



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO EN CIENCIAS FORESTALES

**ESTIMACIÓN REGIONAL DE
BIOMASA AÉREA EN LA PENÍNSULA
DE YUCATÁN MEDIANTE
TECNOLOGÍA GEOESPACIAL**

ALMA DELIA ORTIZ REYES

T E S I S

**PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:**

DOCTORA EN CIENCIAS

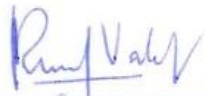
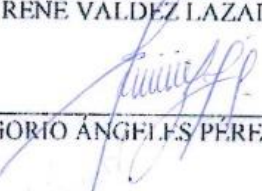
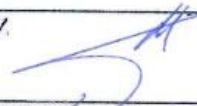

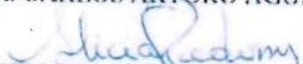
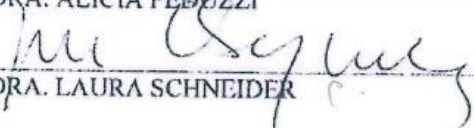
MONTECILLO, TEXCOCO, ESTADO DE MÉXICO

2019

La presente tesis titulada: “**Estimación regional de biomasa aérea en la península de Yucatán mediante tecnología geoespacial**” realizada por la alumna: “**Alma Delia Ortiz Reyes**” bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTORA EN CIENCIAS
CIENCIAS FORESTALES

CONSEJO PARTICULAR

CONSEJERO	 DR. JOSÉ RENÉ VALDEZ LAZALDE
ASESOR	 DR. GREGORIO ANGELES PÉREZ
ASESOR	 DR. HÉCTOR M. DE LOS SANTOS POSADAS
ASESOR	 DR. CARLOS ARTURO AGUIRRE SALADO
ASESORA	 DRA. ALICIA PEDUZZI
ASESORA	 DRA. LAURA SCHNEIDER

Montecillo, Texcoco, Estado de México, noviembre de 2019

ESTIMACIÓN REGIONAL DE BIOMASA AÉREA EN LA PENÍNSULA DE YUCATÁN MEDIANTE TECNOLOGÍA GEOESPACIAL

Alma Delia Ortiz Reyes, Dra.
Colegio de Postgraduados, 2019

RESUMEN

Caracterizar la distribución espacial e incertidumbre de la biomasa aérea en bosques tropicales sobre áreas extensas es factible mediante el uso combinado de datos de campo, LiDAR (Light Detection And Ranging) y datos espectrales. Se colectaron datos en tres niveles y se empleó una estrategia de modelado que permitió la estimación de biomasa aérea y su incertidumbre asociada a nivel regional, en dos tipos de selva mediana subperennifolia (SMSP) y subcaducifolia (SMSC) en la Península de Yucatán, México. En una primera fase, se relacionaron los datos de campo y LiDAR en franjas mediante regresión lineal múltiple y Random Forest (RF). El modelo de regresión transformado explicó en mayor proporción la varianza para ambos tipos de vegetación, mientras que RF arrojó los valores menores de RMSE. En la segunda fase, una cierta combinación de variables espectrales y climáticas explicaron la varianza en 50% mediante RF (RMSE = 34.2 Mg ha⁻¹ SMSP; 26.2 Mg ha⁻¹ SMSC). Finalmente se evaluó la incertidumbre, a nivel de pixel, mediante el algoritmo Quantile Regression Forests (QRF). Se reportan intervalos de incertidumbre de 0 a 250 Mg ha⁻¹ para SMSP y de 0 a 140 Mg ha⁻¹ para SMSC. Los resultados obtenidos apoyan la idea de utilizar datos LiDAR como una herramienta de muestreo, así como, variables auxiliares de diversas fuentes para estimar parámetros forestales de interés en áreas extensas. Contar con mapas de distribución espacial e incertidumbre de biomasa aérea a este nivel, ayudará a evaluar y proponer estrategias frente al desafío que supone el cambio climático global, sobre todo en bosques tropicales distribuidos en áreas geográficas extensas y dada su naturaleza compleja y dinámica.

Palabras clave: Bosque tropical, LiDAR, incertidumbre, Random Forest, Landsat

ESTIMATION OF ABOVE-GROUND BIOMASS AT REGIONAL LEVEL IN THE YUCATAN PENINSULA USING GEOSPATIAL TECHNOLOGY

Alma Delia Ortiz Reyes, Dra.
Colegio de Postgraduados, 2019

ABSTRACT

Characterizing spatial distribution and uncertainty of above-ground biomass in tropical forests over large areas is feasible through combined use of field data, LiDAR (Light Detection And Ranging) and spectral data. Data were collected at three levels and a modeling strategy was used, which allowed the estimation of above-ground biomass and its associated uncertainty at regional level, in two types of tropical forest: semi-evergreen (SETF) and semi-deciduous tropical forest (SDTF) in the Yucatan Peninsula, Mexico. First, field data and LiDAR data in strips were related by multiple linear regression and Random Forest (RF). Second, a particular combination of spectral and climatic variables explained the variance in 50% by RF (RMSE = 34.2 Mg ha⁻¹ SETF; 26.2 Mg ha⁻¹ SDTF). Finally, uncertainty was evaluated, at pixel level, using Quantile Regression Forests (QRF) algorithm. Uncertainty intervals of 0 to 250 Mg ha⁻¹ for SETF and 0 to 140 Mg ha⁻¹ for SMSC were reported. The results support the idea of using LiDAR data as a sampling tool, as well as, auxiliary variables from various sources to estimate forest parameters of interest in large areas. Having maps of spatial distribution and uncertainty of above-ground biomass at this level will help to evaluate and propose strategies to face the challenge by global climate change, especially in tropical forests distributed in large geographical areas and given their complex and dynamic nature.

Key words: Tropical forest, LiDAR, uncertainty, Random Forest, Landsat.

AGRADECIMIENTOS

Al **Consejo Nacional de Ciencia y Tecnología (CONACyT)** por el financiamiento necesario para efectuar esta investigación, otorgado en el marco del Programa Nacional de Posgrados de Calidad.

Al **Colegio de Postgraduados Campus Montecillo**, especialmente a todos los profesores, investigadores y personal del **Programa Forestal** quienes me enseñaron, ayudaron y de alguna u otra forma estuvieron involucrados en el proceso de mi formación académica y desarrollo profesional.

Al **Dr. José René Valdez Lazalde**, por su guía y apoyo para realizar este trabajo.

A cada uno de los integrantes del Comité asesor: **Dr. Gregorio Ángeles Pérez, Dr. Héctor M. De los Santos Posadas, Dr. Carlos Arturo Aguirre Salado, Dra. Alicia Peduzzi, Dra. Laura Schneider**, por el tiempo que dedicaron a asesorías y las aportaciones que hicieron para mejorar este trabajo.

A mis compañeros de posgrado, especialmente a **María de los Ángeles Soriano Luna** y **Ramiro Puc Kauil** por toda la ayuda y confianza que me brindaron.

Al **Centro Nacional de Investigación Disciplinaria en Conservación y Mejoramiento de Ecosistemas Forestales del Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias** por el apoyo otorgado en la última fase de este proceso.

DEDICATORIA

A mi mamita, Delfina Reyes Juárez y a mi papá Ranulfo Ortiz Juárez, con su ejemplo he aprendido a trabajar y seguir adelante.

A mi hermanito Luis Adolfo Ortiz Reyes, de cierta forma por ti estoy aquí, te quiero mucho.

A mi hermanito Jorge Luis Ortiz Reyes†, te quiero y te extraño tanto.

A Daisy Barrera Ortega, no estaría en este punto si no te hubiera encontrado en el camino.
Gracias por todo lo que has hecho por mí.

A la Sra. Marta Ortega y el Sr Lucio Barrera, que me han recibido en su familia y me han ayudado de muchas formas a hacer más fácil mi vida cotidiana.

A Frida y Lila, no lo saben, me hacen enojar y mucho más sonreír.

CONTENIDO

RESUMEN	iii
ABSTRACT.....	iv
LISTA DE FIGURAS.....	ix
LISTA DE CUADROS.....	x
INTRODUCCIÓN GENERAL	1
OBJETIVO GENERAL	4
OBJETIVOS PARTICULARES.....	4
REFERENCIAS	6
CAPÍTULO I. TRANSECTOS DE DATOS LiDAR: UNA ESTRATEGIA DE MUESTREO PARA ESTIMAR BIOMASA AÉREA EN ÁREAS FORESTALES ¹	8
RESUMEN.....	8
ABSTRACT	9
1.1 Introducción.....	10
1.2 Materiales y métodos.....	13
1.3 Resultados.....	21
1.4 Discusión	26
1.5 Conclusiones.....	32
1.6 Literatura citada.....	33
CAPÍTULO II. SINERGIA DE DATOS LANDSAT, CLIMÁTICOS Y LiDAR PARA ESTIMAR Y MAPEAR BIOMASA AÉREA EN SELVAS MEDIANAS DE LA PENÍNSULA DE YUCATÁN, MÉXICO.....	40
RESUMEN.....	40
ABSTRACT	41
2.1 Introducción.....	42
2.2 Materiales y métodos.....	45
2.3 Resultados.....	55
2.4 Discusión	59
2.5 Conclusiones.....	66
2.6 Literatura citada.....	67
2.7 Anexo 1.....	74

CAPÍTULO III. EVALUACIÓN DE LA INCERTIUMBRE ASOCIADA A LAS ESTIMACIONES DE BIOMASA AÉREA EN DOS SELVAS MEDIANAS EN LA PENÍNSULA DE YUCATÁN	78
RESUMEN.....	78
ABSTRACT	79
3.1 Introducción.....	80
3.2 Materiales y métodos.....	84
3.3 Resultados.....	86
3.4 Discusión	90
3.5 Conclusiones.....	96
3.6 Literatura citada.....	96
CONCLUSIONES GENERALES Y COMENTARIOS FINALES	101

LISTA DE FIGURAS

Figura 1.1 Distribución de las unidades de muestreo del Inventario Nacional y de Suelos (INFyS) de México, utilizados en el estudio y de las franjas de datos LiDAR, sobre los dos tipos de vegetación evaluados (Serie V, INEGI, 2013).	15
Figura 1.2 Importancia relativa de a) 20 variables seleccionadas para el modelo Random Forest en la selva mediana subperennifolia y b) de 13 variables en la selva mediana subcaducifolia.....	23
Figura 1.3 Predicción de la biomasa aérea usando los datos LiDAR en franjas	24
Figura 1.4 Mapas de biomasa aérea usando los datos LiDAR en franjas.....	25
Figura 2.1 Área de estudio: selvas mediana subcaducifolia y mediana subperennifolia.....	46
Figura 2.2 Importancia relativa de las variables seleccionadas por el modelo Random Forest en la selva mediana subperennifolia (12 variables) (izq.) y la selva mediana subcaducifolia (15 variables) (der.), respectivamente, para la estimación de biomasa.	56
Figura 2.3 Distribución espacial de la biomasa aérea en las selvas mediana subperennifolia y mediana subcaducifolia de la Península de Yucatán, México.	58
Figura 3.1 Distribución espacial de biomasa aérea en la selva mediana subcaducifolia y mediana subperennifolia de la Península de Yucatán, México.....	85
Figura 3.2 Distribución de la incertidumbre de biomasa aérea (Mg ha^{-1}) en los dos tipos de selva mediana evaluados. La incertidumbre de las estimaciones se expresó como intervalos de predicción al 95%.	88
Figura 3.3 Distribución de la incertidumbre de biomasa aérea (%) expresado como la semiamplitud de los intervalos de predicción al 95% dividido entre su media en los dos tipos de selva mediana evaluados.	89

LISTA DE CUADROS

Cuadro 1.1. Estadísticas descriptivas de los 363 sitios de muestreo que coinciden espacialmente con las franjas de datos LiDAR.....	16
Cuadro 1.2. Resumen de las variables predictoras (métricas LiDAR) empleados en la estimación de biomasa aérea.....	18
Cuadro 1.3. Modelos de regresión lineal múltiple que relacionan las métricas LiDAR con la biomasa aérea por tipo de vegetación.	22
Cuadro 1.4. Resultados de los modelos Random Forest para la estimación de biomasa aérea por tipo de vegetación.....	22
Cuadro 2.1. Características de las escenas Landsat 8 procesadas a nivel de reflectancia en superficie.....	50
Cuadro 2.2. Variables predictoras (espectrales y climáticas) empleadas en el modelamiento de la biomasa aérea.	51
Cuadro 2.3. Parámetros relevantes de los modelos Random Forest para la estimación de biomasa aérea por tipo de selva mediana.....	55

INTRODUCCIÓN GENERAL

Los bosques juegan un rol importante en el sostenimiento y mantenimiento de los sistemas y ciclos ecológicos, de la existencia de los bosques depende la eficacia de múltiples procesos responsables de la captura y almacenamiento del carbono, el suministro constante de agua o el control de la erosión (Patterson & Coelho, 2009; Pan *et al.*, 2013).

Así mismo, es importante conocer con precisión la magnitud de los recursos forestales existentes y los procesos que se detonan a partir de los diversos cambios que se manifiestan en el bosque (Miranda-Aragón *et al.*, 2012). La comprensión que se logre mediante el monitoreo de los procesos de incorporación o de disminución a causa del cambio de uso de la tierra, permitirá a los especialistas y tomadores de decisiones evaluar en qué medida los bosques contribuyen a reducir las emisiones netas de gases de efecto invernadero (Keenan *et al.*, 2015). El conocimiento de estos cambios es esencial para la formulación de políticas, el financiamiento y la toma de decisiones sobre gestión forestal a nivel nacional, regional y mundial (FAO, 2015).

En este sentido, la Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC) ha buscado facilitar el monitoreo forestal a gran escala para la reducción de la deforestación y degradación de bosque, a través de enfoques robustos en la gestión de información forestal y de los inventarios forestales de cada país, haciendo hincapié en el potencial que la percepción remota ofrece durante la ejecución de los mismos, especialmente en aquellos países con infraestructura limitada.

Cuando se requieren estimaciones precisas sobre los cambios y procesos que modifican el paisaje, las herramientas de percepción remota son fundamentales para alcanzar dichos objetivos. De igual forma, son consideradas elementos cruciales en la implementación exitosa de los mecanismos REDD (reducción de emisiones por deforestación y degradación); puesto que

ofrecen mediciones, confiables y precisas de los recursos forestales, lo que resulta en una mejor comprensión sobre el funcionamiento de los ecosistemas (Asner, 2009; Wulder *et al.*, 2012; Pan *et al.*, 2013).

Por otro lado, a pesar de que en muchos países el establecimiento de sus inventarios forestales nacionales está bien definido, faltan muchos objetivos por alcanzar. Los especialistas, técnicos forestales, científicos y tomadores de decisiones concuerdan en que aún se necesitan cuantificaciones y representaciones más precisas sobre las existencias de carbono y biomasa aérea (Asner *et al.*, 2012) y en general de los indicadores y parámetros que estén relacionados a la mitigación de los efectos del cambio climático.

En este sentido, en nuestro país existen estudios tanto a nivel nacional como regional (en diferentes ecosistemas) para cuantificar, estimar y/o mapear las variables esenciales para los reportes sobre el cambio climático, como la cobertura del suelo (Colditz *et al.*, 2014), biomasa aérea y/o carbono aéreo (Aguirre-Salado *et al.*, 2014; Cartus *et al.*, 2014) la identificación de áreas deforestadas (Miranda-Aragón *et al.*, 2012) o los patrones de distribución de especies por mencionar algunos.

De manera particular, se han concentrado esfuerzos para cuantificar las existencias de carbono y biomasa aérea en bosques tropicales dadas sus características de rápido crecimiento, el papel que juegan dentro del ciclo global de carbono, así como, las presiones a las que han sido expuestos debido al cambio de uso de suelo. En este sentido, la Península de Yucatán se caracteriza por su condición de mosaicos sucesionales de bosques secundarios derivado de disturbios antrópicos como las técnicas agrícolas: roza-tumba y quema; la conversión de tierras forestales a agrícolas y/o pecuarias, y por disturbios naturales: como huracanes e incendios (Aryal *et al.*, 2014). Por lo que entender la dinámica de carbono en este tipo de bosques es requerido para el desarrollo de

estrategias de manejo y conservación en bosques tropicales. Sin embargo, como en la mayoría de los casos, las estimaciones de biomasa y/o carbono aéreo en bosques tropicales que cubren áreas extensas aún tienen un alto nivel de incertidumbre.

Debido a la variación natural y constante de la estructura del bosque y biomasa en estos ecosistemas, el uso de parcelas de campo es ineficiente. Se requieren muchas parcelas y visitas periódicas para realizar mediciones. Por lo tanto, se necesitan enfoques que enlacen las mediciones en campo y los datos de observaciones satelitales para afrontar este problema de forma factible.

La tecnología LiDAR (Light Detection And Ranging) se ha propuesto como el puente que enlaza dichas mediciones. Puede ser usado para mapear biomasa aérea en áreas extensas cuando se calibra con datos de parcelas en campo y dichos mapas pueden ser una línea base de análisis cuando se conjunta con observaciones de imágenes satelitales.

Disminuir el tiempo para obtener datos de campo y emplear estrategias de muestreo más robustos hace de LiDAR una herramienta ideal para el mapeo de biomasa aérea. Adicionalmente, las metodologías que aprovechan tales relaciones mejoran aún más cuando se incorporan productos basados en imágenes satelitales. Lo anterior permite un programa de monitoreo y evaluación de las existencias de carbono a escala regional que apoyan los diferentes programas de mitigación de cambio climático, manejo y conservación en regiones tropicales heterogéneas.

El hecho de contar con información y resultados a este nivel puede proporcionar mapas más precisos y con un nivel de incertidumbre asociado de biomasa aérea, lo cual ayuda a evaluar las respuestas que los ecosistemas tropicales muestran ante el escenario de cambio climático global.

Con base en lo anterior el propósito de este documento es proponer una metodología para realizar el escalamiento de las estimaciones de biomasa aérea desde un nivel local a uno regional,

usando datos del Inventario Nacional Forestal y de Suelos (INFyS), de imágenes satelitales de mediana resolución y datos LiDAR, en dos tipos de selva mediana de la Península de Yucatán.

La presente tesis consta de tres capítulos que se presentan como manuscritos individuales en forma de artículos, cada uno de ellos pretende abordar los objetivos que se plantean en la siguiente sección. De manera particular, el capítulo I detalla los dos enfoques estadísticos empleados para establecer relaciones entre datos de sitios de inventario y datos LiDAR para estimar la biomasa aérea en franjas que se registraron en el área de estudio. El capítulo II presenta un método donde a partir de un número mayor de observaciones de biomasa aérea “parcelas LiDAR”, espacialmente continuas, se modela la biomasa aérea en los dos tipos de vegetación evaluados, usando variables auxiliares de imágenes satelitales, como variables predictoras. En el capítulo III se evalúa la incertidumbre asociada, a nivel pixel, a partir de intervalos de predicción. El estudio se limitó al área que cubre la selva mediana subperennifolia y subcaducifolia en la Península de Yucatán. Finalmente, el documento aporta una serie de conclusiones generales que pone de manifiesto los logros y oportunidades de mejora de este trabajo.

OBJETIVO GENERAL

Proponer una metodología para realizar el escalamiento de las estimaciones de biomasa aérea desde un nivel local a uno regional, usando datos del Inventario Nacional Forestal y de Suelos, de imágenes satelitales de mediana resolución y datos LiDAR, en las selvas medianas subperennifolia y subcaducifolia de la Península de Yucatán.

OBJETIVOS PARTICULARES

1.- Estimar y mapear la biomasa aérea en transectos a partir de datos de LiDAR registrados en franjas y mediciones de sitios del Inventario Nacional Forestal y de Suelos.

2.- Combinar la información de las imágenes de mediana resolución (Landsat) con estimaciones de biomasa aérea derivadas de datos LiDAR para evaluar la fusión de datos multiespectrales en las estimaciones y mapeo de biomasa aérea en dos tipos de selva mediana de la Península de Yucatán.

3.- Evaluar el nivel de incertidumbre derivado del escalamiento de las estimaciones de la biomasa aérea en dos tipos de selva mediana de la Península de Yucatán.

REFERENCIAS

- Aguirre-Salado, C. A., Treviño-Garza, E. J., Aguirre-Calderón, O. A., Jiménez-Pérez, J., González-Tagle, M. A., Valdez-Lazalde, J. R., & Miranda-Aragón, L. (2014). Mapping aboveground biomass by integrating geospatial and forest inventory data through a k-nearest neighbor strategy in North Central Mexico. *Journal of Arid Land*, 6(1), 80-96. doi:10.1007/s40333-013-0191-x
- Aryal, D. R., De Jong, B. H. J., Ochoa-Gaona, S., Esparza-Olguin, L., & Mendoza-Vega, J. (2014). Carbon stocks and changes in tropical secondary forests of southern Mexico. *Agriculture, Ecosystems & Environment*, 195, 220-230. doi:10.1016/j.agee.2014.06.005
- Asner, G. P. (2009). Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters*, 4(3). doi:10.1088/1748-9326/4/3/034009
- Asner, G. P., Mascaro, J., Muller-Landau, H. C., Vieilledent, G., Vaudry, R., Rasamoelina, M., . . . van Breugel, M. (2012). A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia*, 168(4), 1147-1160. doi:10.1007/s00442-011-2165-z
- Cartus, O., Kellndorfer, J., Walker, W., Franco, C., Bishop, J., Santos, L., & Fuentes, J. (2014). A National, Detailed Map of Forest Aboveground Carbon Stocks in Mexico. *Remote Sensing*, 6(6), 5559-5588. doi:10.3390/rs6065559
- Colditz, R. R., Pouliot, D., Llamas, R. M., Homer, C., Latifovic, R., Ressler, R. A., & Richardson, K. (2014). Detection of North American land cover change between 2005 and 2010 with 250m MODIS data. *PE&RS, Photogrammetric Engineering & Remote Sensing*, 80(10), 918-924.
- FAO. (2015). *Evaluación de los recursos forestales mundiales 2015*. Food Agriculture Organization of the United Nations.
- Keenan, R. J., Reams, G. A., Achard, F., de Freitas, J. V., Grainger, A., & Lindquist, E. (2015). Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management*, 352, 9-20. doi:10.1016/j.foreco.2015.06.014
- Miranda-Aragón, L., Treviño-Garza, E. J., Jiménez-Pérez, J., Aguirre-Calderón, O. A., González-Tagle, M. A., Pompa-García, M., & Aguirre-Salado, C. A. (2012). Modeling susceptibility to deforestation of remaining ecosystems in North Central Mexico with logistic regression. *Journal of Forestry Research*, 23(3), 345-354. doi:10.1007/s11676-012-0230-z

- Pan, Y., Birdsey, R. A., Phillips, O. L., & Jackson, R. B. (2013). The Structure, Distribution, and Biomass of the World's Forests. *Annual Review of Ecology, Evolution, and Systematics*, 44(1), 593-622. doi:10.1146/annurev-ecolsys-110512-135914
- Patterson, T. M., & Coelho, D. L. (2009). Ecosystem services: Foundations, opportunities, and challenges for the forest products sector. *Forest Ecology and Management*, 257(8), 1637-1646. doi:10.1016/j.foreco.2008.11.010
- Wulder, M. A., White, J. C., Nelson, R. F., Næsset, E., Ørka, H. O., Coops, N. C., . . . Gobakken, T. (2012). Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*, 121, 196-209. doi:10.1016/j.rse.2012.02.001

CAPÍTULO I. TRANSECTOS DE DATOS LiDAR: UNA ESTRATEGIA DE MUESTREO PARA ESTIMAR BIOMASA AÉREA EN ÁREAS FORESTALES¹

RESUMEN

La estimación y el mapeo de la biomasa aérea sobre áreas extensas puede realizarse haciendo uso de las herramientas que ofrece la percepción remota. El objetivo de este estudio fue estimar la biomasa aérea de dos tipos de selva mediana: subperennifolia (SMSP) y subcaducifolia (SMSC) en la Península de Yucatán, México; empleando métricas generadas a partir de datos Light Detection and Ranging (LiDAR). Se usaron datos de 365 unidades de muestreo del Inventario Nacional Forestal y de Suelos (INFyS) de México para calibrar modelos de biomasa aérea usando regresión lineal múltiple y Random Forest (RF). Con estos modelos se mapeó la biomasa aérea sobre franjas de datos LiDAR. El modelo de regresión transformado logró explicar la varianza en un 62% (RMSE = 41.44 Mg ha⁻¹ para SMSP y 36.60 Mg ha⁻¹ para SMSC) para ambos tipos de vegetación. Los modelos generados a través de RF lograron explicar la varianza en un 57% (RMSE = 40.73 Mg ha⁻¹) para la SMSP y sólo de 52% (RMSE = 35.10 Mg ha⁻¹) para la SMSC. El desfase entre la toma de datos en campo y LiDAR, así como el error en la precisión de las coordenadas de los sitios de inventario, son factores reconocidos que influyeron en los resultados. A pesar de lo anterior, las estimaciones obtenidas podrían servir de base para estimar el inventario completo de biomasa en el área de estudio incorporando datos espectrales derivados de un sensor remoto que cubra la totalidad de la misma.

