 2000 to 2014
Biomass data	81.993	Biomass records form years 2000 to 2003 by [3]
Warm cloud	3.731.768	Records form years 2000 to 2014
Cold cloud	27.827.009	Records form years 2000 to 2014
No vegetation	3.459.210	Records form years 2000 to 2014
Ambiguous	11.987.340	Records form years 2000 to 2014
Reflectance valid data	4.071.185	Records form years 2000 to 2014

in addition, the NVDI(normalized difference vegetation index) filter was apply to work with vegetation data only.

The table I shows the data obtained in this process.

D. Regression analysis

The regression analysis was performed using values from Landsat bands acquired from years 2000 and 2003, and biomas values obtained at [3], to acquire a better model the Landsat bands values were grouped and the average of those were calculated to each point, by iterating with values no greater than N samples number, N from 1 to 45 samples, the best model obtained got at least 35 samples per point, this data set has 1009 records. The behavior of the others iterations shows that less quantity of samples generates more records and greater quantity of samples generates low quantity of records, both cases the result is a not good model.

In table II is shown the results of metrics applied to analyzed models with 35 samples and 1009 data, which is the best model got by the research, this table was created using the open source library rminer presented by [12] to R tool.

Also, in this process was used the Boruta R package [13], this is a new features selection algorithm which found all relevant variables. The algorithm is designed as a wrapper to the random forest classification algorithm. The Boruta algorithm was used to establish if every Landsat bands used were relevant to find biomass, in figure 5 can be seen the relevance of Landsat bands to find biomass in the best model.

Tabla II. Metrics applied to analyzed models with 35 samples and 1009 data

	SAE	MAE	RAE	RMSE	COR	R2
ctree	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
pcr	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

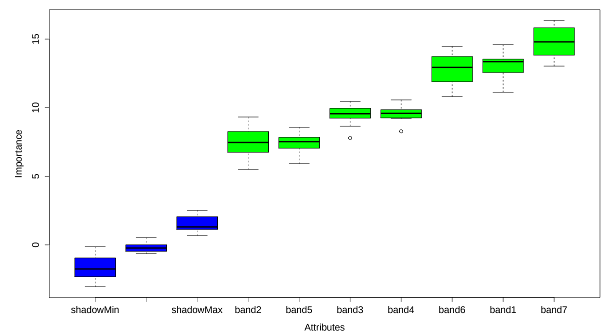


Figura 5. Relevance of Landsat bands int regression analysis

E. Maps construction

To build the biomass maps were used Kriging method, this method provide a solution to the estimation problem based on a continuous model of stochastic spatial variation, the Kriging´s objective is estimate the value of a random variable, Z, in one or more points not sampled or over large blocks.

The Kriging method has as input data from the sample, and a data matrix depending on the desired result image resolution, that is why the data samples were obtained applying the generated model in the regression analysis of grouped data by month, year and general in each point for years 2000 to 2014; and the data matrix was build with regular points with a 450 meters gap.

In figure 6, figure 7, and figure 8 is shown the monthly, yearly and general maps obtained for years 2000 a 2014 respectively.

IV. Conclusiones y Trabajos Futuros

Acknowledgment

This research was possible thanks to resources given by the General Royalty System in the framework of the project “Analysis of energy opportunities from alternative sources at the department of Nariño” implemented by the Electronic Engineering program of the University of 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

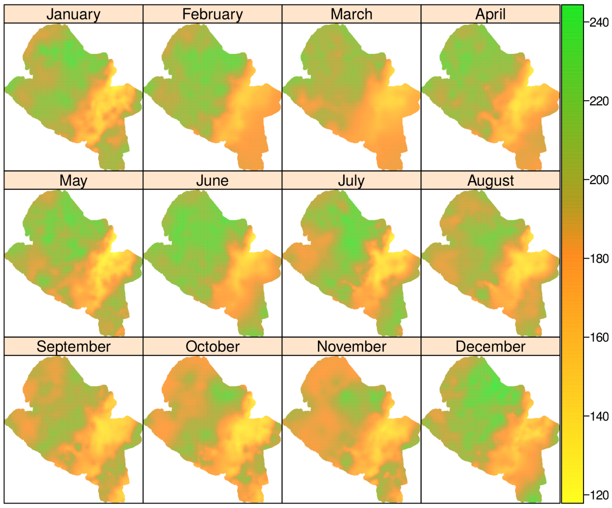


Figura 6. Monthly biomass maps

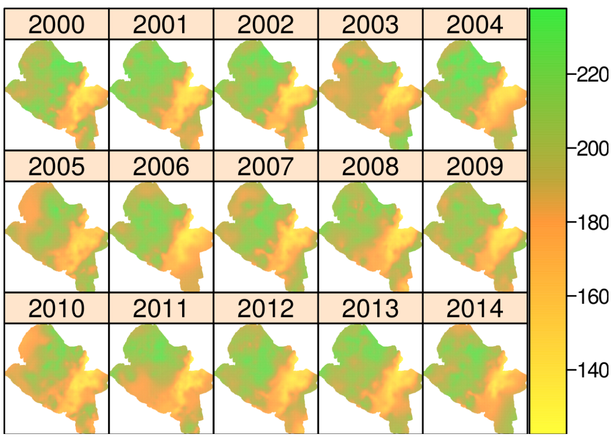


Figura 7. Yearly biomass maps

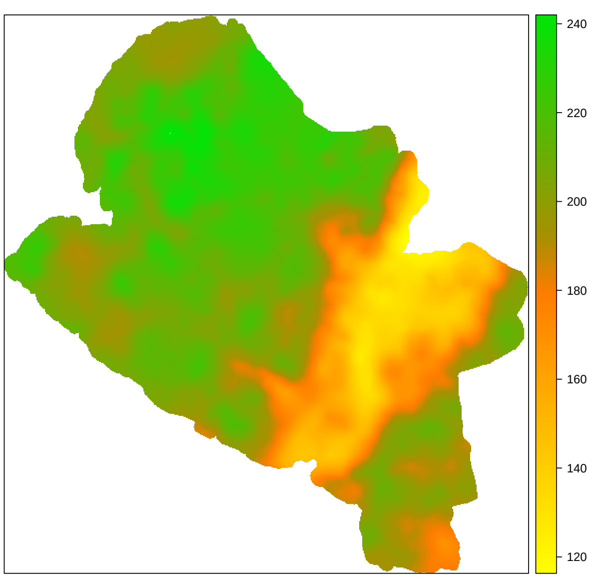


Figura 8. General biomass map years 2000-2014

[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 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.