Palabras clave: franjas LiDAR; inventario forestal; muestreo; Península de Yucatán.

¹ Ortiz-Reyes, A. D., Valdez-Lazalde, J. R., Ángeles-Pérez, G., De los Santos-Posadas, H. M., Schneider, L., Aguirre-Salado, C. A., & Peduzzi, A. (2019). Transectos de datos LiDAR: una estrategia de muestreo para estimar biomasa aérea en áreas forestales. *Madera y Bosques*, 25(3), e2531872. doi: 10.21829/myb.2019.2531872

LiDAR DATA TRANSECTS: A SAMPLING STRATEGY TO ESTIMATE ABOVEGROUND BIOMASS IN FOREST AREAS

ABSTRACT

Estimating and mapping above-ground biomass over large areas can be done using remote sensing tools. The objective of this study was to estimate above-ground biomass of two types of tropical forest: semi-evergreen (SETF) and semi-deciduous tropical forest (SDTF) in the Yucatan Peninsula, Mexico using metrics obtained from LiDAR (Light Detection and Ranging) data. 365 sampling plots obtained from National Forest and Soils Inventory of Mexico were used to calibrate multiple linear regression and Random Forest models to estimate above-ground biomass along LiDAR strips. The transformed regression model explained the variance by 62% (RMSE = 41.44 Mg ha⁻¹ for SETF and 36.60 Mg ha⁻¹ for SDTF) for both types of vegetation. The models of Random Forest explained the variance by 57% (RMSE = 40.73 Mg ha⁻¹) for SETF and only 52% (RMSE = 35.10 Mg ha⁻¹) by SDTF. The mismatch between the field data and LiDAR data, as well as the error in the precision of the coordinates of the inventory plots, were recognized as factors that influenced the results. In spite of the above, the estimates obtained could serve as a basis to estimate the complete biomass inventory in the study area by incorporating spectral data derived from a remote sensor that covers the entire area.

Key words: LiDAR strips; forest inventory; sampling; Yucatan Peninsula.

1.1 INTRODUCCIÓN

Los bosques tropicales son los ecosistemas vegetales más biodiversos del mundo. Cuentan con aproximadamente la mitad de la biomasa viva de los ecosistemas terrestres (Houghton *et al.*, 2009) y juegan un papel importante en el ciclo de carbono global debido a su naturaleza de rápido crecimiento y acumulación de carbono atmosférico (Meyer *et al.*, 2013). Por lo tanto, la estimación precisa de su biomasa aérea es de gran interés para los gestores de estos recursos.

Las mediciones en campo y el muestreo destructivo son consideradas como la metodología más precisa para estimar biomasa (Zolkos *et al.*, 2013); sin embargo, su aplicación es impráctica cuando la región de interés es muy extensa o la frecuencia de evaluación es corta, debido a los costos asociados con la toma de datos. Estos problemas se magnifican en los bosques tropicales debido a la heterogeneidad del dosel, lo cerrado de los estratos y la irregularidad de los fustes (Xu *et al.*, 2016).

El uso de las herramientas de percepción remota ha incrementado la capacidad de realizar las estimaciones, especialmente en los bosques tropicales (Houghton *et al.*, 2009; Gonçalves *et al.*, 2017) que regularmente presentan un patrón de agregación y variabilidad alta en distancias espaciales cortas, lo que resulta en un mosaico complejo que es difícil de muestrear en su totalidad (Barbosa *et al.*, 2014).

Los datos provenientes de sensores LiDAR, con capacidad de derivar información de la estructura del dosel en tres dimensiones pueden usarse para generar información que una vez calibrada y validada con datos de campo, serían una alternativa para el mapeo continuo de la biomasa forestal (Hansen *et al.*, 2014).

De acuerdo con Baltsavias (1999), el sistema LiDAR funciona bajo el principio de medir el tiempo de viaje entre el pulso emitido por el sensor láser y el pulso que recibe del objetivo. La

medición de la distancia se obtiene al multiplicar este intervalo de tiempo por la velocidad de la luz y dividirlo entre dos. En su trayectoria, la señal láser proporciona información de varios objetivos, penetrando a través del dosel, lo que resulta en productos como nubes de puntos de múltiples retornos o formas de onda completas que describen la estructura de la vegetación en tres dimensiones: x, y, z (Næsset & Økland, 2002).

A pesar de la utilidad de los datos LiDAR, no siempre es posible contar con una cobertura completa (de pared a pared) de datos para la zona de interés. Por lo anterior, en la mayoría de los estudios regionales LiDAR se ha usado como un mecanismo de muestreo en fases (Saarela *et al.*, 2015). El muestreo de LiDAR en franjas es una forma de obtener información completa, precisa y uniforme, tanto espacial como temporalmente, teniendo la ventaja (comparado con un inventario tradicional) de reducir el tiempo invertido y la intensidad del trabajo requerido. Para reducir la incertidumbre en las estimaciones de biomasa, es necesario procurar la realización de observaciones espacialmente continuas que sean lo suficientemente finas en capturar la variabilidad sobre una superficie (Hansen *et al.*, 2014; Saarela *et al.*, 2015).

Las estimaciones derivadas de LiDAR pueden ser un recurso clave en los sistemas de monitoreo forestal que se desarrollan en el contexto de los esfuerzos de mitigación del cambio climático como el mecanismo REDD+ (Reducción de Emisiones por Deforestación y Degradación Forestal) (Asner *et al.*, 2012), pues será posible el mapeo y monitoreo de la biomasa y carbono de los bosques tropicales en áreas geográficas grandes.

Existen diversos trabajos de estimación de biomasa aérea en bosques tropicales del mundo. Cada uno emplea diferentes enfoques metodológicos que van desde el modelado estadístico con técnicas paramétricas (Clark *et al.*, 2011; Asner *et al.*, 2012) y no paramétricas (Mascaro *et al.*, 2014). Así mismo utilizan diferentes datos de sensores ópticos pasivos como MODIS (Moderate

Resolution Imaging Spectroradiometer) (Saatchi *et al.*, 2007; Baccini *et al.*, 2008) o Landsat (Foody *et al.*, 2003), sensores activos como LiDAR (Cao *et al.*, 2014; Laurin *et al.*, 2016) y SAR (Ningthoujam *et al.*, 2018) o una combinación de los anteriores (Wang & Qi, 2008; Phua *et al.*, 2017). Todos ellos tratando de encontrar un enfoque de estimación de biomasa que genere los mejores resultados.

En México, recientemente se han realizado estimaciones de biomasa a nivel nacional. En ellos han utilizado datos del INFyS (de diferentes periodos de medición), escalas diferentes y técnicas de aprendizaje de máquinas variadas (Cartus *et al.*, 2014; Urbazaev *et al.*, 2018). Rodríguez-Veiga *et al.* (2016) han combinado sensores ópticos activos y pasivos, y Nelson *et al.* (2017) utilizaron datos de dos tipos de LiDAR, satelital y aerotransportado para probar tres tipos de modelos paramétricos. La varianza explicada por los modelos en los trabajos anteriores va desde 31% hasta 68%.

Los estudios anteriormente señalados analizan puntualmente a la Península de Yucatán como zona clave para centrar esfuerzos y conocer el estado actual de las reservas de carbono, dados los altos niveles de incertidumbre en las estimaciones así como su importancia ecológica. Es una región contrastante con tasas altas de deforestación y un mosaico de rodales secundarios en diferentes estados sucesionales, producto de las rotaciones de cultivo, el abandono de tierras agrícolas o pastizales y los disturbios naturales como incendios o huracanes.

En este trabajo se muestra un enfoque basado en área, en el cual se vinculan los datos de las mediciones en campo, a través del modelado, con las estadísticas y generalidades espaciales derivadas de la distribución vertical de los datos LiDAR, es decir, el recorte de la nube de puntos LiDAR correspondiente a los sitios del INFyS (White *et al.*, 2013). Se empleó una base de datos espaciales y ecuaciones alométricas nuevas generadas para este tipo de selvas. Se ensayaron dos

métodos estadísticos, uno paramétrico y otro no paramétrico para registrar el escalamiento de estimaciones de biomasa aérea de un nivel local (datos de campo) a un segundo nivel de muestreo (franjas). Las estimaciones a nivel franja tienen como propósito aumentar la muestra de campo original y con ella facilitar estimaciones posteriores de biomasa nivel regional. Este trabajo tuvo como objetivo estimar la biomasa aérea en franjas de selva mediana subperennifolia y subcaducifolia en la Península de Yucatán, México. Se utilizaron transectos de datos LiDAR e información del INFyS. De manera particular el estudio se enfocó a 1) evaluar las métricas LiDAR (mejor correlacionadas con la biomasa aérea en campo) y su combinación para la estimación de biomasa aérea en dos tipos de selva mediana, 2) evaluar dos técnicas de modelado para estimar biomasa, un enfoque estadístico de regresión de mínimos cuadrados ordinarios y un enfoque de aprendizaje de máquinas, Random Forest y 3) mapear la distribución espacial de la biomasa aérea en las franjas con datos LiDAR. Lo anterior permitirá incrementar espacialmente la muestra de datos de biomasa y posteriormente realizar estimaciones en toda el área de estudio utilizando datos derivados de imágenes satelitales.

1.2 MATERIALES Y MÉTODOS

Área de estudio

Este estudio se llevó a cabo en dos tipos de selvas medianas: SMSP y SMSC de la Península de Yucatán, México, localizadas en la porción SE de la República Mexicana ($17^{\circ}00'$ y $21^{\circ}45'$ N, $86^{\circ}30'$ y $9^{\circ}30'$ O). La superficie aproximada de la península es de $141\,523\text{ km}^2$, equivalente al 7% del territorio nacional, de los cuales, $28\,171\text{ km}^2$ corresponden a SMSC y $48\,699\text{ km}^2$ a la SMSP (Instituto Nacional de Estadística y Geografía [INEGI], 2013) (Fig. 1.1). La selva mediana subcaducifolia se distribuye en la parte centro y norte de la península de Yucatán, en forma de franja. En la selva mediana subperennifolia, los componentes arbóreos pierden

estacionalmente su follaje en un 25% a 50% durante la época de sequía. Se establece en regiones planas, ligeramente más secas y con drenaje rápido de la península (Pennington, 2005).

Datos de campo

Los datos del INFyS fueron recolectados por la Comisión Nacional Forestal [Conafor] (2015), en el área de estudio durante el periodo 2009 - 2014 (Fig. 1.1) bajo un diseño de muestreo sistemático por conglomerados. Cada conglomerado está integrado por cuatro sitios rectangulares de 400 m² donde se midió y registró el arbolado cuyo diámetro normal a la altura de 1.3 m sobre la superficie del suelo, fue igual o mayor a 7.5 cm (Conafor, 2011). Un total 187 sitios coincidieron espacialmente con las franjas de datos LiDAR, en áreas que corresponden a la SMSC y 176 sitios en la SMSP. Las variables dasométricas consideradas a nivel árbol fueron el diámetro normal (DN), la altura total (AT), el nombre de la especie y la condición del individuo (vivo o muerto). El INFyS indicó que las coordenadas geográficas de los sitios de muestreo tienen un error promedio de 5.9 m (Conafor, 2015).

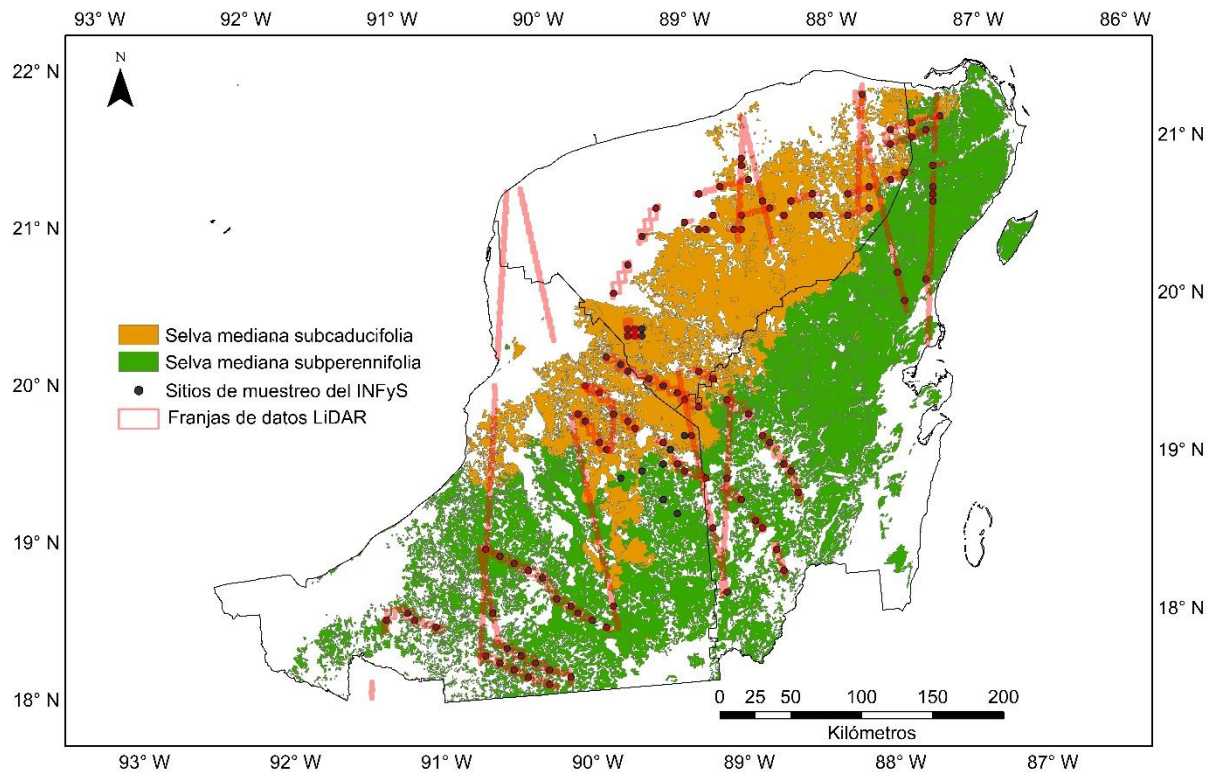


Figura 1.1 Distribución de las unidades de muestreo del Inventario Nacional y de Suelos (INFyS) de México, utilizados en el estudio y de las franjas de datos LiDAR, sobre los dos tipos de vegetación evaluados (Serie V, INEGI, 2013).

El INFyS indicó que las coordenadas geográficas de los sitios de muestreo tienen un error promedio de 5.9 m (Conafor, 2015). Se encontraron superposiciones entre algunos sitios, por lo que se optó por reconstruir los polígonos tomando en cuenta las medidas de diseño para el establecimiento de las parcelas en campo (Conafor, 2011). En el Cuadro 1.1 se presenta un resumen de los datos dasométricos de los sitios de muestreo.

Cuadro 1.1. Estadísticas descriptivas de los 363 sitios de muestreo que coinciden espacialmente con las franjas de datos LiDAR.

Variable	Selva mediana subperennifolia (176)		Selva mediana subcaducifolia (187)	
	Media	Desviación estándar	Media	Desviación estándar
Área basal (m ² ha ⁻¹)	16.27	9.32	14.14	8.48
Altura media (m)	9.24	2.18	8.91	1.52
Biomasa aérea (Mg ha ⁻¹)	89.38	62.26	71.50	50.84

Los datos de campo sirvieron de base para calcular la biomasa aérea mediante ecuaciones alométricas. La biomasa de individuos con DN < 10 cm se estimó con ecuaciones para la especie cuando estuvo disponible (Cairns *et al.*, 2003; Puc-Kauil *et al.*, 2019), cuando la especie no contaba con una ecuación específica, se empleó la ecuación local de Puc-Kauil *et al.* (2019) para diámetros menores a 10 cm. Para individuos con DN > 10 cm se usaron las ecuaciones disponibles para cada especie (Cairns *et al.*, 2003), caso contrario, se investigó su valor de densidad de la madera y, tomando como referencia la base de datos de densidad de la madera global de Zanne *et al.* (2013), se aplicó la ecuación global de Urquiza-Haas *et al.* (2007). El valor de densidad de la madera mejora las predicciones de biomasa, debido a que las formas de crecimiento de las especies están en función de ésta (Chave *et al.*, 2004). Para las especies que no tenían registrado un valor de densidad de la madera y no tenían una ecuación específica se aplicó la ecuación global de Cairns *et al.* (2003) para diámetros mayores a 10 cm. En el caso de palmas y lianas, se usaron las ecuaciones propuestas por Chave *et al.* (2003).

Datos LiDAR

Los datos LiDAR fueron adquiridos en 2013 por la NASA (National Aeronautics & Space Administration) mediante el sistema aerotransportado G-LiHT (Goddard's LiDAR, Hyperspectral & Thermal Imager). Este utilizó un escáner láser Riegl VQ-480, con una

frecuencia de pulso de hasta 300 Khz. La velocidad de medición efectiva es de hasta 150 Khz a lo largo de una franja perpendicular a la dirección de vuelo de 60°. La huella del láser fue de 10 cm de diámetro, con un máximo de retornos por pulso de ocho y una densidad de seis pulsos por m² en la superficie. La altitud de vuelo fue de 335 m (Cook *et al.*, 2013). Los datos se registraron a lo largo de franjas y clasificados en rutas para un mejor control de la información (Fig. 1.1). También se utilizaron los modelos digitales del terreno disponibles para normalizar la nube de puntos a alturas por encima del terreno.

Métricas LiDAR

Los datos LiDAR se procesaron en el software FUSION/LDV Versión 3.60+ (McGaughey, 2016) para extraer, a nivel de unidad de muestreo, las métricas que posteriormente se utilizaron como variables predictoras de los modelos de biomasa aérea. Las métricas representan la distribución de altura de la vegetación y se dividen en tres grupos: 1) estadísticas que describen la distribución de las alturas correspondientes a los retornos del láser en una parcela o área determinada (p. ej. media, desviación estándar, moda), 2) percentiles de la distribución de la altura (P05, P10, P50, P70, P95) y 3) métricas de densidad que corresponden a la proporción de retornos del láser dentro de cierta porción de altura o por encima o debajo de cierta altura, conocida como altura de corte (Cuadro 1.2). La descripción detallada de las métricas se encuentra en el manual del software FUSION (McGaughey, 2016). Se definió una altura de corte de 2 m para disminuir el ruido de los puntos próximos al suelo, causado por el sotobosque y fallas en el filtrado de los puntos del terreno. Se obtuvieron productos ráster por cada métrica calculada (Cuadro 2) y corresponden a las franjas de datos LiDAR. El tamaño de pixel fue de 20 m, equivalente al tamaño de las parcelas en campo (400 m²).

Cuadro 1.2. Resumen de las variables predictoras (métricas LiDAR) empleados en la estimación de biomasa aérea.

Tipo de variable predictorora	Descripción	Notación
Estadísticas descriptivas de altura.	Altura media	elev_ave_2plus_20METERS
	Altura máxima	elev_max_2plus_20METERS
	Altura media generalizada cuadrática	elev_quadratic_mean_20METERS
	Altura media generalizada cúbica	elev_cubic_mean_20METERS
	Desviación estándar de la altura	elev_stddev_2plus_20METERS
Percentiles de altura (14)	Percentiles 05-99 de altura	elev_P05-P99_2plus_20METERS
Métricas de densidad	Tasa de relieve del dosel	elev_canopy_relief_ratio_20METERS
	Primeros retornos encima de la altura media	X1st_cnt_above_mean_20METERS
	% de los primeros retornos encima de la altura media	X1st_cover_above_mean_20METERS
	% de todos los retornos encima de 2 m	all_cover_above2_20METERS
	% de los primeros retornos encima de 2 m	X1st_cover_above2_20METERS
	% de todos los retornos encima de la altura media	all_cover_above_mean_20METERS
	(Todos los retornos encima de 2 m) / (Total de primeros retornos)*100	all_1st_cover_above2_20METERS
	(Todos los retornos encima de la altura media) / (Total de primeros retornos) * 100	all_1st_cover_above_mean_20METERS

Estimación de biomasa aérea basada en datos LiDAR

Análisis de regresión

Inicialmente se identificaron correlaciones de Pearson ($r \geq 0.5$) entre las métricas de LiDAR (Cuadro 1.2) obtenidas en las unidades de muestreo y las estimaciones de biomasa para los mismos sitios calculadas mediante ecuaciones alométricas. Se probaron modelos lineales múltiples para estimar la biomasa aérea en función de las métricas LiDAR, empleando la función “lm” del paquete “stats” del software R (R Development Core Team, 2013). Sin embargo, se

optó por construir los modelos con base en lo señalado por Ortiz-Reyes *et al.* (2015), en donde incorporaron características notorias al modelo, es decir, variables que describen tanto la estructura horizontal como vertical del dosel. La estructura horizontal considera la densidad del arbolado. La estructura vertical es descrita por la distribución de altura. Las métricas “primeros retornos por encima de la altura media” y el “percentil 95” son ejemplos de cada uno de estos planos.

Los primeros modelos no cumplían con los supuestos de la regresión, por lo que se consideró mejorarlos mediante la transformación Box-Cox a través de la librería MASS del software R, para conseguir una varianza constante y residuales distribuidos en forma normal. Adicionalmente se utilizó el factor de inflación de la varianza (FIV) como indicador para detectar problemas de colinealidad entre variables predictoras (Peduzzi *et al.*, 2012). Sin embargo, el hecho de usar sólo dos métricas (correspondientes al plano vertical y horizontal) no significó gran problema de colinealidad entre ellas.

Random Forest

Random Forest (RF) es una técnica avanzada de aprendizaje de máquinas utilizada para modelar asociaciones estadísticas complejas de datos de diferentes fuentes. Es un método de regresión (también de clasificación) basado en la agregación de un número grande de árboles de decisión. RF es un conjunto de árboles construidos a partir de un conjunto de datos de entrenamiento y validados internamente para generar una predicción de la variable de respuesta dados los predictores. Cada árbol es construido a partir de una muestra bootstrap extraída con reemplazo del conjunto de datos original y las predicciones de los árboles resultantes son producto del promedio de todos ellos, en el caso de la regresión (Cutler *et al.*, 2012).

El rendimiento de RF no se ve afectado por la colinealidad de las variables predictoras, ni por la falta de normalidad de la variable de respuesta (Wilkes *et al.*, 2015). Es de fácil aplicación y con capacidad de procesar eficientemente bases de datos grandes, lo que permite ser una opción en estudios regionales (Baccini *et al.*, 2012; Mascaró *et al.*, 2014).

Para reconocer las métricas LiDAR que aportan información significativa al modelo RF, es necesario medir la importancia de las variables predictoras identificando la influencia que tiene cada predictor sobre el Mean Square Error (MSE) del modelo. El algoritmo se describe ampliamente en Cutler *et al.* (2012). Sin embargo, en términos prácticos, si la métrica LiDAR está contribuyendo al modelo y se permuta (o cambia) aleatoriamente por otro de los predictores en el grupo de prueba, entonces el MSE del modelo RF aumentará debido a que está perdiendo la información que aportaba dicha variable (Wilkes *et al.*, 2015)

Las variables predictoras o métricas LiDAR empleadas en RF para estimar la biomasa aérea en la SMSP y la SMSC fueron las que registraron una $r \geq 0.5$. Sin embargo, se eliminaron variables conforme la precisión del modelo mejoraba y se identificaba la influencia de cada variable predictora en el MSE. Se utilizó el paquete “randomForest” del software R.

Los modelos de biomasa fueron desarrollados usando el enfoque basado en área (White *et al.*, 2013). La precisión de los diferentes modelos fue evaluada en términos de la raíz del error cuadrático medio (RMSE por sus siglas en inglés) y RMSE relativo (%), así como el coeficiente de determinación (R^2) entre las observaciones y predicciones.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \qquad RMSE\% = 100 * \frac{RMSE}{\bar{y}}$$

Donde y_i es la biomasa observada, \hat{y}_i es la biomasa estimada, \bar{y} es la biomasa promedio observada, y n es el número de observaciones.

Mapeo de la biomasa aérea en franjas con datos LiDAR.

Los mapas de biomasa aérea para los dos tipos de selva se obtuvieron a través de la aplicación de ambos enfoques de regresión: regresión lineal múltiple y Random Forest. Se emplearon los productos ráster correspondientes a las métricas elegidas por el modelo de regresión lineal múltiple y el modelo RF, a lo largo de las franjas de datos LiDAR con un tamaño de pixel de 20 m. Se utilizó el paquete “raster” del software R.

1.3 RESULTADOS

Estimación de biomasa mediante modelo de regresión lineal

En total fueron 24 métricas LiDAR que mejor correlacionaron con la biomasa aérea estimada en campo de la SMSP. Los valores de las correlaciones de Pearson fueron ≥ 0.5 y hasta 0.7. La altura media fue la métrica que tuvo la mayor correlación con la biomasa estimada para este tipo de vegetación. En el caso de la SMSC fueron 23 métricas con correlación similar, el percentil 90 fue la variable con mayor correlación con la biomasa aérea estimada en campo.

Los resultados de los modelos de regresión lineal múltiple y transformado para la SMSP y SMSC se presentan en el Cuadro 1.3. En ambos casos se utilizaron sólo dos métricas para predecir la biomasa aérea. La altura media y el porcentaje de los primeros retornos encima de 2 m para la SMSP. En el caso de la SMSC las métricas fueron el percentil 95 de la altura y el porcentaje de los primeros retornos encima de la altura media. Los valores más altos de R^2 se alcanzaron cuando se transformó a la biomasa aérea mediante el procedimiento Box Cox. Cuando se agregaron más variables explicativas a los modelos, se alcanzó una R^2 marginalmente más grande, sin embargo, esos modelos tendieron a sobre ajustar los datos observados, resultando en valores altos de FIV entre las métricas explicativas de LiDAR similares.

Cuadro 1.3. Modelos de regresión lineal múltiple que relacionan las métricas LiDAR con la biomasa aérea por tipo de vegetación.

Tipo de vegetación	Modelo	R ²	RMSE	RMSE %
SMSP	$B = -92.377 + \text{Altura media} * 15.09 + \% \text{ de los primeros retornos encima de } 2 \text{ m} * 0.728$	0.56	40.79	45.63
SMSP Box Cox	$B_t = -4.79733 + \text{Altura media} * 1.034 + \% \text{ de los primeros retornos encima de } 2 \text{ m} * 0.08617$	0.62	41.44	46.36
SMSC	$B = -110.435 + \text{Percentil } 95 \text{ de la altura} * 12.939 + \% \text{ de los primeros retornos encima de la altura media} * 0.573$	0.53	34.68	48.50
SMSC Box Cox	$B_t = -2.46 + \text{Percentil } 95 \text{ de la altura} * 0.567 + \% \text{ de los primeros retornos encima de la altura media} * 0.032$	0.62	36.50	51.05

*** B= Biomasa aérea, B_t=Biomasa aérea transformada.

Estimación de biomasa mediante Random Forest

Los resultados de los modelos de RF para la SMSP y SMSC se presentan en el Cuadro 1.4. El número de variables predictoras empleadas en el modelo de SMSP fueron 20 métricas LiDAR, correspondientes a 13 percentiles de altura (P05 - P99), tres estadísticas descriptivas de la altura y cuatro tasas de retorno. Para el modelo de la SMSC fueron tres percentiles de altura (P75, P90 y P95), cuatro estadísticas descriptivas de la altura y seis tasas de retorno.

Cuadro 1.4. Resultados de los modelos Random Forest para la estimación de biomasa aérea por tipo de vegetación.

	selva mediana subperennifolia	selva mediana subcaducifolia
% de varianza explicada	56.95	52.07
r	0.75	0.73
RMSE	40.73	35.10
# de variables predictoras	20	13

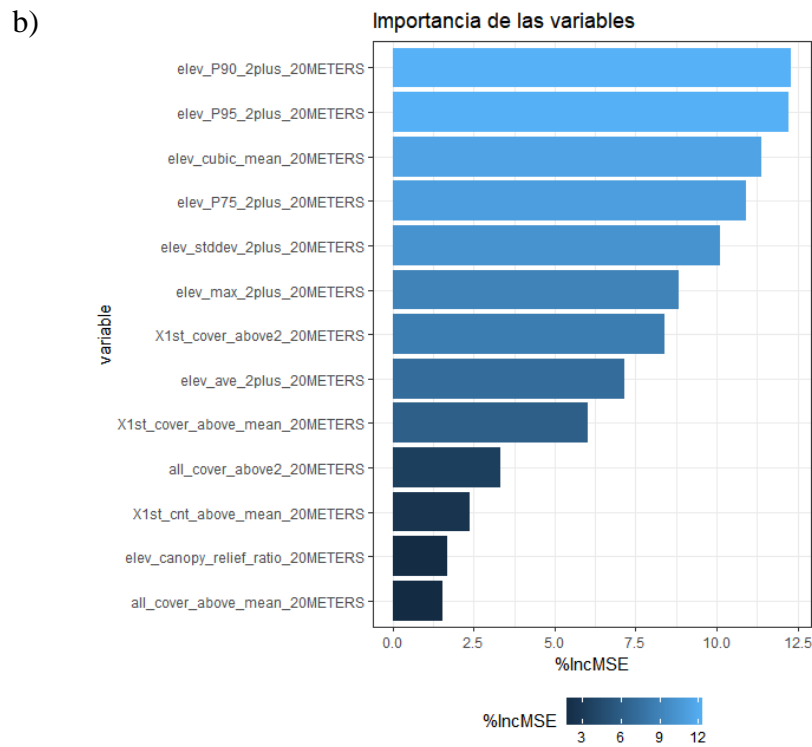
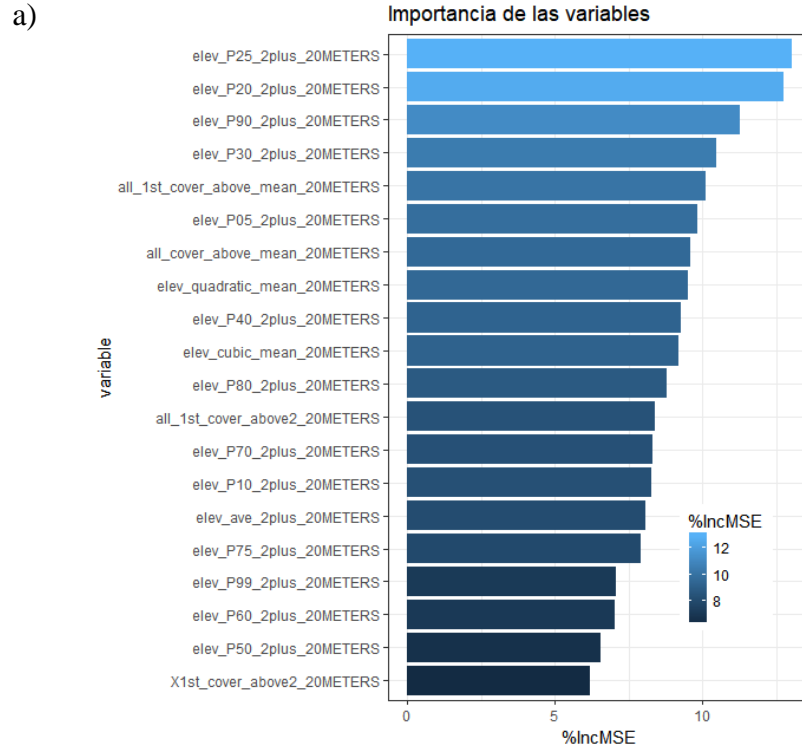


Figura 1.2 Importancia relativa de a) 20 variables seleccionadas para el modelo Random Forest en la selva mediana subperennifolia y b) de 13 variables en la selva mediana subcaducifolia. %IncMSE: Incremento porcentual del Mean Square Error (error cuadrático medio).

La Figura 1.3 muestra la relación entre la biomasa aérea predicha a partir de los modelos de regresión (a y b) y Random Forest (c y d) con la biomasa aérea de referencia calculada aplicando las ecuaciones alométricas a partir de las mediciones del INFyS. Los cuatro gráficos sugieren que los valores altos de biomasa aérea se subestimaron, mientras que los valores pequeños fueron sobreestimados.

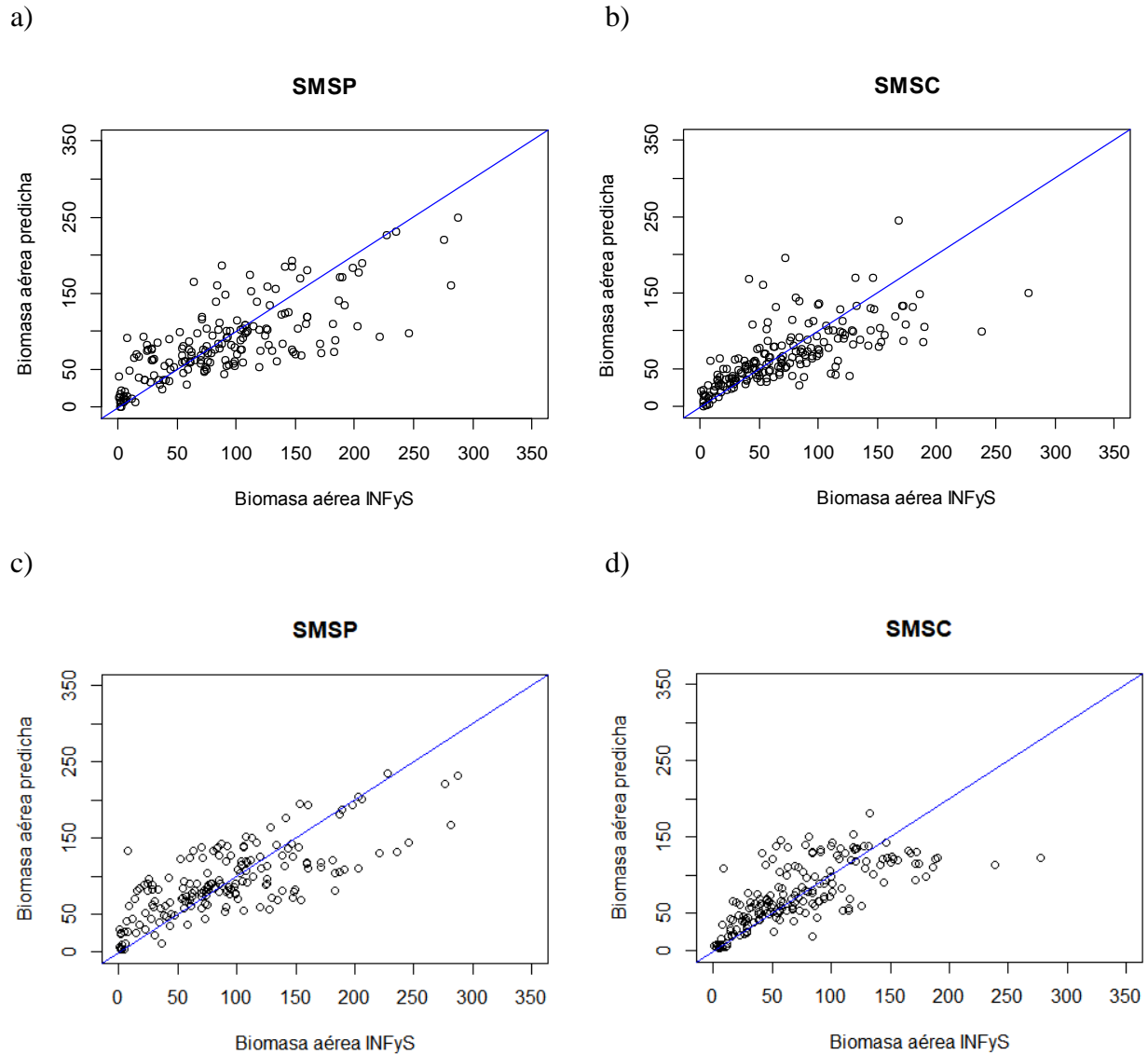


Figura 1.3 Predicción de la biomasa aérea usando los datos LiDAR en franjas: a) Biomasa aérea en la selva Mediana Subperennifolia y b) selva mediana subcaducifolia mediante regresión lineal y c) SMSP y d) SMSC mediante Random Forest. Se indica la línea 1:1.

Mapas de biomasa aérea mediante los dos enfoques

En la Figura 1.4 se presentan un acercamiento de los mapas de biomasa aérea en franjas generada mediante regresión lineal (a) y mediante RF (b) en la SMSC. c) un acercamiento de la biomasa aérea generada mediante regresión lineal y d) mediante RF para la SMSP.

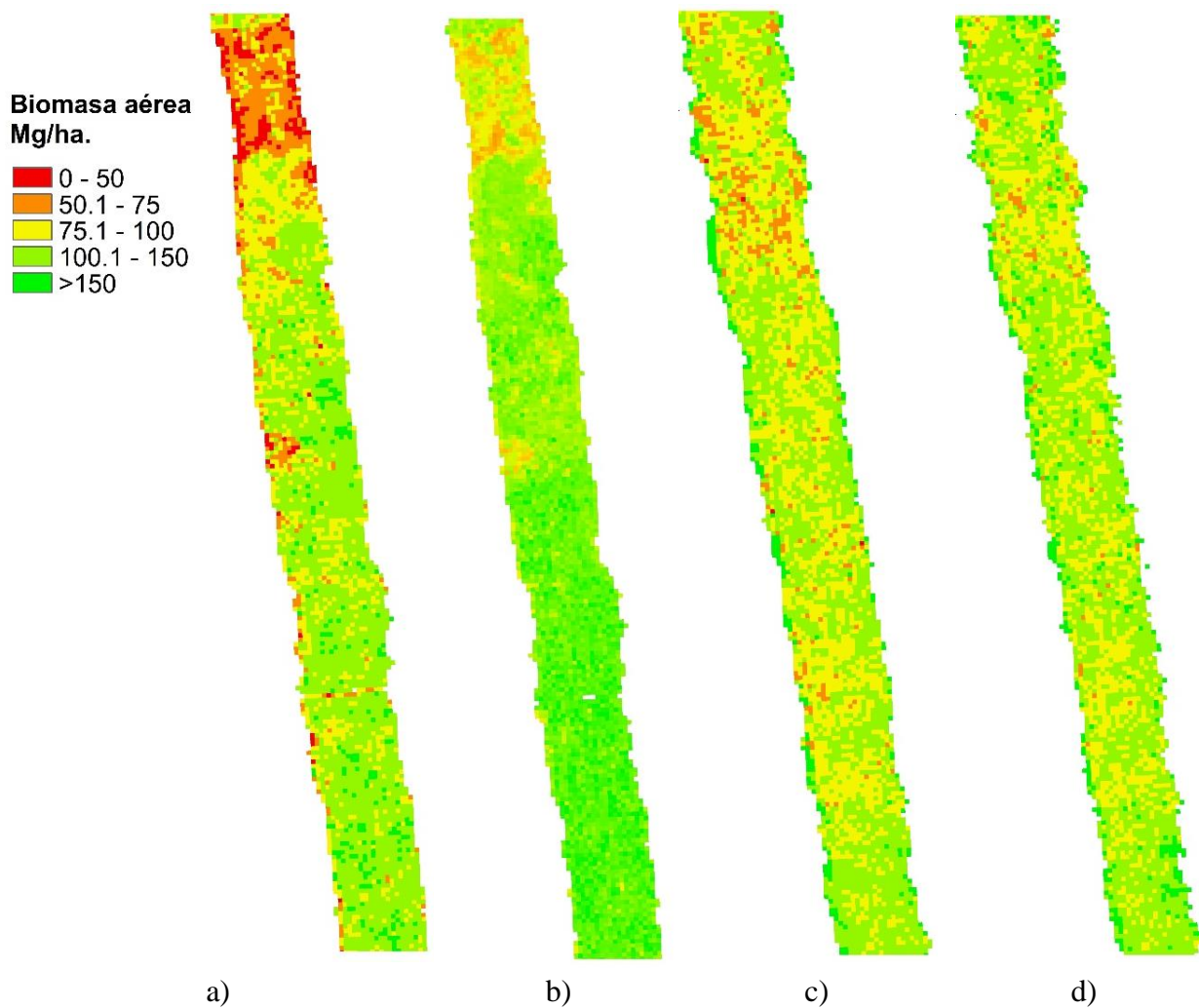


Figura 1.4 Mapas de biomasa aérea usando los datos LiDAR en franjas: a) biomasa aérea mediante regresión múltiple y b) mediante RF en la SMSC; c) biomasa aérea mediante regresión múltiple y d) mediante Random Forest en la SMSP.

1.4 DISCUSIÓN

En el presente estudio, bajo el enfoque de regresión lineal, las métricas de LiDAR que más se correlacionaron con la biomasa aérea de la SMSC fueron los percentiles de altura 80, 90 y 95. El P90 y P95 son lo más parecido a contar con la altura máxima del dosel (Goetz & Dubayah, 2011). Los resultados de este estudio son consistentes con otros de estimación de biomasa aérea en bosques tropicales. Cada una de estas tres métricas (P80, P90 y P95) se combinan con otras métricas de densidad o estadísticas de la distribución de altura para conformar los modelos generados por Clark *et al.* (2011) en Costa Rica, en manglares de Brasil (Rocha de Souza *et al.*, 2018), así como en bosques templados (Lefsky *et al.*, 2002).

En el caso de la SMSP, la altura media y dos percentiles bajos (P20 y P25) fueron los que correlacionaron mejor con la biomasa de este tipo de selva, lo cual es consistente con lo indicado por Véga *et al.* (2015). En el estudio de d'Oliveira *et al.* (2012), realizado con el propósito de estimar la biomasa aérea en un bosque tropical en Brasil, los autores señalaron que el P25 y la varianza de la altura fueron las mejores variables explicativas de los modelos.

El comportamiento de mayor correlación entre los percentiles P20 y P25 (4.79 – 5.32 m) con la biomasa aérea de la SMSP podría explicarse a que la nube de puntos LiDAR logró registrar adecuadamente los retornos correspondientes a este conjunto de árboles de alturas menores. Echeverría *et al.* (2014) y Zamora-Crescencio *et al.* (2017) estudiaron la composición y estructura de diferentes porciones de selva mediana subperennifolia en el estado de Campeche. En el primer estudio indicaron que 77.6% de los individuos tienen alturas menores a 5 m, mientras que la caracterización vertical del segundo estudio indicó que 42% de los individuos se concentraron en la primera clase (2 - 4.9 m). Sin embargo hay que considerar que el arbolado fue medido a partir de $DN > 1$ cm. En cuanto a la altura media (7.9 m) también fue registrada

apropiadamente por la nube de puntos LiDAR. El estudio de Read & Lawrence (2003) señala que las alturas promedio en la clase diamétrica de 5 - 9.9 cm se encuentran en un intervalo de 5 a 8 m. Para la siguiente clase (>10 cm) el intervalo es de más de 6 m y menos de 12 m.

Algo similar sucedió en la SMSC. Los percentiles P80 y P90 (9.6 – 10.6 m) que correlacionaron mejor con la biomasa estimada, corresponde al patrón de altura promedio para este tipo de vegetación. En el estudio de Dupuy *et al.* (2012) el patrón de altura (determinado principalmente por el arbolado adulto) se encuentra en un intervalo similar a dichos percentiles. Hay que tener en cuenta que en dicho estudio se muestrearon individuos con DN > 5 cm y los datos del INFyS empleados en este trabajo registran los valores del DN > 7.5 cm.

En este trabajo se estimó la biomasa aérea a nivel regional, empleando franjas que atraviesan largas porciones de la Península de Yucatán. Los sitios coincidentes con las franjas LiDAR son muy diferentes entre sí a pesar de pertenecer a un mismo tipo de vegetación, sin embargo, los datos LiDAR fueron capaces de registrar la altura promedio de las selvas evaluadas. Se debe tener presente que la estructura y composición de los tipos de selva estudiados son característicos de cada lugar; la estructura varía con respecto a su posición topográfica, la disponibilidad de agua en el suelo, la intervención antropogénica a la que ha sido sometida, la etapa de sucesión en la que se encuentra entre muchos más factores.

Por otro lado, en los dos tipos de selva estudiados el percentil 95 y la altura media, además de dos métricas relacionadas a la cobertura del dosel (el porcentaje de los primeros retornos encima de 2 m y el porcentaje de los primeros retornos encima de la altura media) contribuyeron significativamente a predecir la biomasa mediante modelos de regresión lineal. La idea principal de usar dos métricas en cada modelo fue crear una relación simple que involucre la distribución de altura y la profundidad en que los retornos LiDAR penetran el arbolado, como lo señala

Rocha de Souza *et al.* (2018), además es importante señalar que la adición de más variables en los modelos no incrementó significativamente la variación explicada.

Para el enfoque Random Forest, el percentil 90 (elev_P90_2plus_20METERS) fue una métrica común entre las variables más importantes seleccionadas por el modelo. La tercera más importante para la SMSP y la primera para la SMSC. En un estudio realizado por Li *et al.* (2014) este percentil, así como el percentil 80, fueron las variables más importantes en los modelos evaluados para estimar la biomasa de bosques templados mediante técnicas de aprendizaje de máquinas.

Los resultados obtenidos en este estudio mediante el enfoque de regresión usando variables transformadas funcionó mejor que RF en cuanto a la varianza explicada, registrando 62% para ambos tipos de vegetación (RMSE = 41.44 y 36.60 Mg ha⁻¹ para SMSP y SMSC respectivamente); mientras que con RF fue de 57% (40.73 Mg ha⁻¹) para la SMSP y de 52% (35.10 Mg ha⁻¹) para la SMSC. En general, los estudios realizados de estimación de biomasa a partir de datos LiDAR han señalado precisiones (RMSE) altamente variables que oscilan entre 20 y 200 Mg ha⁻¹ en escalas de parcelas de 30 m (900 m²) hasta 1 ha (Goetz & Dubayah, 2011).

No obstante, cuando se comparan los resultados de este trabajo con estudios similares de estimación de biomasa aérea en bosques tropicales existen diferencias notables. Por ejemplo, en un estudio realizado en la amazonia brasileña, d'Oliveira *et al.* (2012) logró explicar la varianza hasta un 72% (RMSE = 40.2 Mg ha⁻¹); mientras que Véga *et al.* (2015) estimaron la biomasa aérea en un bosque tropical húmedo perenne en Western Ghats, India, con una R²= 0.96 (RMSEvc= 28.83 Mg ha⁻¹). Por otro lado, Meyer *et al.* (2013), utilizando sitios de muestreo del mismo tamaño que este estudio (0.04 ha), en un bosque tropical húmedo, hallaron una R² de 0.19

(RMSE = 184.6 Mg ha⁻¹) y 0.28 (RMSE = 173.8 Mg ha⁻¹) para dos tipos de sensores LiDAR, respectivamente.

Hernández-Stefanoni *et al.* (2014) estimó la biomasa aérea en dos sitios ubicados en los mismos tipos de selva analizados en este trabajo empleando regresión múltiple. La varianza explicada para ambos sitios fue de 0.49 (RMSE_{VC}= 37.4 Mg ha⁻¹ para el estudio situado en la SMSC y 60.2 Mg ha⁻¹ para el situado en la SMSP) con un tamaño de 0.04 ha. El RMSE obtenido en este trabajo a través del enfoque de regresión (36.50 Mg ha⁻¹) fue similar al que obtuvieron Hernández *et al.* (2014) en la SMSC. Incluso la RMSE para la SMSP fue menor en éste estudio (41.44 Mg ha⁻¹ con una diferencia de 18.8 Mg ha⁻¹) a pesar de que la varianza explicada no logró un valor mayor.

Por otro lado Urbazaev *et al.* (2018) estimaron la biomasa en todo México mediante un modelo que explicó la varianza en un 68% (RMSE = 22.22 Mg ha⁻¹) empleando simulaciones Monte Carlo y el algoritmo de aprendizaje de máquinas denominado Cubist. La varianza obtenida en su trabajo es parecida a la que se obtuvo en este estudio, la diferencia más grande es de 16%. Probablemente la escala empleada de 1 ha (100 m tamaño del pixel) con respecto a los 400 m² (20 m) de este estudio es un factor importante que explica estas diferencias. Los errores en la biomasa estimada con datos tridimensionales disminuyen cuando se utilizan parcelas más grandes debido al promedio espacial de los errores (Goetz & Dubayah, 2011; Hernández-Stefanoni *et al.*, 2014).

Además del promedio espacial, existen factores adicionales como el efecto de borde, en donde las copas de los árboles fuera del sitio se extienden dentro de la parcela (White *et al.* 2013). La nube de puntos LiDAR los detecta, pero en el inventario estos árboles no se registran. Dicho

efecto de borde representa una discrepancia entre la cantidad de biomasa estimada y observada siendo más evidente en las parcelas pequeñas (Zolkos *et al.*, 2013).

Se ha mencionado reiteradamente las ventajas de utilizar datos LiDAR para realizar las estimaciones de los parámetros forestales de interés, sin embargo, en este trabajo no se logró que los modelos explicaran la varianza más allá del 65%, lo que hace suponer que existen otros motivos que agrega error al modelo.

Uno de estos factores puede ser la poca precisión con la que se geo localizaron los centros de los sitios del inventario, 5.9 m según el INFyS (Conafor, 2015), lo que resulta en una falta de coincidencia espacial con la nube de puntos LiDAR. De esta situación depende que árboles o partes de árboles se encuentran dentro del sitio y cuáles de ellos se miden o no (Mascaro *et al.*, 2011; Gonçalves *et al.*, 2017).

En este sentido, estudios de propósito similar (Deo *et al.*, 2017; Rocha de Souza *et al.*, 2018) registran errores de geo posicionamiento de los sitios de muestreo menores a 1 m. No obstante White *et al.* (2013) recomiendan errores de geo posicionamiento de hasta 5 m en los sitios de muestreo en campo, además de otros métodos para disminuir el error, como registrar con el GPS un mínimo de 500 puntos por parcela; aplicar corrección diferencial o incrementar el tamaño de parcela para asegurar un traslape suficiente.

Por otro lado, el uso de ecuaciones globales, que a pesar de ser desarrolladas en la zona puede ser otra fuente de error. Aunque en este estudio se utilizaron ecuaciones por especie, una gran cantidad de especies en los bosques tropicales no cuentan con un modelo específico. No obstante, se esperaría un buen resultado con el uso de las ecuaciones generales. Cuando se usan dichos modelos alométricos, la capacidad de predecir la biomasa aérea a partir de datos LiDAR se rige por la intensidad de la relación diámetro normal - altura a nivel de individuo (Asner,

2009; Chen *et al.*, 2015). Sin embargo, en todo el proceso de estimación de biomasa existen más fuentes de error en el modelamiento de biomasa aérea: el tamaño de la muestra, el error en las mediciones en campo, la incertidumbre de muestreo relacionado al tamaño de la parcela, así como la representatividad de una red de parcelas en todo el paisaje que sin duda, están influyendo en la precisión de las estimaciones presentadas.

Otro factor importante que podría explicar los valores bajos de R^2 obtenidos en los modelos es el desfase temporal entre la toma de datos del inventario y la recolección de datos LiDAR (Mascaro *et al.*, 2011); durante ese tiempo ha ocurrido una serie de eventos que modificaron las condiciones de la vegetación, ya sea por mortalidad, deforestación, perturbación, etc. (Goetz & Dubayah, 2011). La mayoría de sitios se remidieron en 2009 y 2010 y el registro de datos LiDAR se llevó a cabo durante el año 2013. Los errores que se agregan en este punto podrían minimizarse al realizar la colecta de datos de campo y LiDAR simultáneamente (Andersen *et al.*, 2011; Woods *et al.*, 2011).

Es importante resaltar una constante en los modelos evaluados, los valores altos de biomasa aérea fueron subestimados mientras que los valores bajos fueron sobreestimados. Con Random Forest el proceso de promediado crea resultados sesgados hacia la media de la muestra, entonces los valores pequeños o grandes de biomasa aérea comúnmente son sub o sobre estimados (Xu *et al.*, 2016). En el caso de regresión múltiple probablemente es consecuencia de no contar con suficiente información en los extremos. El valor de referencia fue calculado solo considerando los árboles medidos dentro del sitio, eliminado árboles pequeños y vegetación menor, sin embargo, los datos LiDAR consideran toda la información registrada dentro de la parcela.

A pesar de la limitada capacidad explicativa de los modelos generados, los resultados de este trabajo indican que la información obtenida es de alta calidad y es adecuada para mejorar las

estimaciones de biomasa aérea a nivel regional. Los errores registrados no exceden los 50 Mg ha⁻¹, aunque se esperaría que los errores obtenidos en este tipo de estudios y utilizando datos de percepción remota fueran próximos a 20 Mg ha⁻¹ o estar dentro del intervalo del 20% de las estimaciones obtenidas en campo (Zolkos *et al.*, 2013). No obstante se espera una mejora al conjuntar esta información con datos auxiliares como los que proporcionan las imágenes satelitales (Lu *et al.*, 2012) u otros sensores activos (Rodríguez-Veiga *et al.*, 2016).

El proceso de obtener información extensa y de alta calidad de la estructura del bosque es muy difícil y costosa, principalmente en los ecosistemas tropicales, dados sus características y factores estacionales adversos como inundaciones o huracanes durante la temporada de lluvias. Dichos factores pueden comprometer las mediciones recolectadas o la cantidad de sitios muestreados.

Aunque el problema tampoco se resolverá inmediatamente con los datos de LiDAR, debido a su limitada cobertura espacial, es necesario hacer uso de pequeñas muestras de datos que incrementan la precisión de la variable de interés evaluada, y a través de la conexión con los datos ópticos de satélite es posible realizar estimaciones en superficies mayores sin tener que invertir el tiempo y el dinero que implica realizar el inventario forestal sobre esa misma gran superficie (Chi *et al.*, 2017).

1.5 CONCLUSIONES

Este trabajo demostró que es posible realizar una estimación adecuada de biomasa aérea en bosques tropicales a través de una muestra de datos LiDAR y datos de inventario, a pesar de la complejidad y diversidad de este tipo de ecosistemas. Las métricas de altura y cobertura de LiDAR son las variables que correlacionaron mejor y contribuyeron en gran medida a estimar la biomasa aérea mediante los modelos empleados. Utilizando dos enfoques de regresión: lineal

múltiple y Random Forest fue posible estimar la biomasa aérea con resultados similares. El enfoque de regresión múltiple (transformado) logró explicar mejor la varianza que el método de Random Forest, no obstante los valores de RMSE en ambos casos fueron los más altos (41.44 y 36.50 Mg ha⁻¹ para la SMSP y SMSC respectivamente). A partir de los modelos generados, se obtuvieron los mapas de biomasa aérea sobre las franjas de datos LiDAR, lo que permitió conocer su distribución espacial de forma continua y homogénea.

Este estudio pone en evidencia la importancia de corregir algunas deficiencias de diseño en la fase de campo que podrían mejorar la precisión de los modelos ajustados para predecir la biomasa aérea: el desfase entre la toma de datos en campo y el registro de datos LiDAR, la precisión en la geolocalización de las parcelas en campo, el uso de ecuaciones locales o adecuadas para estimar biomasa aérea y la necesidad de que la muestra incluya la totalidad de las diferentes condiciones de densidad existentes en el área de interés.

1.6 LITERATURA CITADA

- Andersen, H.-E., Strunk, J., & Temesgen, H. (2011). Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the Upper Tanana Valley of Interior Alaska. *Western Journal of Applied Forestry*, 26(4), 157-164. doi: 10.1093/wjaf/26.4.157
- Asner, G. P. (2009). Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters*, 4(3). doi: 10.1088/1748-9326/4/3/034009
- Asner, G. P., Mascaro, J., Muller-Landau, H. C., Vieilledent, G., Vaudry, R., Rasamoelina, M., & van Breugel, M. (2012). A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia*, 168(4), 1147-1160. doi: 10.1007/s00442-011-2165-z
- Baccini, Goetz, S., Walker, W., Laporte, N., Sun, M., Sulla-Menashe, D., & Friedl, M. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, 2(3), 182. doi: 10.1038/nclimate1354
- Baccini, Laporte, N., Goetz, S., Sun, M., & Dong, H. (2008). A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environmental Research Letters*, 3(4), 045011. doi: 10.1088/1748-9326/3/4/045011

- Baltsavias, E. P. (1999). Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2), 199-214. doi: 10.1016/S0924-2716(99)00015-5
- Barbosa, J. M., Broadbent, E. N., & Bitencourt, M. D. (2014). Remote Sensing of Aboveground Biomass in Tropical Secondary Forests: A Review. *International Journal of Forestry Research*, 2014, 1-14. doi: 10.1155/2014/715796
- Cairns, M. A., Olmsted, I., Granados, J., & Argaez, J. (2003). Composition and aboveground tree biomass of a dry semi-evergreen forest on Mexico's Yucatan Peninsula. *Forest Ecology and Management*, 186(1-3), 125-132. doi: 10.1016/s0378-1127(03)00229-9
- Cao, L., Coops, N. C., Innes, J., Dai, J., & She, G. (2014). Mapping above-and below-ground biomass components in subtropical forests using small-footprint LiDAR. *Forests*, 5(6), 1356-1373. doi: 10.3390/f5061356
- Cartus, O., Kellndorfer, J., Walker, W., Franco, C., Bishop, J., Santos, L., & Fuentes, J. (2014). A National, Detailed Map of Forest Aboveground Carbon Stocks in Mexico. *Remote Sensing*, 6(6), 5559-5588. doi: 10.3390/rs6065559
- Chave, Condit, R., Aguilar, S., Hernandez, A., Lao, S., & Perez, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1443), 409-420. doi: 10.1098/rstb.2003.1425
- Chave, Condit, R., Lao, S., Caspersen, J. P., Foster, R. B., & Hubbell, S. P. (2003). Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. *Journal of Ecology*, 91(2), 240-252. doi: 10.1046/j.1365-2745.2003.00757.x
- Chen, Q., Lu, D., Keller, M., dos-Santos, M., Bolfe, E., Feng, Y., & Wang, C. (2015). Modeling and Mapping Agroforestry Aboveground Biomass in the Brazilian Amazon Using Airborne Lidar Data. *Remote Sensing*, 8(12). doi: 10.3390/rs8010021
- Chi, H., Sun, G., Huang, J., Li, R., Ren, X., Ni, W., & Fu, A. (2017). Estimation of Forest Aboveground Biomass in Changbai Mountain Region Using ICESat/GLAS and Landsat/TM Data. *Remote Sensing*, 9(12). doi: 10.3390/rs9070707
- Clark, M. L., Roberts, D. A., Ewel, J. J., & Clark, D. B. (2011). Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sensing of Environment*, 115(11), 2931-2942. doi: 10.1016/j.rse.2010.08.029
- Comisión Nacional Forestal [Conafor]. (2011). Manual y procedimientos para el muestreo de campo. Remuestreo 2012. Inventario Nacional Forestal y de Suelos. Zapopan, Jalisco, México: Conafor. Secretaría del Medio Ambiente y Recursos Naturales.
- Comisión Nacional Forestal [Conafor]. (2015). Base de datos del Inventario Nacional Forestal y de Suelos Remedición 2009-2014.

- Cook, B., Corp, L., Nelson, R., Middleton, E., Morton, D., McCorkel, J., & Montesano, P. (2013). NASA Goddard's LiDAR, Hyperspectral and Thermal (G-LiHT) Airborne Imager. *Remote Sensing*, 5(12), 4045-4066. doi: 10.3390/rs5084045
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. En C. Zhang, & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (pp. 157-175). Nueva York, Estados Unidos: Springer.
- d'Oliveira, M. V., Reutebuch, S. E., McGaughey, R. J., & Andersen, H.-E. (2012). Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. *Remote Sensing of Environment*, 124, 479-491. doi: 10.1016/j.rse.2012.05.014
- Deo, R., Russell, M., Domke, G., Andersen, H.-E., Cohen, W., & Woodall, C. (2017). Evaluating Site-Specific and Generic Spatial Models of Aboveground Forest Biomass Based on Landsat Time-Series and LiDAR Strip Samples in the Eastern USA. *Remote Sensing*, 9(12). doi: 10.3390/rs9060598
- Dupuy, J. M., Hernández- Stefanoni, J. L., Hernández- Juárez, R. A., Tetetla- Rangel, E., López- Martínez, J. O., Leyequién- Abarca, E., & May- Pat, F. (2012). Patterns and correlates of tropical dry forest structure and composition in a highly replicated chronosequence in Yucatan, Mexico. *Biotropica*, 44(2), 151-162.
- Echeverría, F., Arreola, J., Esparza, L. G., Morales, V., y López, J. L. (2014). Análisis de la composición y estructura de la selva mediana subperennifolia del Ejido Pachuitz, Hopelchén, Campeche, México. *Foresta Veracruzana*, 16(2).
- Foody, G. M., Boyd, D. S., & Cutler, M. E. J. (2003). Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment*, 85(4), 463-474. doi: 10.1016/s0034-4257(03)00039-7
- Goetz, S. & Dubayah, R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Management*, 2(3), 231-244. doi: 10.4155/cmt.11.18
- Gonçalves, F., Treuhaft, R., Law, B., Almeida, A., Walker, W., Baccini, A., & Graça, P. (2017). Estimating Aboveground Biomass in Tropical Forests: Field Methods and Error Analysis for the Calibration of Remote Sensing Observations. *Remote Sensing*, 9(12). doi: 10.3390/rs9010047
- Hansen, A. J., Phillips, L. B., Dubayah, R., Goetz, S., & Hofton, M. (2014). Regional-scale application of lidar: Variation in forest canopy structure across the southeastern US. *Forest Ecology and Management*, 329, 214-226. doi: 10.1016/j.foreco.2014.06.009
- Hernández-Stefanoni, J., Dupuy, J., Johnson, K., Birdsey, R., Tun-Dzul, F., Peduzzi, A., y López-Merlín, D. (2014). Improving Species Diversity and Biomass Estimates of Tropical Dry Forests Using Airborne LiDAR. *Remote Sensing*, 6(12), 4741-4763. doi: 10.3390/rs6064741

- Houghton, R. A., Hall, F., & Goetz, S. J. (2009). Importance of biomass in the global carbon cycle. *Journal of Geophysical Research: Biogeosciences*, 114(G2), doi: 10.1029/2009jg000935
- Instituto Nacional de Estadística y Geografía [INEGI]. (2013). *Conjunto Nacional de Uso del Suelo y Vegetación a escala 1:250,000*. Aguascalientes, México: INEGI.
- Laurin, G. V., Puletti, N., Chen, Q., Corona, P., Papale, D., & Valentini, R. (2016). Above ground biomass and tree species richness estimation with airborne lidar in tropical Ghana forests. *International Journal of Applied Earth Observation and Geoinformation*, 52, 371-379. doi: 10.1016/j.jag.2016.07.008
- Lefsky, M. A., Cohen, W. B., Harding, D. J., Parker, G. G., Acker, S. A., & Gower, S. T. (2002). Lidar remote sensing of above- ground biomass in three biomes. *Global ecology and biogeography*, 11(5), 393-399. doi: 10.1046/j.1466-822x.2002.00303.x
- Li, M., Im, J., Quackenbush, L. J., & Liu, T. (2014). Forest Biomass and Carbon Stock Quantification Using Airborne LiDAR Data: A Case Study Over Huntington Wildlife Forest in the Adirondack Park. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(7), 3143-3156. doi: 10.1109/jstars.2014.2304642
- Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., & Saah, D. (2012). Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *International Journal of Forestry Research*, 2012, 1-16. doi: 10.1155/2012/436537
- Mascaro, Asner, G. P., Knapp, D. E., Kennedy-Bowdoin, T., Martin, R. E., Anderson, C., & Chadwick, K. D. (2014). A tale of two "forests": random forest machine learning AIDS tropical forest carbon mapping. *PLoS One*, 9(1), 5993. doi: 10.1371/journal.pone.0085993
- Mascaro, Detto, M., Asner, G. P., & Muller-Landau, H. C. (2011). Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sensing of Environment*, 115(12), 3770-3774. doi: 10.1016/j.rse.2011.07.019
- McGaughey, R. J. (2016). *FUSION/LDV: software for LiDAR data analysis and visualization*. US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Seattle, WA, USA.
- Meyer, V., Saatchi, S. S., Chave, J., Dalling, J. W., Bohlman, S., Fricker, G. A., & Hubbell, S. (2013). Detecting tropical forest biomass dynamics from repeated airborne lidar measurements. *Biogeosciences*, 10(8), 5421-5438. doi: 10.5194/bg-10-5421-2013
- Næsset, E. & Økland, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, 79(1), 105-115. doi: 10.1016/S0034-4257(01)00243-7

- Nelson, R., Margolis, H., Montesano, P., Sun, G., Cook, B., Corp, L., & Prisley, S. (2017). Lidar-based estimates of aboveground biomass in the continental US and Mexico using ground, airborne, and satellite observations. *Remote Sensing of Environment*, 188, 127-140. doi: 10.1016/j.rse.2016.10.038
- Ningthoujam, R. K., Joshi, P. K., & Roy, P. S. (2018). Retrieval of forest biomass for tropical deciduous mixed forest using ALOS PALSAR mosaic imagery and field plot data. *International Journal of Applied Earth Observation and Geoinformation*, 69, 206-216. doi: 10.1016/j.jag.2018.03.007
- Ortiz-Reyes, A. D., Valdez-Lazalde, J. R., los Santos-Posadas, D., Héctor, M., Ángeles-Pérez, G., Paz-Pellat, F., y Martínez-Trinidad, T. (2015). Inventario y cartografía de variables del bosque con datos derivados de LiDAR: comparación de métodos. *Madera y Bosques*, 21(3), 111-128. doi: 10.21829/myb.2015.213461
- Peduzzi, A., Wynne, R. H., Fox, T. R., Nelson, R. F., & Thomas, V. A. (2012). Estimating leaf area index in intensively managed pine plantations using airborne laser scanner data. *Forest Ecology and Management*, 270, 54-65. doi: 10.1016/j.foreco.2011.12.048
- Pennington, T. D. (2005). *Árboles tropicales de México: manual para la identificación de las principales especies*. México, D.F.: UNAM.
- Phua, M.-H., Johari, S. A., Wong, O. C., Ioki, K., Mahali, M., Nilus, R., & Hashim, M. (2017). Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests. *Forest Ecology and Management*, 406, 163-171. doi: 10.1016/j.foreco.2017.10.007
- Puc-Kauil, R., Ángeles-Pérez, G., Valdez-Lazalde, J. R., Reyes-Hernández, V. J., Dupuy-Rada, J. M., Schneider, L., García-Cuevas, X., Pérez-Rodríguez, P. (2019, aceptado para publicación). Species-specific biomass equations for small-size tree species in secondary tropical forests. *Tropical and Subtropical Agroecosystems*, 22(x): xx-xx.
- R Development Core Team. (2013). R: A language and environment for statistical computing. 2009-2017.
- Read, L. & Lawrence, D. (2003). Recovery of biomass following shifting cultivation in dry tropical forests of the Yucatan. *Ecological Applications*, 13(1), 85-97.
- Rocha de Souza, P. F., Kampel, M., Gomes Soares, M., Estrada, G., Bentz, C., & Vincent, G. (2018). Reducing Uncertainty in Mapping of Mangrove Aboveground Biomass Using Airborne Discrete Return Lidar Data. *Remote Sensing*, 10(4). doi: 10.3390/rs10040637
- Rodríguez-Veiga, P., Saatchi, S., Tansey, K., & Balzter, H. (2016). Magnitude, spatial distribution and uncertainty of forest biomass stocks in Mexico. *Remote Sensing of Environment*, 183, 265-281. doi: 10.1016/j.rse.2016.06.004
- Saarela, S., Grafström, A., Ståhl, G., Kangas, A., Holopainen, M., Tuominen, S., & Hyyppä, J. (2015). Model-assisted estimation of growing stock volume using different combinations

- of LiDAR and Landsat data as auxiliary information. *Remote Sensing of Environment*, 158, 431-440. doi: 10.1016/j.rse.2014.11.020
- Saatchi, S. S., Houghton, R., Dos Santos Alvala, R., Soares, J. V., & Yu, Y. (2007). Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology*, 13(4), 816-837. doi: 10.1111/j.1365-2486.2007.01323.x
- Urbazaev, M., Thiel, C., Cremer, F., Dubayah, R., Migliavacca, M., Reichstein, M., & Schmillius, C. (2018). Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. *Carbon Balance and Management*, 13(1), 5. doi: 10.1186/s13021-018-0093-5
- Urquiza-Haas, T., Dolman, P. M., & Peres, C. A. (2007). Regional scale variation in forest structure and biomass in the Yucatan Peninsula, Mexico: Effects of forest disturbance. *Forest Ecology and Management*, 247(1-3), 80-90. doi: 10.1016/j.foreco.2007.04.015
- Véga, C., Vepakomma, U., Morel, J., Bader, J.-L., Rajashekar, G., Jha, C., & Dadhwal, V. (2015). Aboveground-Biomass Estimation of a Complex Tropical Forest in India Using Lidar. *Remote Sensing*, 7(12), 10607-10625. doi: 10.3390/rs70810607
- Wang, C. & Qi, J. (2008). Biophysical estimation in tropical forests using JERS- 1 SAR and VNIR imagery. II. Aboveground woody biomass. *International Journal of Remote Sensing*, 29(23), 6827-6849. doi: 10.1080/01431160802270123
- White, J. C., Wulder, M., Vastaranta, M., Coops, N., Pitt, D., & Woods, M. (2013). The Utility of Image-Based Point Clouds for Forest Inventory: A Comparison with Airborne Laser Scanning. *Forests*, 4(3), 518-536. doi: 10.3390/f4030518
- Wilkes, P., Jones, S. D., Suarez, L., Mellor, A., Woodgate, W., Soto-Berelov, M., & Skidmore, A. K. (2015). Mapping forest canopy height across large areas by upscaling ALS estimates with freely available satellite data. *Remote Sensing*, 7(9), 12563-12587. doi: 10.3390/rs70912563
- Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D., & Treitz, P. (2011). Operational implementation of a LiDAR inventory in Boreal Ontario. *The Forestry Chronicle*, 87(4), 512-528.
- Xu, L., Saatchi, S. S., Yang, Y., Yu, Y., & White, L. (2016). Performance of non-parametric algorithms for spatial mapping of tropical forest structure. *Carbon Balance Manag*, 11(1), 18. doi: 10.1186/s13021-016-0062-9
- Zamora-Crescencio, P., Rico-Gray, V., Barrientos-Medina, R. C., Puc-Garrido, E. C., Villegas, P., Domínguez-Carrasco, M. d. R., y Gutiérrez-Báez, C. (2017). Estructura y composición florística de la selva mediana subperennifolia en Bethania, Campeche, México. *Polibotánica*, 43, 67-86.
- Zanne, A., Lopez-Gonzalez, G., Coomes, D., Ilic, J., Jansen, S., Lewis, S., & Chave, J. (2013). *Global wood density database 2009*.

Zolkos, S. G., Goetz, S. J., & Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment*, 128, 289-298. doi: 10.1016/j.rse.2012.10.017

CAPÍTULO II. SINERGIA DE DATOS LANDSAT, CLIMÁTICOS Y LiDAR PARA ESTIMAR Y MAPEAR BIOMASA AÉREA EN SELVAS MEDIANAS DE LA PENÍNSULA DE YUCATÁN, MÉXICO

RESUMEN

Los bosques tropicales constituyen ecosistemas complejos y dinámicos que cubren áreas grandes y donde el acceso para obtener información de la vegetación se vuelve difícil y costoso. Además, pueden representar una fuente de dióxido de carbono terrestre hacia la atmósfera por las altas tasas de deforestación que ocurren en ellos. Por lo tanto, es importante determinar su contenido de biomasa, que corresponde a casi la mitad de carbono en bosques tropicales, haciendo uso de los beneficios que ofrece la percepción remota. En este trabajo se usó “Random Forest”, un algoritmo de aprendizaje de máquinas basado en árboles de decisión, para definir la relación existente entre variables explicatorias obtenidas de imágenes satelitales Landsat 8 y variables climáticas contra la variable de respuesta denominada biomasa aérea, estimada previamente a partir de franjas de datos LiDAR (Light Detection And Ranging) y muestreo en campo en la península de Yucatán, México. Una cierta combinación de variables climáticas, bandas espectrales e índices de vegetación mostró capacidad aceptable para estimar la biomasa en la selva mediana subperennifolia (SMSP) y mediana subcaducifolia (SMSC) con una varianza explicada de 50% y un valor de RMSE de 34.2 Mg ha⁻¹ y 26.2 Mg ha⁻¹ respectivamente. En general, las variables climáticas fueron las más importantes para predecir la biomasa aérea en los dos tipos de selva mediana analizados, lo que pone en evidencia la influencia de los efectos climáticos sobre la biomasa aérea a nivel regional.

Palabras clave: bosque tropical, imágenes de satélite, índices de vegetación, Random Forest

ABSTRACT

Tropical forests represent complex and dynamic ecosystems that cover large areas, thus obtaining vegetation from them information becomes difficult and expensive. In addition, they can represent a source of terrestrial carbon dioxide into the atmosphere due to the high rates of deforestation that occur in them. Therefore, it is important to determine its biomass, which almost half is carbon in tropical forest, making use of the benefits offered by remote sensing. In this, "Random Forest", a machine learning algorithm based on decision trees, was used to define the relationship between explanatory variables obtained from Landsat 8 satellite images and climatic variables with aboveground biomass, previously estimated from LiDAR data strips and field sampling at Yucatan Peninsula, Mexico. A certain combination of climatic variables, spectral bands and vegetation index showed acceptable capacity to estimate aboveground biomass in semi-evergreen tropical forest (SETF) and semi-deciduous tropical forest (SDTF) with an explained variance of 50% and RMSE value 34.2 Mg ha⁻¹ y 26.2 Mg ha⁻¹, respectively. In general, climatic variables were the most important for predicting aboveground biomass in the two tropical forest types analyzed, which highlights the influence of climate effects on aboveground biomass at regional level.

Key words: tropical forest, satellite images, vegetation index, Random Forest

2.1 INTRODUCCIÓN

Existe un gran interés en conseguir estimaciones confiables de biomasa aérea en los bosques tropicales, debido a su naturaleza compleja y dinámica (Barbosa *et al.*, 2014), su condición como principal fuente de emisión de carbono terrestre a la atmósfera (Houghton *et al.*, 2015) y su amplia extensión, que en muchos casos es de difícil acceso. Anteriormente, dichas estimaciones se lograban con un gran número de sitios medidos en campo para lograr presiones aceptables (Asner & Mascaro, 2014), sin embargo, en los últimos años la mayoría de estudios relacionados a estimaciones de biomasa hacen uso de la percepción remota para determinar relaciones entre la densidad de biomasa en campo y los rasgos que registran los sensores remotos (Saatchi *et al.*, 2011).

Los sistemas de percepción remota actualmente disponibles ofrecen diversas capacidades que operan desde distintas plataformas (White *et al.*, 2016). Son capaces de adquirir datos de manera sinóptica, sistemática y repetitiva, además de estar georreferenciados y ofrecer información a través de diferentes escalas espaciales (Boyd & Danson, 2005) que, en muchos casos, no implica altos costos (Foody *et al.*, 2001). Son una herramienta esencial para cuantificar las propiedades y la estructura de los bosques, particularmente donde la información es escasa y difícil de adquirir, como en los bosques tropicales (Lu *et al.*, 2004; Palace *et al.*, 2015).

En este sentido, las imágenes satelitales (sensores pasivos) de la plataforma Landsat se han convertido en una fuente cuantiosa de datos e información para aplicaciones de observación y monitoreo de la tierra. Ofrecen una cobertura global y continua que permite realizar diversos estudios como detección de cambios (Wulder *et al.*, 2012; Caughlin *et al.*, 2016), clasificación de cobertura del suelo o la estimación de parámetros del bosque (Feeley *et al.*, 2005; Hernández-Stefanoni *et al.*, 2014).

Las imágenes Landsat, por sí solas permiten mapear el tipo y condición de la vegetación existente en el paisaje. Las bandas NIR (near infrared - infrarrojo cercano) y SWIR (shortwave Infrared - infrarrojo de onda corta) son útiles para mapear las condiciones de humedad de las plantas y del suelo, así como la calidad del agua en humedales, ríos y ambientes costeros (Young *et al.*, 2017). Así mismo, se puede aprovechar la información combinada de los valores espectrales, cuando se crean los índices de vegetación, como indicadores biofísicos (Foody *et al.*, 2001), ya sea de la cantidad o el vigor de la vegetación en un pixel o incluso pueden ser asociados a la diversidad de especies arbóreas y la biomasa forestal (Freitas *et al.*, 2005).

No obstante, utilizar los índices espectrales de las imágenes satelitales representa un desafío cuando la densidad de la vegetación es alta, debido a que pueden perder sensibilidad a las propiedades biofísicas de la vegetación, lo que sucede a menudo en bosques tropicales (Foody *et al.*, 2001; Hernández-Stefanoni *et al.*, 2014). Sin embargo, existe una variedad importante de índices espectrales (NDVI normalized difference vegetation index - índice de vegetación de diferencia normalizada, SAVI soil-adjusted vegetation index- índice de vegetación ajustado al suelo, NDMI normalized difference moisture index - índice normalizado de la diferencia de humedad) que se han utilizado para el modelamiento de biomasa, la cuantificación de la estructura y la composición de bosques tropicales (Baccini *et al.*, 2004; Deo *et al.*, 2016; Ghosh & Behera, 2018).

Otro tipo de índices son obtenidos a través de la Transformación Tasseled Cap (TTC), mediante la cual se reducen las bandas de reflectancia a tres índices ortogonales: brillo, verdor y humedad (Pflugmacher *et al.*, 2012). Son capaces de distinguir cambios fenológicos y atributos clave del dosel, además de ser confiables y robustos en diferentes tipos y condiciones del bosque. (Healey *et al.* 2005; Deo *et al.*, 2016).

Por otra parte, el sensor activo LiDAR es considerado una de las tecnologías más precisas para estudiar la estructura del bosque, brindando el detalle y precisión espacial que caracteriza la composición del dosel (White *et al.*, 2016). Sin embargo, presenta restricciones en cuanto a su cobertura espacial y temporal al momento de monitorear grandes áreas, debido a su costo todavía alto, la logística en la obtención de datos, así como, la gran cantidad de datos que se almacenan a esta escala. Por lo que en muchos casos la única opción viable es contar con transectos o franjas establecidas estratégicamente (Wulder *et al.*, 2012; Saarela *et al.*, 2015; Wilkes *et al.*, 2015) y apoyarse en datos provenientes de otras fuentes, como las que proporcionan las imágenes satelitales, para estimar variables estructurales del bosque a una escala regional (Saatchi *et al.*, 2011). Los datos LiDAR proporcionan información sobre la estructura vertical del bosque, mientras que las imágenes satelitales lo hacen sobre la estructura horizontal del dosel (Mora *et al.*, 2013).

Los transectos de datos LiDAR pueden funcionar como una herramienta de muestreo, cubriendo una fracción pequeña del área de estudio (Ene *et al.*, 2016), si se establece una relación estadística (modelo) entre las métricas LiDAR y variables provenientes de imágenes satelitales u otras variables espacialmente sinópticas (precipitación, clima, elevación). El modelo construido puede posteriormente alimentarse con las variables que cubren completamente el área de estudio y así realizar un escalamiento de la variable de interés fuera de las franjas que cubrieron los datos LiDAR (Saatchi *et al.*, 2011; Wilkes *et al.*, 2015).

Para realizar estimaciones a nivel regional, hay que considerar que la densidad de biomasa aérea forestal varía en función del tipo de ecosistema. Esta variabilidad es producto de las limitaciones ambientales: los nutrientes del suelo, la distribución temporal de precipitación y temperatura así como de los procesos de disturbio y recuperación (Houghton *et al.*, 2009; Saatchi *et al.*, 2011).

En este sentido, White & Hood (2004) apuntan que los gradientes climáticos regionales influyen en las asociaciones vegetales sobre áreas extensas. Y particularmente en la Península de Yucatán los factores determinantes de dichas asociaciones son: la precipitación escasa y variable, el sustrato geológico y el suelo pobre, así como los disturbios naturales y antrópicos (Martínez & Leal, 2002).

En razón de lo expuesto, y aprovechando la información de diversas fuentes, se plantearon los objetivos de este trabajo, que fueron la identificación de variables espectrales y climáticas para realizar la estimación de biomasa aérea en dos tipos de selva mediana en la Península de Yucatán. Posteriormente se evaluó el rendimiento de los modelos ajustados con el algoritmo Random Forest (RF) para finalmente realizar el mapeo de biomasa en ambos tipos de vegetación del área de estudio. Tomando en cuenta que las variables climáticas influyen en la variabilidad de la vegetación a esta escala, se espera que sean las principales variables que determinen la estimación de biomasa en los dos tipos de selva mediana. Cabe aclarar que este estudio fue posible debido a que se emplearon estimaciones de biomasa aérea sobre franjas, derivadas de datos LiDAR, con el propósito de aumentar la muestra de campo original y con ello facilitar las estimaciones de biomasa nivel regional.

2.2 MATERIALES Y MÉTODOS

Área de estudio

Este estudio se llevó a cabo en dos tipos de selvas medianas: SMSP y SMSC de la Península de Yucatán, México localizadas en la porción sureste de la República Mexicana (17°00' y 21°45' N, 86°30' y 9°30' O). La clasificación por tipo de vegetación se tomó con base en la carta de uso de suelo y vegetación Serie V de INEGI (Fig. 2.1). La superficie aproximada de la península es de 141 523 km², equivalente al 7% del territorio nacional, de los cuales, 28 171 km²

corresponden a SMSC y 48 699 km² a la SMSP (Instituto Nacional de Estadística y Geografía [Inegi], 2013).

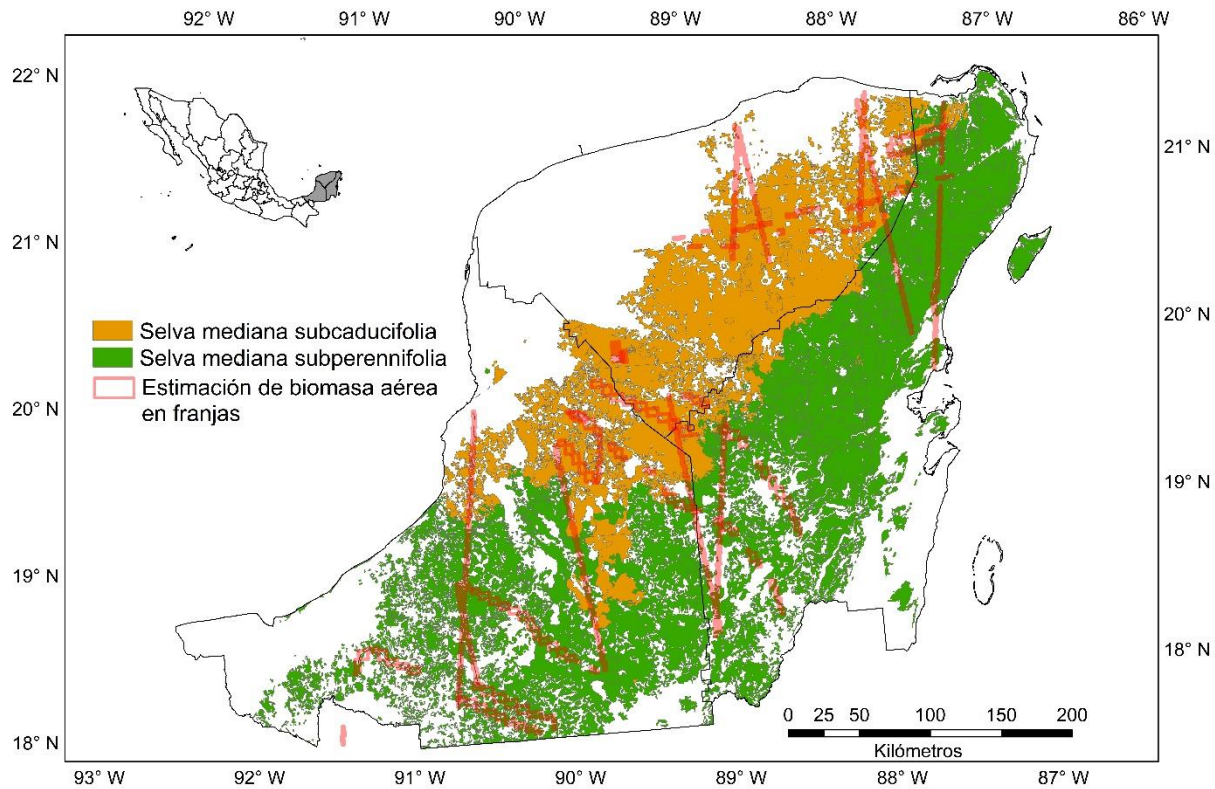


Figura 2.1 Área de estudio: selvas mediana subcaducifolia y mediana subperennifolia —Serie V, INEGI, 2013— de la Península de Yucatán, México. Las franjas rojas indican la ubicación de datos de biomasa estimados a partir de datos LiDAR (Ortiz-Reyes *et al.* 2019).

La selva mediana subcaducifolia se distribuye en la parte centro y norte de la península de Yucatán, en forma de franja (Pennington, 2005). Las especies importantes de este tipo de selva son: *Hymenaea courbaril*, *Hura polyandra*, *Brosimum alicastrum*, *Lysiloma bahamensis*, *Enterolobium cyclocarpum*, *Piscidia piscipula*, *Bursera simaruba*. El clima es cálido subhúmedo (Aw1) con lluvia en verano (mayo-octubre) y una estación seca (noviembre-abril). La temperatura media anual es de 26° C (Dupuy-Rada *et al.*, 2012). Con respecto a la selva mediana subperennifolia, los componentes arbóreos pierden estacionalmente su follaje en un 25 a 50% durante la época de sequía, debido a las condiciones marginales de humedad. Se establece en

regiones planas, con drenaje rápido y suelos desarrollados (Martínez & Galindo, 2002). Las especies importantes de este tipo de selva son: *Lysiloma latisiliquum*, *Brosimum alicastrum*, *Bursera simaruba*, *Manilkara zapota*, *Lysiloma spp.*, *Vitex gaumeri*, *Bucida buceras*, *Alseis yucatanensis*, *Psidium sartorianum*, *Carpodiptera floribunda* (Pennington, 2005). El clima es cálido subhúmedo con lluvias en verano, con una temperatura media anual de 26° C (INEGI, 2013). En la mayor parte de la península de Yucatán se presenta un gradiente de precipitación general de seco hacia el noroeste (600 mm) y más húmedo hacia el sureste (1400 mm) (Márdero *et al.*, 2012).

Datos de biomasa aérea derivados de transectos de datos LiDAR

Se calculó la biomasa aérea en las franjas LiDAR siguiendo el procedimiento de Ortiz-Reyes *et al.* (2019), que emplea un enfoque basado en área el cual consiste en vincular mediciones en campo (datos del Inventario Nacional Forestal y de Suelos de México - INFyS), con métricas LiDAR (generadas a partir del recorte de la nube de puntos coincidentes con los sitios del INFyS) a través del algoritmo Random Forest (Cutler *et al.*, 2012). El modelo construido permitió generar estimaciones de biomasa aérea a nivel pixel (20 m) Fig. 2.1.

Sin embargo, se consideró remover la información de ciertos sitios que excedían un rango de 2 veces la desviación estándar residual, con la finalidad de mejorar el modelo, tomando en cuenta el desfase temporal entre la toma de datos en campo y el registro de datos LiDAR que alcanzó un máximo de 4 años. Por lo que, de 176 sitios que se utilizaron para entrenar el modelo de biomasa aérea en la SMSP (ver capítulo I) se descartaron 8 sitios con residuales por arriba de dos desviaciones estándar. Esto permitió entrenar nuevamente los modelos sólo con 168 sitios para la SMSP y 173 sitios para la SMSC y obtener la distribución espacial de la biomasa aérea sobre las franjas eliminando en la medida de lo posible, la discrepancia entre la adquisición de los datos de

campo y el registro de datos LiDAR. Dicha discordancia temporal conlleva cambios estructurales importantes que pueden ser debido a disturbios tanto naturales como antrópicos (Urbazaev *et al.* 2018).

El propósito principal de Ortiz-Reyes *et al.* (2019), en su trabajo de estimación de biomasa aérea en franjas, fue subsanar en lo posible la desventaja principal de estimar la biomasa aérea a escala regional a partir de una muestra de campo limitada, en ese caso, los datos de los sitios del INFyS existentes para el área de interés. La biomasa aérea estimada a nivel pixel (20 m de resolución espacial) en las franjas censadas por un dispositivo LiDAR, generó una superficie muestral de alta calidad y para una fracción mayor del área de estudio. De acuerdo con White *et al.* (2013) este tipo de estimaciones son adecuadas ya que frecuentemente cumplen o superan los estándares de precisión operacional. Para este trabajo, cada pixel (20 m x 20 m) de la franja es una “parcela LiDAR”, similar a las parcelas en campo, que proporcionan información relevante que puede ser usada de forma independiente o integrada con otros recursos como las imágenes satelitales. Las “parcelas LiDAR” son un medio para aumentar la distribución y el tamaño de muestra de los datos locales e informar sobre la estructura del bosque (Wulder *et al.*, 2012).

Posteriormente, las franjas con la estimación de biomasa aérea fueron remuestreadas a un tamaño de pixel de 30 m, empleando el método del vecino más cercano, para hacerlos coincidentes con las dimensiones de pixel de las imágenes Landsat.

Datos de imágenes Landsat y climáticos

Se descargaron de la página <https://espa.cr.usgs.gov/ordering/new/> (Servicio Geológico de Estados Unidos —USGS) seis imágenes Landsat 8, producto Operational Land Imager (OLI), procesadas a nivel de reflectancia en superficie (RS). Los detalles sobre el algoritmo L8SR

(Landsat 8 Surface Reflectance) empleado para derivar dichos productos se encuentran en Vermote *et al.* (2016).

El área de estudio abarca escenas correspondientes a los path 19 y 20 y row 45, 46 y 47. Se estableció un año como periodo de búsqueda en las imágenes (1 de abril del 2013 hasta el 30 de abril del 2014). Lo anterior debido a que se estableció una cercanía entre las condiciones de la vegetación que capturaron los datos LiDAR (los datos se registraron entre abril y mayo de 2013) con las imágenes Landsat. Freitas *et al.* (2005) sugiere que se puede encontrar un relación más sólida cuando el periodo de adquisición de la imagen es similar al tiempo en que se toman las medidas del bosque. Sin embargo, debido a las condiciones climáticas presentes en la zona, que frecuentemente provoca nubes y sombras, así como, las características particulares de las selvas analizadas (subperennifolia y subcaducifolia), utilizar imágenes que coincidieran precisamente con la fecha de toma de datos LiDAR (abril – mayo) podría significar no encontrar diferencias entre los valores espectrales de las imágenes. La respuesta espectral en bosques deciduos y perennes es similar en abril y mayo, (cuando empiezan la temporada de lluvias) mientras que en los meses más secos cuando las hojas empiezan a caer, los bosques tropicales deciduos pueden alcanzar un alto grado de separabilidad en su respuesta espectral con respecto a otros tipos de vegetación, llegando a ser similar a la respuesta de suelo desnudo (Simonetti *et al.* 2014). Tomando en cuenta lo anterior y que las escenas tuvieran el menor porcentaje de nubosidad posible (menor al 20%), se descargaron aquellas imágenes correspondientes a la época de invierno (Cuadro 2.1).

Cuadro 2.1. Características de las escenas Landsat 8 procesadas a nivel de reflectancia en superficie.

Identificador de la escena Landsat	WRS Path	WRS Row	Cobertura de nubes de la escena (%)	Fecha de adquisición
LC80190452014046LGN01SR		45	2.08	15 feb 2014
LC80190462014046LGN01SR	19	46	1.97	15 feb 2014
LC80190472014046LGN01SR		47	10.79	15 feb 2014
LC80200452014005LGN01SR		45	0.54	05 ene 2014
LC80200462014021LGN01SR	20	46	8.3	21 ene 2014
LC80200472014005LGN01SR		47	20.9	05 ene 2014

Se aplicó una máscara de nubes a cada escena empleando el filtro Pixel QA (Quality Assessment), el cual solo está disponible para los productos L8SR (Vermote *et al.*, 2016). La aplicación de las máscaras se realizó en el software QGIS Versión: 3.6 Noosa (QGIS, 2019), empleando el complemento CloudMasking que hace el enmascaramiento de nubes, sombras, aerosoles, hielo/nieve y agua, utilizando diferentes procesos y filtros. Las áreas de nubes y sombras fueron excluidas de los análisis posteriores.

A continuación, se crearon los índices espectrales: NDVI, MSAVI (modified soil-adjusted vegetation index - índice de vegetación ajustado al suelo modificado), SAVI y EVI (enhanced vegetation index - índice de vegetación mejorado) con las imágenes pre-procesadas.

Foody *et al.* (2001) indica que los índices de vegetación a veces fallan porque no aprovechan todo los datos espectrales disponibles, usando solo una o dos bandas, dejando de lado la información disponible de las demás, es por eso que la transformación Tasseled Cap se consideró como una forma de aprovechar la información de más bandas. Para obtener los índices de brillo, verdor y humedad de TTC se utilizaron los coeficientes para productos Landsat con reflectancia en superficie que señala Crist (1985). Para este procedimiento se utilizó el paquete “raster” del software R (R Development Core Team, 2013). También se empleó la información de las bandas

espectrales por sí solas (2-7), como variables independientes en el modelo RF para estimar biomasa aérea.

Los datos climáticos se obtuvieron de la página <http://worldclim.org/>. Esta dispone de datos de temperatura mensual promedio, mínima y máxima, así como datos de precipitación para un periodo de 1970 a 2000. Se descargaron los datos de temperatura mensual promedio para enero, febrero, noviembre y diciembre, así como los datos de precipitación de los mismos meses y dos variables bioclimáticas (temperatura media anual -°C y precipitación anual -mm), derivadas de valores de temperatura y precipitación mensuales que permite generar variables biológicamente más significativas (Fick & Hijmans, 2017). Todas las variables se descargan a una resolución de ~1km², por lo que fueron remuestreadas a 30 m para hacerlo coincidente con las demás variables, usando el método del vecino más cercano. Se eligieron estas variables por su relevancia probada en otros estudios de estimación de parámetros forestales (Chen & Hay, 2011; Brososke *et al.*, 2014; Ahmed *et al.*, 2015; Wilkes *et al.*, 2015). La lista de variables predictoras procesadas se presenta en el Cuadro 2.2. Y en el anexo 1, se muestra la variación espacial de cada una de las variables.

Cuadro 2.2. Variables predictoras (espectrales y climáticas) empleadas en el modelamiento de la biomasa aérea.

Variable (abreviación)	Características/Fórmula	Propiedad
Banda 2 (B2fc)	B2 Azul (λ : 0.452 - 0.512 μm)	Puede diferenciar el suelo de la vegetación, y la vegetación decidua de coníferas (USGS, 2019)
Banda 3 (B3fc)	B3 Verde (λ : 0.533 - 0.590 μm)	Útil para evaluar el vigor de la planta (USGS, 2019)
Banda 4 (B4fc)	B4 Rojo (λ : 0.636 - 0.673 μm)	Discrimina las pendientes de vegetación (USGS, 2019)
Banda 5 (B5fc)	B5 Infrarrojo cercano (λ : 0.851 - 0.879 μm)	Enfatiza el contenido de biomasa y las costas (USGS, 2019)
Banda 6	B6 Infrarrojo de onda corta 1. (λ : 1.566	Relación negativa fuerte con

Variable (abreviación)	Características/Fórmula	Propiedad
(B6fc)	- 1.651 μm)	biomasa, discrimina el contenido de humedad del suelo y vegetación (Lu, 2006; USGS, 2019)
Banda 7 (B7fc)	B7 Infrarrojo de onda corta 2 (λ : 2.107 - 2.294 μm)	Mejora el contenido de humedad del suelo y la vegetación (USGS, 2019)
NDVI (bNDVIfc)	$\text{NDVI} = (\text{Banda } 5 - \text{Banda } 4) / (\text{Banda } 5 + \text{Banda } 4)$	Sensible a la actividad fotosintética (DeVries <i>et al.</i> , 2016)
MSAVI (bMSAVIfc)	$\text{MSAVI} = (2 * \text{Banda } 5 + 1 - \text{sqrt}((2 * \text{Banda } 5 + 1)^2 - 8 * (\text{Banda } 5 - \text{Banda } 4))) / 2$	Sensible a la cantidad de vegetación (Qi <i>et al.</i> , 1994)
SAVI (bSAVIfc)	$\text{SAVI} = ((\text{Banda } 5 - \text{Banda } 4) / (\text{Banda } 5 + \text{Banda } 4 + 0.5)) * (1.5)$	Altamente correlacionado con la dinámica de la cobertura vegetal (Gao <i>et al.</i> , 2000)
EVI (bEVIfc)	$\text{EVI} = 2.5 * ((\text{Banda } 5 - \text{Banda } 4) / (\text{Banda } 5 + 6 * \text{Banda } 4 - 7.5 * \text{Banda } 2 + 1))$.	Sensible a variaciones estructurales del dosel (Gao <i>et al.</i> , 2000; Vieilledent <i>et al.</i> , 2016)
TCB (brighVal)	$\text{TCB} = 0.2043\text{Banda } 2 + (0.4158)\text{Banda } 3 + (0.5524)\text{Banda } 4 + (0.5741)\text{Banda } 5 + (0.3124)\text{Banda } 6 + (0.2303)\text{Banda } 7$	Sensible al brillo de la superficie (Crist, 1985)
TCV (GreenVal)	$\text{TCV} = (-0.1603)\text{Banda } 2 + (-0.2819)\text{Banda } 3 + (-0.4934)\text{Banda } 4 + (0.7940)\text{Banda } 5 + (-0.0002)\text{Banda } 6 + (-0.1446)\text{Banda } 7$	Sensible al verdor de la vegetación (Crist, 1985)
TCH (WetVal)	$\text{TCH} = (0.0315)\text{Banda } 2 + (0.2021)\text{Banda } 3 + (0.3102)\text{Banda } 4 + (0.1594)\text{Banda } 5 + (-0.6806)\text{Banda } 6 + (-0.6109)\text{Banda } 7$	Sensible al contenido de humedad de la vegetación (Crist, 1985)
Temperatura media anual (Var_Bio1)	Datos desde 1970 – 2000 $^{\circ}\text{C}$ a 30" de resolución espacial (~1km ²)	Influye en los procesos de crecimiento y mortalidad de la vegetación (Kardol <i>et al.</i> , 2010; Alvarez-Davila <i>et al.</i> , 2017).
Precipitación anual (Var_Bio12)	Datos desde 1970 – 2000 mm a 30" de resolución espacial (~1km ²)	Relación positiva con la biomasa (Huxman <i>et al.</i> , 2004). Influye en los procesos de crecimiento y mortalidad de la vegetación (Alvarez-Davila <i>et al.</i> , 2017)

Variable (abreviación)	Características/Fórmula	Propiedad
Temperatura promedio de los meses de enero (TemAv_M01), febrero (TemAv_M02), noviembre (TemAv_M11) y diciembre (TemAv_M12) °C	°C a 30'' de resolución espacial (~1km ²)	Influye en la activación de los procesos de crecimiento en las plantas (Fick & Hijmans, 2017).
Precipitación promedio de los meses de enero (Prec_M01), febrero (Prec_M02), noviembre (Prec_M11) y diciembre (Prec_M12)	mm a 30'' de resolución espacial (~1km ²)	Influye en la activación de los procesos de crecimiento en las plantas (Fick & Hijmans, 2017).

Estimación de biomasa aérea mediante el algoritmo Random Forest.

Foody *et al.* (2001) señala que los datos adquiridos en diferentes longitudes de onda del espectro electromagnético están fuertemente correlacionados entre sí. Por lo tanto, se aprovechó los beneficios del algoritmo Random Forest para estimar la biomasa aérea en los dos tipos de selva evaluados. RF no es afectado por la colinealidad de las variables predictoras, ni por la falta de normalidad de la variable de respuesta (Wilkes *et al.*, 2015), es de fácil aplicación y con capacidad de procesar eficientemente bases de datos grandes, lo que permite ser una opción en estudios regionales (Mascaro *et al.*, 2014).

Sin embargo, hay que decir que RF puede verse afectado por la autocorrelación espacial, es decir, el algoritmo no está diseñado para tomar en cuenta las ubicaciones espaciales de las observaciones y esto puede llevar a predicciones sesgadas. Hay estudios como el de Mascaro et

al. (2014) que utilizan la información de coordenadas como covariables del modelo, no obstante, puede no mejorar del todo el proceso de predicción y entonces sería necesario hacer uso de enfoques nuevos que subsanen este aspecto importante del modelado mediante este tipo de algoritmos (Hengl *et al.*, 2018).

A pesar de las limitaciones que se pudieran presentar, en este trabajo se utilizó RF en su modo de regresión, empleando el paquete “randomForest” del software R (R Development Core Team, 2013). Donde la biomasa aérea estimada sobre las franjas fue la variable dependiente mientras que los datos climáticos y espectrales son las variables independientes. En este punto es importante aclarar que la biomasa estimada en franjas corresponde a un conjunto de píxeles (con resolución espacial de 30 m) que caen ya sea en la SMSP o en la SMSC. Debido a que el número total de píxeles (por tipo de vegetación) estuvo por encima de 250 000, se realizó una muestra aleatoria sin reemplazo de 0.02% (por tipo de vegetación) para que finalmente se contara con una muestra de 5000 píxeles. Utilizar más observaciones implica sobreajuste del modelo y mayor esfuerzo computacional.

La selección de las variables predictoras se realizó con base en su importancia dentro del modelo. Después de cada iteración se evaluó el rendimiento del modelo para cada tipo de vegetación analizado, con base en el RMSE (root mean square error – raíz del error cuadrático medio), el número de predictores empleados, así como el porcentaje de varianza explicado. A pesar que RF no es sensible al número de variables predictoras, eliminar variables no relevantes puede mejorar el rendimiento del modelo (McInerney & Nieuwenhuis, 2009).

Mapeo de biomasa aérea en los dos tipos de selva mediana

Los mapas de biomasa aérea para los dos tipos de selva se obtuvieron a través de la aplicación del modelo RF ajustado. Se emplearon los productos ráster correspondientes a las variables

espectrales y climáticas elegidas por el modelo. Para este procedimiento se utilizó el paquete “raster” del software R (R Development Core Team, 2013).

2.3 RESULTADOS

Estimación de biomasa aérea en dos tipos de selva mediana

Los modelos ajustados para cada tipo de selva tienen capacidad de predecir la biomasa aérea con una varianza explicada de 50% y valores de RMSE = 34.2 y 26.2 Mg ha⁻¹ para la SMSP y SMSC respectivamente (Cuadro 2.3).

Cuadro 2.3. Parámetros relevantes de los modelos Random Forest para la estimación de biomasa aérea por tipo de selva mediana.

	Selva Mediana Subperennifolia (5000 pixeles)	Selva Mediana Subcaducifolia (5000 pixeles)
% de varianza explicada	50	50
r	0.71	0.70
RMSE	34.2	26.2
# de variables predictoras	12	15

El modelo ajustado para la SMSC mostró un valor de RMSE substancialmente menor que el correspondiente a la SMSP, sin embargo, el número de variables fue mayor en el primer caso (15) contra las 12 variables empleadas para la SMSP.

A continuación, se presentan los gráficos que evalúan la fuerza e importancia de las variables dentro del modelo para predecir las biomásas aéreas en los dos tipos de selva evaluadas (Figura 2.2).

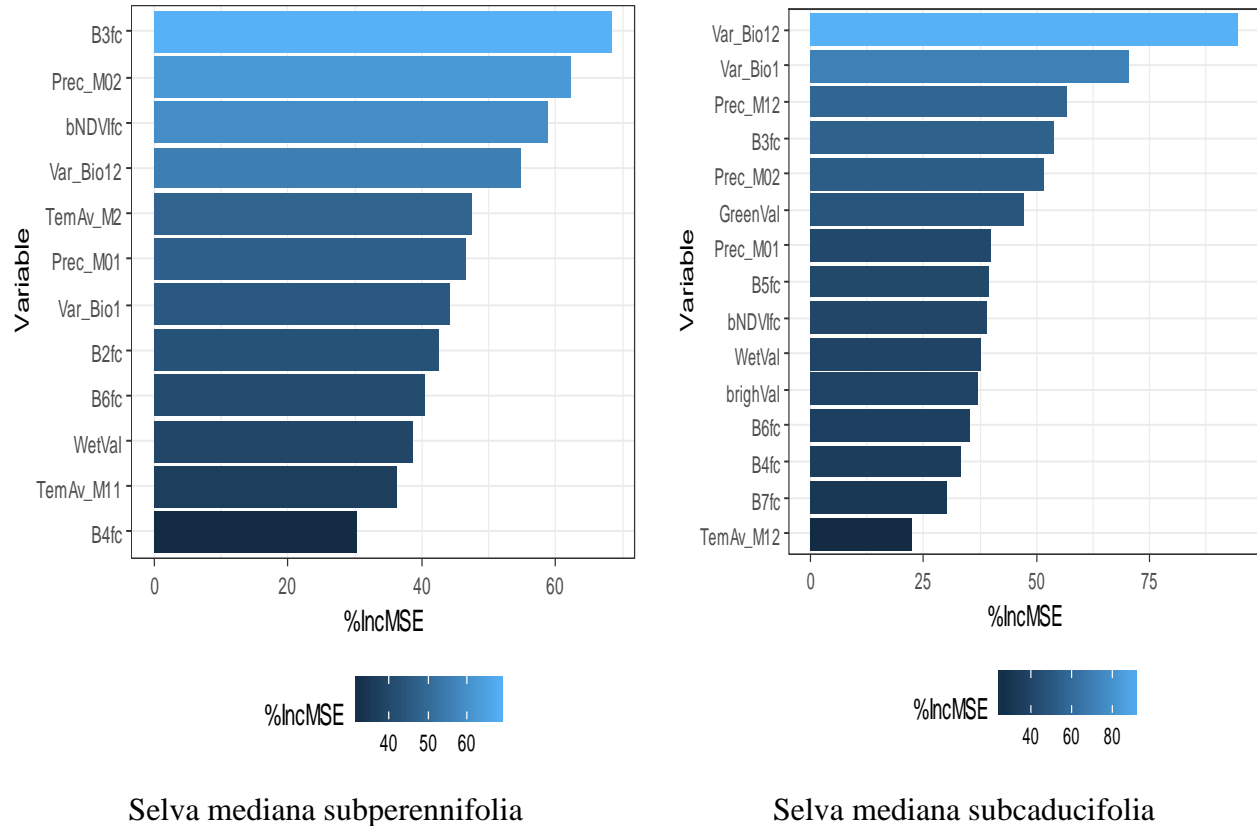


Figura 2.2 Importancia relativa de las variables seleccionadas por el modelo Random Forest en la selva mediana subperennifolia (12 variables) (izq.) y la selva mediana subcaducifolia (15 variables) (der.), respectivamente, para la estimación de biomasa. **B3fc**: reflectancia de la Banda 3 - verde del sensor Landsat 8 OLI; **Prec_M02**: Precipitación del mes 2 (febrero) (mm); **bNDVlfc**: Índice de vegetación de diferencia normalizada; **Var_Bio12**: Precipitación anual (mm); **TemAv_M2**: Temperatura promedio del mes 02 (febrero) (°C); **Prec_M01**: Precipitación del mes 1 (enero) (mm); **Var_Bio1**: Temperatura media anual (°C); **B2fc**: reflectancia de la Banda 2 - azul del sensor Landsat 8 OLI; **B6fc**: reflectancia de la Banda 6 – infrarrojo de onda corta 1 del sensor Landsat 8 OLI; **WetVal**: Humedad en la transformación Tasseled Cap; **TemAv_M11**: Temperatura promedio del mes 11 (noviembre) (°C); **B4fc**: reflectancia de la Banda 4 – roja del sensor Landsat 8 OLI; **Prec_M12**: Precipitación del mes 12 (diciembre) (mm); **GreenVal**: Verdor en la transformación Tasseled Cap; **B5fc**: reflectancia de la Banda 5 - infrarrojo cercano del sensor Landsat 8 OLI; **brighVal**: Brillo en la transformación Tasseled Cap; **B7fc**: reflectancia de la Banda 7 – infrarrojo de onda corta 2 del sensor Landsat 8 OLI; **TeamAv_12**: Temperatura promedio del mes 12 (diciembre) (°C). %IncMSE: incremento porcentual del (MSE) error cuadrático medio.

El incremento porcentual del MSE es un indicador de la contribución de las variables a la precisión predictiva. Es decir, si la variable está contribuyendo al modelo y se permuta (o cambia) aleatoriamente por otro de los predictores en el grupo de prueba, entonces el MSE del

modelo RF aumentará de forma porcentual, debido a que está perdiendo la información que aportaba dicha variable (McInerney & Nieuwenhuis, 2009; Ahmed *et al.*, 2015).

Los resultados mostraron que la banda verde (B3fc) es la que explica en mayor medida la variabilidad de la biomasa aérea para la SMSP mientras que la precipitación anual (Var_Bio12) lo fue para la SMSC. Esto sugiere una fuerte relación entre la banda verde y la biomasa aérea de la SMSP, mientras que para la biomasa de SMSC lo fue con la precipitación anual (mm). Sin embargo, en ambos modelos, las variables climáticas fueron las más sobresalientes para predecir la biomasa aérea en los dos tipos de vegetación.

Mapas de biomasa aérea en los dos tipos de selva mediana

En la Figura 2.3 se presenta la distribución espacial de la biomasa aérea en los dos tipos de selva mediana evaluados. La SMSP presentó un intervalo de valores de biomasa entre 3.5 y 180.0 Mg ha⁻¹ (media = 87.1; desviación estándar = 24.7), los valores más altos se presentan en la parte este y sur de la península de Yucatán. En el caso de la biomasa aérea de la SMSC el intervalo se encontró entre 5.5 a 130.0 Mg ha⁻¹ (media = 57.36; desviación estándar = 22.6).

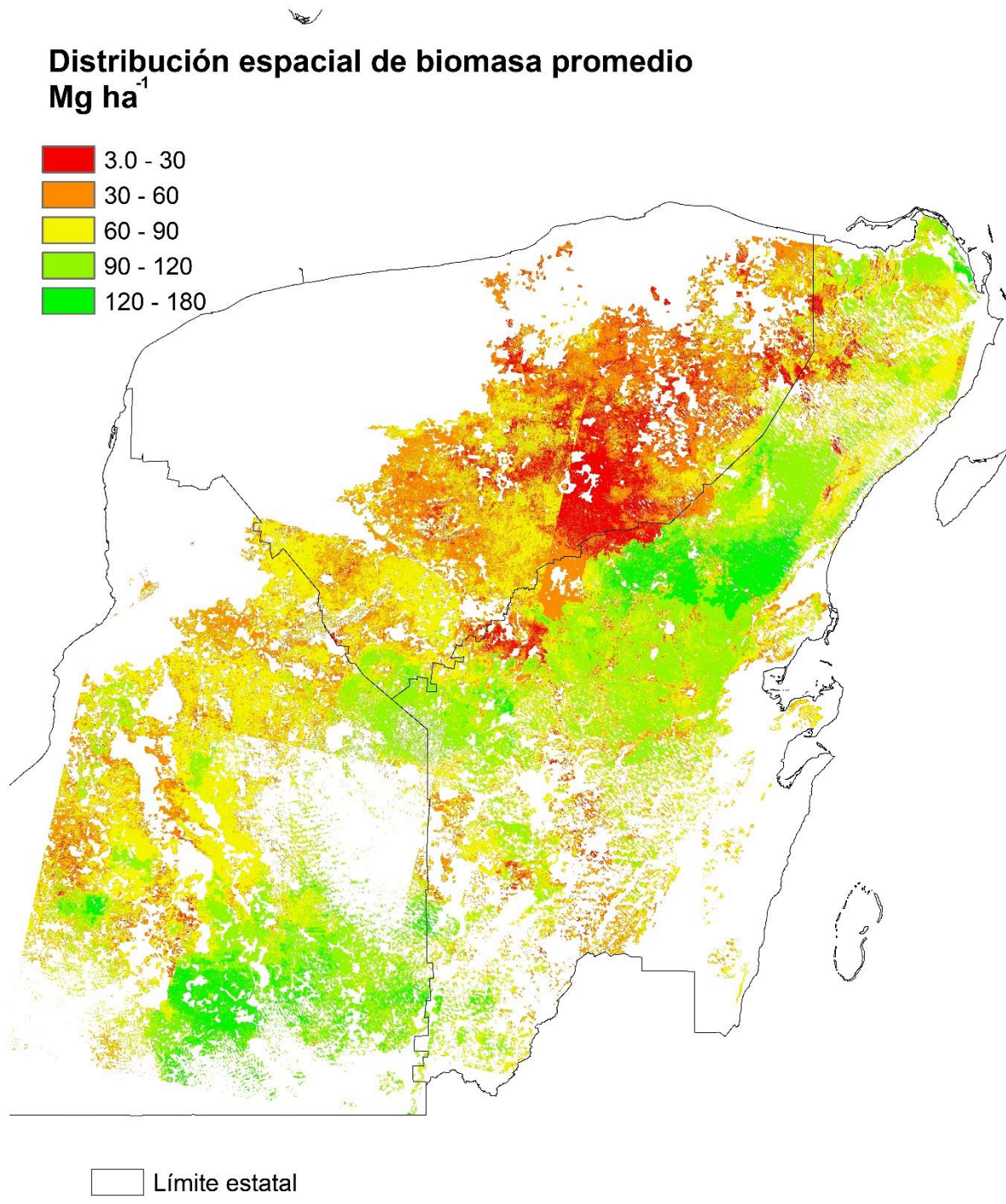


Figura 2.3 Distribución espacial de la biomasa aérea en las selvas mediana subperennifolia y mediana subcaducifolia de la Península de Yucatán, México.

2.4 DISCUSIÓN

Usando una muestra de 5000 píxeles (cada píxel es producto de las estimaciones de biomasa aérea en franjas derivada de datos LiDAR) por tipo de vegetación, se obtuvieron modelos, a partir del algoritmo RF, para estimar la biomasa aérea en dos tipos de selva mediana en la Península de Yucatán. Ambos modelos explican el porcentaje de varianza en 50%, con un RMSE = 34.2 Mg ha⁻¹ para la SMSP y 26.2 Mg ha⁻¹ para la SMSC.

El rendimiento de los modelos de este trabajo se encuentra dentro del intervalo de resultados que exponen estudios previos para estimar la biomasa aérea en bosques tropicales alrededor del mundo ($R^2 = 0.5 - 0.92$). Dichos estudios emplean datos y/o enfoques similares, inclusive comparten características de bosques en diferentes etapas de sucesión, como el reportado por Lu *et al.* (2012). Los autores estimaron la biomasa aérea en el bosque tropical húmedo de la cuenca Amazónica. Usaron medidas de textura de Landsat TM. Mediante regresión múltiple diferenciaron entre bosques maduros ($R^2 = 0.50$) y bosques en sucesión secundaria ($R^2 = 0.76$). El estudio evidenció los problemas de saturación en las imágenes provocado por la estructura compleja en bosques maduros (biomasa media = 248 Mg ha⁻¹). Los autores señalan que la estimación de biomasa aérea usando imágenes Landsat es, en muchos casos, dependiente del sitio debido a la variación en la fenología, así como el tipo y estructura de la vegetación.

Otras investigaciones se han realizado en superficies menores o áreas protegidas, condiciones diversas a las presentadas en este trabajo. Por ejemplo, Basuki *et al.* (2013) usaron datos de PALSAR (Phased Array type L-band Synthetic Aperture Radar) e imágenes Landsat ETM+ para estimar biomasa aérea en bosques tropicales bajo manejo de Indonesia. Emplearon modelos de regresión que explicaron la varianza en un 75% (RMSE = 78.9 Mg ha⁻¹). Phua *et al.* (2017) usaron métricas de LiDAR e imágenes Landsat 8 OLI para estimar la biomasa aérea en un

bosque tropical en Malasia (4294 ha). A través de regresión múltiple, la varianza fue explicada por las métricas LiDAR en 63%, mientras que 18% por las variables de textura de la banda verde (RMSE = 112.15 Mg ha⁻¹). Aplicando RF y Gradient Boosting (GB), Ghosh y Behera (2018) estimaron la biomasa aérea de dos especies plantadas en un bosque tropical en India. Usaron datos SAR (Synthetic Aperture Radar) e información de imágenes Sentinel-2A. La varianza explicada fue 60% y 71% (RMSE = 79.45 Mg ha⁻¹; 105.02 Mg ha⁻¹) con RF y GB respectivamente. A escala regional, Asner y Mascaro (2014) estimaron la densidad de carbono aéreo en 14 ecorregiones tropicales de cinco países. Mediante el ajuste de modelos de probabilidad máxima no lineal explicaron la varianza en 92.3% (RMSE = 17.12 Mg C ha⁻¹).

En los casos anteriores se puede apreciar que el rendimiento de los diferentes métodos de predicción es superior al presentado en este trabajo, sin embargo, los errores obtenidos también son altos (RMSE entre 78.9 Mg ha⁻¹ y 112.15 Mg ha⁻¹) en comparación con los obtenidos en este estudio (RMSE = 34.2 Mg ha⁻¹; 26.2 Mg ha⁻¹ para la SMSP y SMSC respectivamente). No obstante, si se comparan los resultados con los de Asner y Mascaro (2014), que abarcan superficies mayores, el error reportado es similar en términos de biomasa aérea.

Como se puede notar, la precisión de los trabajos depende en gran medida de los sensores empleados, los datos de campo, las condiciones de las áreas de estudio, así como la elección de los enfoques estadísticos como lo señala Lu (2006). Existen variables en común, sin embargo no existe un enfoque generalizado para la estimación de biomasa aérea tomando en cuenta la gran variedad de condiciones en bosques tropicales (Stegen *et al.*, 2011; Deo *et al.*, 2016).

En cuanto a las variables seleccionadas para predecir la biomasa aérea mediante el algoritmo RF, las variables climáticas fueron generalmente primordiales en los dos tipos de selva mediana. Esto sugiere una influencia de los procesos climáticos sobre la biomasa estimada por tipo de

vegetación. La relación entre variables climáticas y la biomasa aérea de bosques tropicales en el mundo ha sido reconocida en diferentes trabajos. Stegen *et al.* (2011) encontraron correlaciones débiles entre la biomasa de bosques tropicales y variables climáticas. Vieilledent *et al.* (2016) señalaron que las variables climáticas sobresalían en importancia, seguida de índices espectrales y la altitud, para predecir la densidad de carbono aéreo en tres tipos de vegetación tropical de Madagascar ($R^2 = 64\%$; $RMSE = 44 \text{ Mg ha}^{-1}$). Por su parte, Alvarez-Davila *et al.* (2017) relacionaron la biomasa aérea y variables del clima en los principales bosques tropicales del noroeste de Sudamérica. La variable disponibilidad de agua (precipitación anual – evapotranspiración potencial) fue el mejor determinante en los modelos cuadráticos generalizados de biomasa aérea evaluados por los autores.

Referente a la precipitación, se reconoce que la principal limitante de las selvas secas es el agua en el suelo (Dupuy *et al.*, 2012), esto podría explicar porque la variable precipitación anual (Var_Bio12) fue la primera para el modelo de la SMSC, mientras que, los promedios mensuales de precipitación (Prec_M01, Prec_M02) permanecieron en el modelamiento de ambos tipos de vegetación. En este sentido, Cao *et al.* (2015) complementan que el crecimiento de este tipo de bosques no solo varía con la edad, sino con otros factores específicos como la precipitación, el tipo de suelo o el antecedente del uso del suelo.

Para el caso de las variables relacionadas a la temperatura, el algoritmo RF seleccionó la temperatura media anual ($^{\circ}\text{C}$) en ambos modelos y una combinación diferente de promedios mensuales tanto para la SMSP como la SMSC. Lo anterior pone en evidencia lo sugerido por Alvarez-Davila *et al.* (2017), que la temperatura influye en la biomasa ya que la actividad fotosintética es dependiente de la temperatura.

Lo anterior, sugiere la importancia de la cantidad de agua disponible en cada ecosistema y su interacción con la temperatura para influir sobre los procesos de crecimiento y mortalidad. En tal caso, tanto la precipitación como la temperatura estarían ejerciendo un control superior sobre el contenido de biomasa aérea presente en los bosques tropicales (Alvarez-Davila *et al.* 2017).

Así mismo, se debe tomar en cuenta que los efectos climáticos, que influyen sobre la biomasa aérea, varían regionalmente y pueden ser dependientes de la escala. Saatchi *et al.* (2011) señalan que la variabilidad espacial de la biomasa aérea depende de factores como el clima, procesos de disturbio y recuperación provocados de forma natural y por el hombre, el tipo de suelo y variaciones en la topografía. Martínez & Leal, (2002) apuntan que la alta variabilidad espacial y temporal de la precipitación, el sustrato geológico y el escaso desarrollo del suelo, son factores decisivos en la distribución de la vegetación en la región de Calakmul, área coincidente con este estudio.

En este trabajo, los modelos fueron capaces de reconocer la importancia de las variables climáticas en ambos tipos de selva. Lo anterior, apoya la idea que a esta escala los factores climáticos son los que dictan la variación en la vegetación (Martínez & Leal, 2002; White & Hood 2004).

En cuanto a las variables espectrales que el algoritmo RF seleccionó para la predicción de biomasa en la SMSC, son las mismas o similares que han mostrado buena capacidad predictiva en bosques de condiciones semejantes. Freitas *et al.* (2005) señala que el NDVI es un buen indicador de la biomasa en bosques secos y deciduos, sin embargo, la sensibilidad de los índices de vegetación a la biomasa varía entre los diferentes ambientes (Foody *et al.*, 2001). El NDVI fue el único de los índices espectrales que se mantuvo para ambos modelos, los demás se fueron

eliminando a medida que se entrenaban los modelos, ya que aparecían como las variables con menor contribución a la explicación de la varianza.

Sin embargo, algunos autores señalan que el uso individual del NDVI puede no ser un buen indicador de las características biofísicas en bosques tropicales densos debido al problema de saturación, sugiriendo complementar al modelo seleccionado con otros índices que aprovechen la información de las demás bandas infrarrojas (Foody *et al.*, 2001). Dicho autor señaló que después de realizar un análisis de sensibilidad, el orden de importancia de las variables empleadas fueron: (las bandas 5, 2 y 4 -SWIR1, Verde y NIR-) de Landsat 5 (Thematic Mapper -TM-), mientras que la banda 3 (roja) fue la menos significativa.

En este trabajo, las bandas azul, verde e infrarrojos fueron componentes comunes con los índices de vegetación que Foody *et al.* (2001) evaluó. De ahí que el autor señaló la importancia de considerar todas las bandas útiles que registra el sensor y no solo los índices que se construyen principalmente con la banda roja. Las bandas verde, roja e infrarroja (B3fc, B6fc y B4fc) se mantuvieron en ambos tipos de vegetación. La B3fc, primera y cuarta en la lista de importancia para la SMSP y SMSC respectivamente, es capaz de enfatizar los picos de reflectancia de la vegetación permitiendo evaluar el vigor de las plantas. Esta relevancia también fue encontrada en los estudios de Foody *et al.* (2001) y Phua *et al.* (2017) para bosques tropicales. Por su parte, las características de B6fc (SWIR1) hacen que sea posible detectar la humedad de las hojas y la estructura del dosel (Lu, 2006). Freitas *et al.* (2005) señala que dicha banda podría ser una variable adecuada para estimar la madurez estructural del bosque en los trópicos.

En cuanto a los valores registrados de biomasa aérea, los resultados de este trabajo (3.5 – 180 Mg ha⁻¹, media 87.1 Mg ha⁻¹) para la SMSP se encuentran dentro de los resultados reportados por otros autores. El estudio de Puc (2014), en una selva mediana subperennifolia, reportó que la

biomasa aérea promedio osciló entre 39.75 Mg ha⁻¹ y 175.35 Mg ha⁻¹ en rodales con 7 y 46 años de abandono respectivamente, mientras que en el bosque maduro, la biomasa promedio fue de 179.14 Mg ha⁻¹. Se puede decir que los resultados de dicho autor, en relación con la biomasa, son mayores porque incluyó árboles con diámetros normales a partir de 2.5 cm, así mismo, los sitios clasificados como rodales maduros se establecieron en un área bajo conservación. Y en el caso de este estudio, las franjas con la estimación de biomasa aérea, atraviesan estratos en diferente estadio sucesional, lo que hace que exista una mayor variabilidad espacial que se refleja en los datos.

Por su parte, Aryal *et al.* (2014) indicaron que la biomasa aérea en áreas aledañas a la Reserva de la Biosfera de Calakmul, Campeche presentó una variación de 11.72 Mg C ha⁻¹ en bosques secundarios de 4 años a 99.56 Mg C ha⁻¹ en bosques maduros, lo cual en términos de biomasa aérea, corresponde al intervalo reportado en este estudio para la SMSP.

En relación con los resultados reportados en otros estudios correspondientes a la SMSC, Dupuy-Rada *et al.* (2012) evaluaron la biomasa aérea en 352 km² de bosques tropicales secos de la Península de Yucatán. Concluyeron que la edad sucesional fue el factor que más influyó sobre la biomasa aérea con una R²= 0.68. La condición actual de la biomasa aérea encontrada en dicho estudio fue de 56 Mg ha⁻¹, lo que es similar a la media reportada en este trabajo (57.36 Mg ha⁻¹) para el mismo tipo de vegetación. Se debe considerar que en ambos trabajos, los datos de biomasa aérea fueron obtenidos a partir de mosaicos de fragmentos de bosque en diferentes edades sucesionales y arreglo espacial.

Sin embargo, también existen estimaciones que reportan el doble de la cantidad de biomasa aérea encontrada en este estudio para la SMSC. Hernández-Stefanoni *et al.* (2014) reportaron una media de 109.71 Mg ha⁻¹ en sitios muestreados de 1000 m². Mientras que para la SMSP el

mismo autor reportó una media de 376.77 Mg ha⁻¹. Cabe aclarar que dicha investigación se realizó a nivel local en áreas de 9 km², una de ellas es un área privada protegida (SMSC), mientras que la otra (SMSP) se establece en tierras comunales, y este estudio se realizó a nivel regional analizando datos de biomasa de los transectos que se extienden sobre parches de bosques abiertos y degradados, además de bosques mejor conservados; es decir, existe una alta variabilidad cubierta por las franjas.

Por su parte, Dai *et al.* (2014) simularon valores de biomasa aérea mediante el modelo de procesos Forest-DeNitrification-DeComposition en una selva mediana subcaducifolia en la Península de Yucatán. Obtuvieron una eficiencia del modelo de $R^2 = 0.83$. Posteriormente simularon la biomasa aérea para toda el área de estudio (350 km²), reportando un intervalo de 5.0 a 115.0 Mg C ha⁻¹ con una media de 56.6 Mg C ha⁻¹. Estos resultados nuevamente corresponden al doble de lo que este estudio estimó, sin embargo, cabe notar que el estudio de dicho autor se llevó a cabo con datos del área privada protegida que se mencionó líneas arriba. Midieron todos los individuos con $DN \geq 1$ cm, además que este tipo de modelos son ajustados mediante una serie de parámetros tanto de la vegetación como del suelo. No obstante, los resultados mostrados en este estudio siguen dentro del intervalo reportado para este tipo de bosques tropicales.

Este trabajo se desarrolló en una región contrastante con tasas altas de deforestación y un mosaico de rodales secundarios en diferentes estados sucesionales, producto de las rotaciones de cultivo, el abandono de tierras agrícolas o pastizales y disturbios naturales como incendios o huracanes (Dupuy-Rada *et al.*, 2012; McGroddy *et al.*, 2013; Aryal *et al.*, 2014; López-Jiménez *et al.*, 2019). Dichos estados sucesionales representan una alta complejidad al momento de modelar. Los bosques jóvenes son frecuentemente más irregulares y complejos en cuanto a

estructura del dosel que los bosques maduros, lo que finalmente repercute en la precisión de los modelos de variables en bosques tropicales (Ahmed *et al.*, 2015).

De acuerdo con Ewel (1980), la sucesión ecológica en las selvas secas (SMSC) con relación a las selvas húmedas (SMSP), puede suceder en un lapso de tiempo menor y entonces todas las etapas de la sucesión no se presentarían, resultando en una estructura menos compleja y diversidad menor. Por su parte Cao *et al.* (2015) señalan que mapear los bosques tropicales secos durante la estación seca es mejor debido a que los parámetros biofísicos tienden a ser más pronunciados y distintivos, independientemente del sensor empleado (activo o pasivo). Estas características podrían ser la causa por la que el modelo de la SMSC presentara un menor error (RMSE = 22.6 Mg ha⁻¹), aunque la varianza explicada no fue precisamente mayor.

Así mismo, existen otros elementos en los bosques tropicales que incluyen lianas, así como las sombras provocadas por la estructura compleja del bosque, que pueden ser fuentes de incertidumbre en el espectro del pixel. La presencia o ausencia de estas, pueden alterar en gran medida la respuesta espectral de los bosques (Basuki *et al.*, 2013; Cao *et al.*, 2015).

A pesar de la limitada capacidad explicativa de los modelos presentados, los resultados son adecuados en estimaciones a escala regional. Los errores registrados no exceden los 35 Mg ha⁻¹, aunque se esperaría que los errores obtenidos en este tipo de estudios y utilizando datos de percepción remota fueran próximos a 20 Mg ha⁻¹ o estar dentro del intervalo del 20% de las estimaciones obtenidas en campo (Zolkos *et al.*, 2013).

2.5 CONCLUSIONES

El uso combinado de datos espectrales obtenidos, tanto por sensores pasivos (ópticos – multiespectrales) como activos (p. ej. LiDAR), que recojan información del vigor y la estructura del bosque respectivamente, será siempre la mejor alternativa para estudiar ecosistemas de

vegetación densa en aquellas áreas de estudio de tamaño considerable, donde la saturación de los índices de vegetación tradicionales ha representado un reto para la adecuada modelación geoespacial de la densidad forestal. A pesar de que la disponibilidad de datos espectrales para este tipo de ecosistemas es afectada frecuentemente por la nubosidad, siempre habrá una alternativa instrumental y metodológica para el adecuado monitoreo de sus recursos forestales, que permita conocer la distribución espacial de las existencias reales en términos de alguna medida de densidad forestal, ya sea volumen, biomasa o carbono.

Desde el punto de vista de la distribución espacial de la biomasa en la península de Yucatán, los bosques tropicales maduros están siendo reemplazados por paisajes complejos compuestos de áreas agrícolas, pastizales y parches de vegetación en diferentes grados de sucesión. Esta variabilidad fue capturada por los datos de entrenamiento (estimaciones de biomasa aérea en franjas de datos LiDAR) repercutiendo en el desempeño de los modelos para predecir la biomasa aérea en ambos tipos de vegetación.

Finalmente, los modelos generados permitieron obtener un mapa continuo que detalla la distribución espacial de biomasa aérea a nivel de pixel (30 m de resolución espacial) en dos tipos de selvas medianas. Dicha distribución fue explicada en mayor proporción por las variables climáticas dada su relación a esta escala. La metodología empleada optimizó el potencial utilitario que brindan las diferentes fuentes de información y los diferentes enfoques de percepción remota, minimizando los costos operativos y alcanzando una precisión razonable.

2.6 LITERATURA CITADA

Ahmed, O. S., Franklin, S. E., Wulder, M. A., & White, J. C. (2015). Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 89-101. doi:10.1016/j.isprsjprs.2014.11.007

- Alvarez-Davila, E., Cayuela, L., Gonzalez-Caro, S., Aldana, A. M., Stevenson, P. R., Phillips, O., . . . Rey-Benayas, J. M. (2017). Forest biomass density across large climate gradients in northern South America is related to water availability but not with temperature. *PLoS One*, 12(3), e0171072. doi:10.1371/journal.pone.0171072
- Aryal, D. R., De Jong, B. H. J., Ochoa-Gaona, S., Esparza-Olguin, L., & Mendoza-Vega, J. (2014). Carbon stocks and changes in tropical secondary forests of southern Mexico. *Agriculture, Ecosystems & Environment*, 195, 220-230. doi:10.1016/j.agee.2014.06.005
- Asner, G. P., & Mascaro, J. (2014). Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment*, 140, 614-624. doi:10.1016/j.rse.2013.09.023
- Baccini, A., Friedl, M. A., Woodcock, C. E., & Warbington, R. (2004). Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters*, 31(10), n/a-n/a. doi:10.1029/2004gl019782
- Barbosa, J. M., Broadbent, E. N., & Bitencourt, M. D. (2014). Remote Sensing of Aboveground Biomass in Tropical Secondary Forests: A Review. *International Journal of Forestry Research*, 2014, 1-14. doi:10.1155/2014/715796
- Basuki, T. M., Skidmore, A. K., Hussin, Y. A., & Van Duren, I. (2013). Estimating tropical forest biomass more accurately by integrating ALOS PALSAR and Landsat-7 ETM+ data. *International Journal of Remote Sensing*, 34(13), 4871-4888. doi:10.1080/01431161.2013.777486
- Boyd, D., & Danson, F. (2005). Satellite remote sensing of forest resources: three decades of research development. *Progress in Physical Geography*, 29(1), 1-26.
- Broszofski, K. D., Froese, R. E., Falkowski, M. J., & Banskota, A. (2014). A Review of Methods for Mapping and Prediction of Inventory Attributes for Operational Forest Management. *Forest Science*, 60(4), 733-756. doi:10.5849/forsci.12-134
- Cao, S., Yu, Q., Sanchez-Azofeifa, A., Feng, J., Rivard, B., & Gu, Z. (2015). Mapping tropical dry forest succession using multiple criteria spectral mixture analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109, 17-29. doi:10.1016/j.isprsjprs.2015.08.009
- Caughlin, T. T., Rifai, S. W., Graves, S. J., Asner, G. P., Bohlman, S. A., Nagendra, H., & Fatoyinbo, L. (2016). Integrating LiDAR-derived tree height and Landsat satellite reflectance to estimate forest regrowth in a tropical agricultural landscape. *Remote Sensing in Ecology and Conservation*, 2(4), 190-203. doi:10.1002/rse2.33
- Chen, G., & Hay, G. J. (2011). A support vector regression approach to estimate forest biophysical parameters at the object level using airborne lidar transects and quickbird data. *Photogrammetric Engineering & Remote Sensing*, 77(7), 733-741.
- Crist, E. P. (1985). A TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sensing of Environment*, 17(3), 301-306. doi:10.1016/0034-4257(85)90102-6

- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In *Ensemble machine learning* (pp. 157-175): Springer.
- Dai, Z., Birdsey, R. A., Johnson, K. D., Dupuy, J. M., Hernandez-Stefanoni, J. L., & Richardson, K. (2014). Modeling Carbon Stocks in a Secondary Tropical Dry Forest in the Yucatan Peninsula, Mexico. *Water, Air, & Soil Pollution*, 225(4). doi:10.1007/s11270-014-1925-x
- Deo, R. K., Russell, M. B., Domke, G. M., Woodall, C. W., Falkowski, M. J., & Cohen, W. B. (2016). Using Landsat Time-Series and LiDAR to Inform Aboveground Forest Biomass Baselines in Northern Minnesota, USA. *Canadian Journal of Remote Sensing*, 43(1), 28-47. doi:10.1080/07038992.2017.1259556
- DeVries, B., Pratihast, A. K., Verbesselt, J., Kooistra, L., & Herold, M. (2016). Characterizing Forest Change Using Community-Based Monitoring Data and Landsat Time Series. *PLoS One*, 11(3), e0147121. doi:10.1371/journal.pone.0147121
- Dupuy-Rada, J., Hernández-Stefanoni, J., Hernández-Juárez, R., Tun-Dzul, F., & May-Pat, F. (2012). Efectos del cambio de uso del suelo en la biomasa y diversidad de plantas leñosas en un paisaje de bosque tropical seco en Yucatán. *Investigación Ambiental Ciencia y Política Pública*, 4, 130-140.
- Dupuy, J. M., Hernández-Stefanoni, J. L., Hernández-Juárez, R. A., Tetetla-Rangel, E., López-Martínez, J. O., Leyequién-Abarca, E., . . . May-Pat, F. (2012). Patterns and correlates of tropical dry forest structure and composition in a highly replicated chronosequence in Yucatan, Mexico. *Biotropica*, 44(2), 151-162.
- Ene, L. T., Næsset, E., & Gobakken, T. (2016). Simulation-based assessment of sampling strategies for large-area biomass estimation using wall-to-wall and partial coverage airborne laser scanning surveys. *Remote Sensing of Environment*, 176, 328-340. doi:10.1016/j.rse.2016.01.025
- Ewel, J. (1980). Tropical succession: manifold routes to maturity. *Biotropica*, 12(2), 2-7. doi:10.2307/2388149
- Feeley, K. J., Gillespie, T. W., & Terborgh, J. W. (2005). The Utility of Spectral Indices from Landsat ETM+ for Measuring the Structure and Composition of Tropical Dry Forests 1. *Biotropica: The Journal of Biology and Conservation*, 37(4), 508-519.
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302-4315. doi:10.1002/joc.5086
- Foody, G. M., Cutler, M. E., Mcmorrow, J., Pelz, D., Tangki, H., Boyd, D. S., & Douglas, I. (2001). Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Global ecology and biogeography*, 10(4), 379-387.

- Freitas, S. R., Mello, M. C. S., & Cruz, C. B. M. (2005). Relationships between forest structure and vegetation indices in Atlantic Rainforest. *Forest Ecology and Management*, 218(1-3), 353-362. doi:10.1016/j.foreco.2005.08.036
- Gao, X., Huete, A. R., Ni, W., & Miura, T. (2000). Optical–biophysical relationships of vegetation spectra without background contamination. *Remote Sensing of Environment*, 74(3), 609-620.
- Ghosh, S. M., & Behera, M. D. (2018). Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. *Applied Geography*, 96, 29-40. doi:10.1016/j.apgeog.2018.05.011
- Healey, S., Cohen, W., Zhiqiang, Y., & Krankina, O. (2005). Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. *Remote Sensing of Environment*, 97(3), 301-310. doi:10.1016/j.rse.2005.05.009
- Hengl, T., Nussbaum, M., Wright, M., Heuvelink, G., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ Preprints*, 6:e26693v3, doi.org/10.7287/peerj.preprints.26693v3
- Hernández-Stefanoni, J., Dupuy, J., Johnson, K., Birdsey, R., Tun-Dzul, F., Peduzzi, A., . . . López-Merlín, D. (2014). Improving Species Diversity and Biomass Estimates of Tropical Dry Forests Using Airborne LiDAR. *Remote Sensing*, 6(12), 4741-4763. doi:10.3390/rs6064741
- Houghton, R. A., Byers, B., & Nassikas, A. A. (2015). A role for tropical forests in stabilizing atmospheric CO₂. *Nature Climate Change*, 5(12), 1022-1023. doi:10.1038/nclimate2869
- Houghton, R. A., Hall, F., & Goetz, S. J. (2009). Importance of biomass in the global carbon cycle. *Journal of Geophysical Research: Biogeosciences*, 114(G2), n/a-n/a. doi:10.1029/2009jg000935
- Huxman, T. E., Smith, M. D., Fay, P. A., Knapp, A. K., Shaw, M. R., Loik, M. E., . . . Weltzin, J. F. (2004). Convergence across biomes to a common rain-use efficiency. *Nature*, 429(6992), 651.
- INEGI. (2013). *Conjunto Nacional de Uso del Suelo y Vegetación a escala 1:250,000*. Aguascalientes, México: INEGI.
- Ingram, J. C., Dawson, T. P., & Whittaker, R. J. (2005). Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. *Remote Sensing of Environment*, 94(4), 491-507. doi:10.1016/j.rse.2004.12.001
- Kardol, P., Company, C. E., Souza, L., Norby, R. J., Weltzin, J. F., & Classen, A. T. (2010). Climate change effects on plant biomass alter dominance patterns and community evenness in an experimental old-field ecosystem. *Global Change Biology*, 16(10), 2676-2687. doi:10.1111/j.1365-2486.2010.02162.x

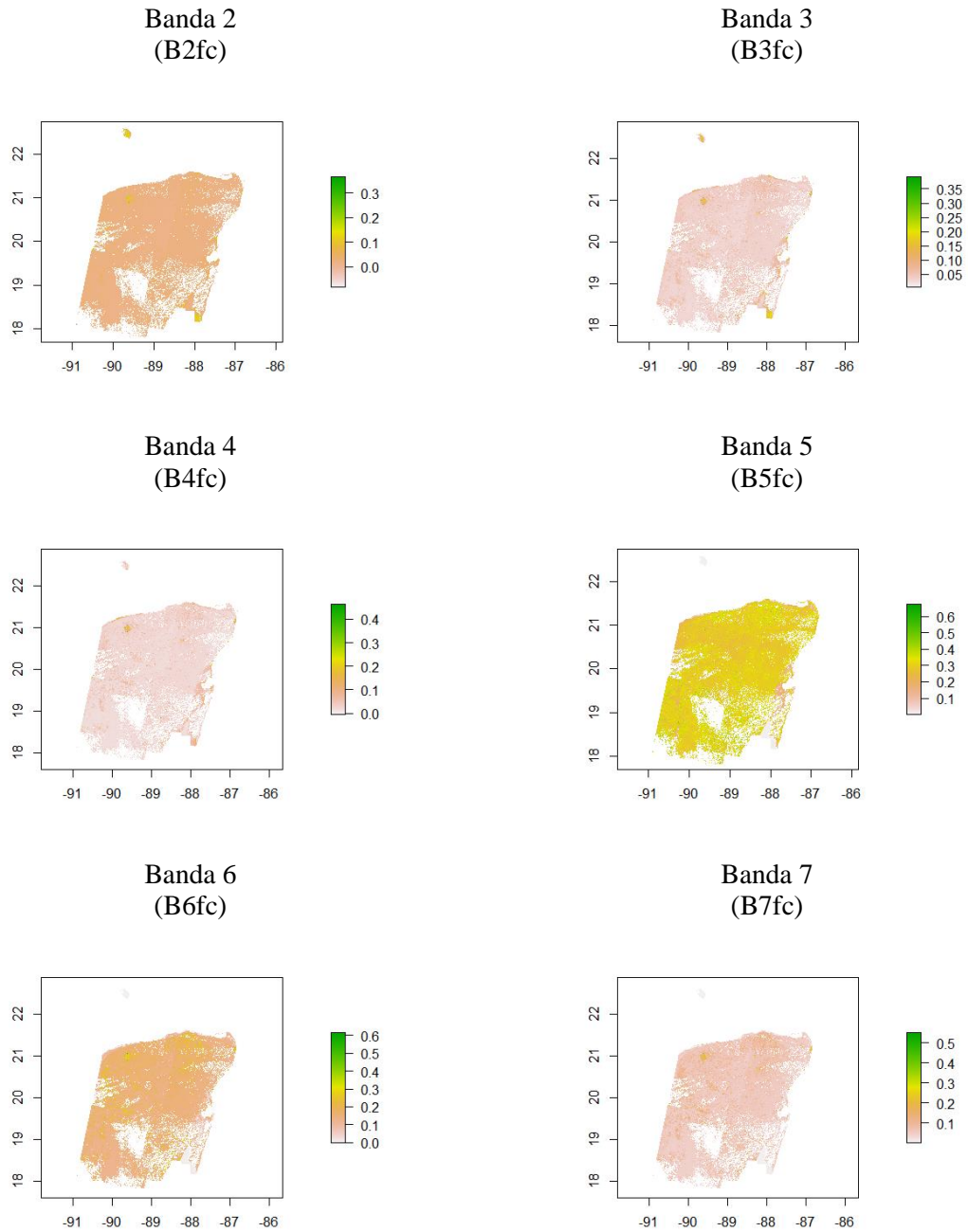
- López-Jiménez, L. N., Durán-García, R., & Dupuy-Rada, J. M. (2019). Recuperación de la estructura, diversidad y composición en una selva mediana subperennifolia en Yucatán, México. *Madera y bosques*, 25(1). doi:10.21829/myb.2019.2511587
- Lu, D. (2006). The potential and challenge of remote sensing- based biomass estimation. *International Journal of Remote Sensing*, 27(7), 1297-1328. doi:10.1080/01431160500486732
- Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., . . . Saah, D. (2012). Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *International Journal of Forestry Research*, 2012, 1-16. doi:10.1155/2012/436537
- Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest Ecology and Management*, 198(1-3), 149-167. doi:10.1016/j.foreco.2004.03.048
- Márdero, S., Nickl, E., Schmook, B., Schneider, L., Rogan, J., Christman, Z., & Lawrence, D. (2012). Sequías en el sur de la península de Yucatán: análisis de la variabilidad anual y estacional de la precipitación. *Investigaciones Geográficas*(78), 19-33.
- Martínez, E., & Leal, C. G. (2002). La vegetación de Calakmul, Campeche, México: clasificación, descripción y distribución. *Boletín de la Sociedad Botánica de México*(71), 7-32.
- McGroddy, M., Lawrence, D., Schneider, L., Rogan, J., Zager, I., & Schmook, B. (2013). Damage patterns after Hurricane Dean in the southern Yucatán: Has human activity resulted in more resilient forests? *Forest Ecology and Management*, 310, 812-820. doi:10.1016/j.foreco.2013.09.027
- McInerney, D. O., & Nieuwenhuis, M. (2009). A comparative analysis of kNN and decision tree methods for the Irish National Forest Inventory. *International Journal of Remote Sensing*, 30(19), 4937-4955. doi:10.1080/01431160903022936
- Mora, B., Wulder, M. A., Hobart, G. W., White, J. C., Bater, C. W., Gougeon, F. A., . . . Coops, N. C. (2013). Forest inventory stand height estimates from very high spatial resolution satellite imagery calibrated with lidar plots. *International Journal of Remote Sensing*, 34(12), 4406-4424. doi:10.1080/01431161.2013.779041
- Ortiz-Reyes, A. D., Valdez-Lazalde, J. R., Ángeles-Pérez, G., De los Santos-Posadas, H. M., Schneider, L., Aguirre-Salado, C. A., & Peduzzi, A. (2019). Transectos de datos LiDAR: una estrategia de muestreo para estimar biomasa aérea en áreas forestales. *Madera y Bosques*, 25(3), e2531872. doi: 10.21829/myb.2019.2531872
- Palace, M. W., Sullivan, F. B., Ducey, M. J., Treuhaft, R. N., Herrick, C., Shimbo, J. Z., & Mota-E-Silva, J. (2015). Estimating forest structure in a tropical forest using field measurements, a synthetic model and discrete return lidar data. *Remote Sensing of Environment*, 161, 1-11. doi:10.1016/j.rse.2015.01.020

- Pennington, T. D. (2005). *Árboles tropicales de México: manual para la identificación de las principales especies*: UNAM.
- Pflugmacher, D., Cohen, W. B., & E. Kennedy, R. (2012). Using Landsat-derived disturbance history (1972–2010) to predict current forest structure. *Remote Sensing of Environment*, 122, 146-165. doi:10.1016/j.rse.2011.09.025
- Phua, M.-H., Johari, S. A., Wong, O. C., Ioki, K., Mahali, M., Nilus, R., . . . Hashim, M. (2017). Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests. *Forest Ecology and Management*, 406, 163-171. doi:10.1016/j.foreco.2017.10.007
- Puc Kauil, R. (2014). *Acumulación de biomasa y carbono aéreo en bosques tropicales secundarios del sur de Quintana Roo, México*. (Tesis de Maestría), Colegio de Postgraduados, Texcoco, México.
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119-126. doi:10.1016/0034-4257(94)90134-1
- Quantum, G. (2019). QGIS Geographic Information System. Open Source Geospatial Foundation Project. (Version 3.6 Noosa). Retrieved from <https://qgis.org>.
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. 2009-2017.
- Saarela, S., Grafström, A., Ståhl, G., Kangas, A., Holopainen, M., Tuominen, S., . . . Hyypä, J. (2015). Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information. *Remote Sensing of Environment*, 158, 431-440. doi:10.1016/j.rse.2014.11.020
- Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., . . . Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc Natl Acad Sci U S A*, 108(24), 9899-9904. doi:10.1073/pnas.1019576108
- Simonetti, E., Simonetti, D. & Preatoni, D. (2014). Phenology-based land cover classification using Landsat 8 time series. Publications Office of the European Union-Scientific and Technical Research Reports. Intitute for Environment and Sustainability of the Joint Research Center (JRC). Ispra, Italy. 58 p.
- Stegen, J. C., Swenson, N. G., Enquist, B. J., White, E. P., Phillips, O. L., Jørgensen, P. M., . . . Núñez Vargas, P. (2011). Variation in above-ground forest biomass across broad climatic gradients. *Global ecology and biogeography*, 20(5), 744-754. doi:10.1111/j.1466-8238.2010.00645.x

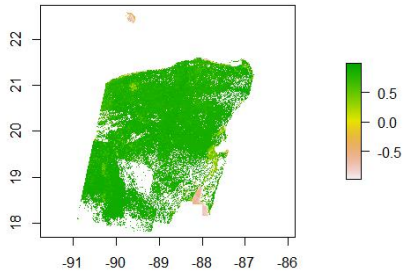
- USGS. (2019). What are the best Landsat spectral bands for use in my research? Retrieved from https://www.usgs.gov/faqs/what-are-best-landsat-spectral-bands-use-my-research?qt-news_science_products=7#qt-news_science_products
- Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, 46-56.
- Vieilledent, G., Gardi, O., Grinand, C., Burren, C., Andriamanjato, M., Camara, C., . . . Lines, E. (2016). Bioclimatic envelope models predict a decrease in tropical forest carbon stocks with climate change in Madagascar. *Journal of Ecology*, 104(3), 703-715. doi:10.1111/1365-2745.12548
- White, D. A., & Hood, C. S. (2004). Vegetation patterns and environmental gradients in tropical dry forests of the northern Yucatan Peninsula. *Journal of Vegetation Science*, 15(2), 151-160. doi:10.1111/j.1654-1103.2004.tb02250.x
- White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., & Tompalski, P. (2016). Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Canadian Journal of Remote Sensing*, 42(5), 619-641. doi:10.1080/07038992.2016.1207484
- White, J. C., Wulder, M., Vastaranta, M., Coops, N., Pitt, D., & Woods, M. (2013). The Utility of Image-Based Point Clouds for Forest Inventory: A Comparison with Airborne Laser Scanning. *Forests*, 4(3), 518-536. doi:10.3390/f4030518
- Wilkes, P., Jones, S. D., Suarez, L., Mellor, A., Woodgate, W., Soto-Berelov, M., . . . Skidmore, A. K. (2015). Mapping forest canopy height across large areas by upscaling ALS estimates with freely available satellite data. *Remote Sensing*, 7(9), 12563-12587. doi:10.3390/rs70912563
- Wulder, M., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, 122, 2-10.
- Wulder, M. A., White, J. C., Bater, C. W., Coops, N. C., Hopkinson, C., & Chen, G. (2012). Lidar plots—A new large-area data collection option: Context, concepts, and case study. *Canadian Journal of Remote Sensing*, 38(5), 600-618.
- Wulder, M. A., White, J. C., Nelson, R. F., Næsset, E., Ørka, H. O., Coops, N. C., . . . Gobakken, T. (2012). Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*, 121, 196-209. doi:10.1016/j.rse.2012.02.001
- Young, N. E., Anderson, R. S., Chignell, S. M., Vorster, A. G., Lawrence, R., & Evangelista, P. H. (2017). A survival guide to Landsat preprocessing. *Ecology*, 98(4), 920-932.
- Zolkos, S. G., Goetz, S. J., & Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment*, 128, 289-298. doi:10.1016/j.rse.2012.10.017

2.7 Anexo 1

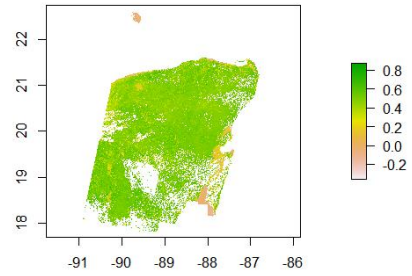
Distribución espacial de las variables espectrales y climáticas empleadas para modelar biomasa aérea en dos tipos de selva medianas de la Península de Yucatán



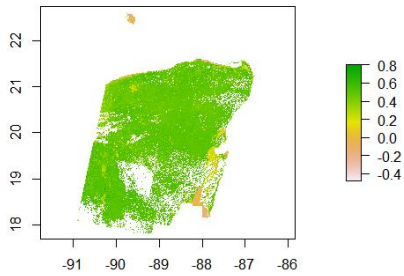
NDVI
(bNDVIfc)



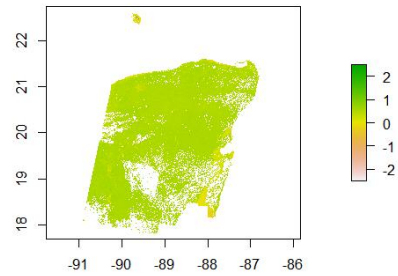
MSAVI
(bMSAVIfc)



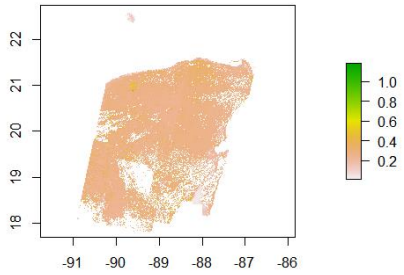
SAVI
(bSAVIfc)



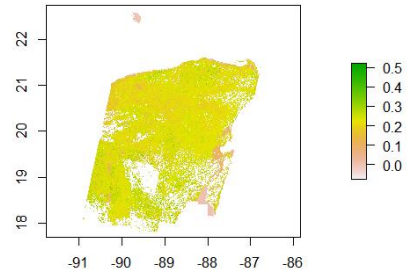
EVI
(bEVIfc)



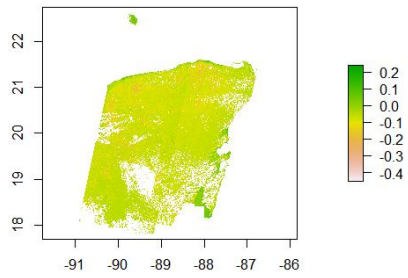
TCB
(brighVal)



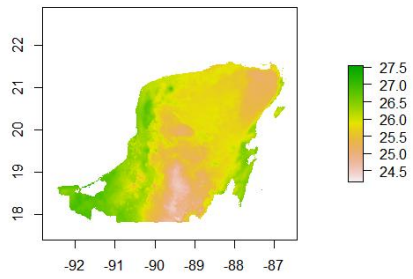
TCV
(GreenVal)



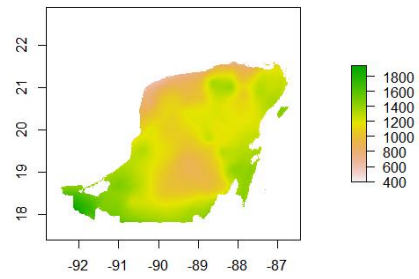
TCH
(WetVal)



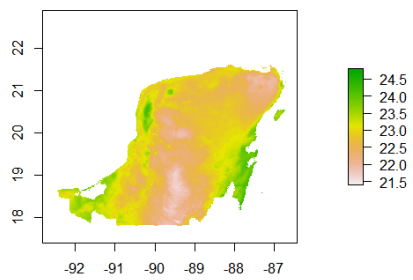
Temperatura media anual (Var_Bio1) en °C



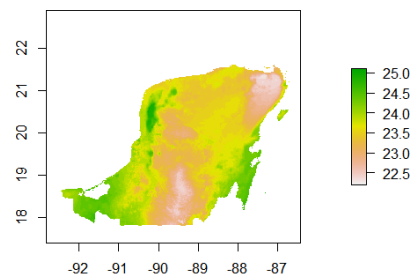
Precipitación anual (Var_Bio12) en mm



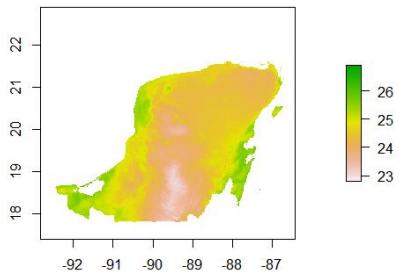
Temperatura promedio del mes de enero
(TemAv_M01) en °C



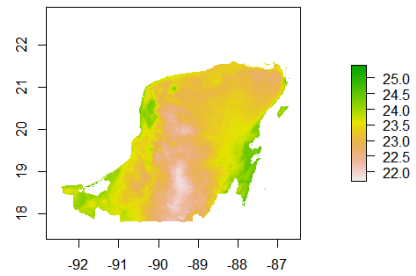
Temperatura promedio del mes de febrero
(TemAv_M02) en °C



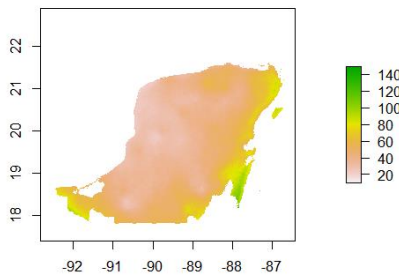
Temperatura promedio del mes de noviembre
(TemAv_M11) en °C



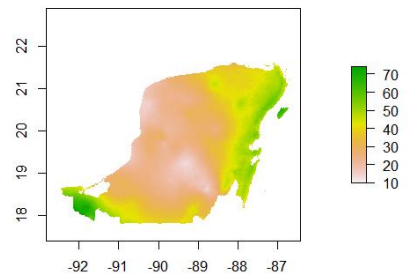
Temperatura promedio del mes de diciembre
(TemAv_M12) en °C



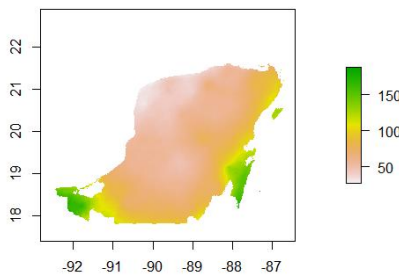
Precipitación promedio del mes de enero
(Prec_M01) en mm



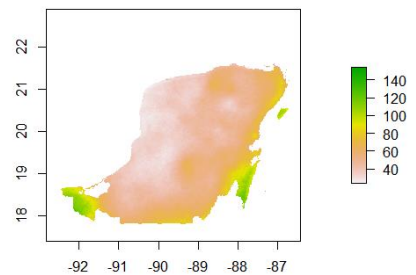
Precipitación promedio del mes de febrero
(Prec_M02) en mm



Precipitación promedio del mes de noviembre
(Prec_M11) en mm



Precipitación promedio del mes de diciembre
(Prec_M12) en mm



B3fc – B7fc Valores de reflectancia para cada banda

Índices NDVI, MSAVI, SAVI, EVI, TCB, TCV y TCH son adimensionales.

**CAPÍTULO III. EVALUACIÓN DE LA INCERTIUMBRE ASOCIADA A LAS
ESTIMACIONES DE BIOMASA AÉREA EN DOS SELVAS MEDIANAS EN LA
PENÍNSULA DE YUCATÁN**

RESUMEN

La estimación de la biomasa aérea forestal y su distribución espacial en grandes áreas, así como la incertidumbre asociada permite entender el comportamiento del ciclo del carbono y ofrecer directrices a las políticas de cambio climático. El uso combinado de datos de campo, LiDAR y datos espectrales han mostrado ser de gran ayuda para estimar la biomasa aérea de áreas forestales desde un nivel de parcela hasta nivel regional. Desafortunadamente, durante el proceso de modelación se generan errores (incertidumbre) que es necesario evaluar y reportar a los usuarios. En este estudio se evaluó la incertidumbre, a nivel de pixel, asociada a la estimación de biomasa aérea en selvas medianas mediante el algoritmo Quantile Regression Forests (QRF) que calcula intervalos de predicción. Se reportan intervalos de predicción que van desde 0 a 250 Mg ha⁻¹ para la selva mediana subperennifolia (SMSP) y de 0 a 140 Mg ha⁻¹ para la selva mediana subcaducifolia (SMSC). Los resultados representan una primera aproximación a la evaluación de incertidumbre a diferentes niveles de estimación de biomasa aérea.

Palabras clave: Intervalos de predicción, error, bosque tropical, Quantile Regression Forests

ABSTRACT

The estimation of forest above-ground biomass and its spatial distribution in large areas, as well as, the associated uncertainty allows us to understand the behavior of the carbon cycle and offer guidelines to climate change policies. The combined use of field data, LiDAR and spectral data has proven to be of great help in estimating above-ground biomass of forest areas from a plot level to a regional level. Unfortunately, errors (uncertainty) are generated during the modeling process that need to be evaluated and reported to users. This study evaluated the uncertainty, at the pixel level, associated with the estimation of above-ground biomass in tropical forests using Quantile Regression Forests (QRF) algorithm that calculates prediction intervals. Prediction intervals ranging from 0 to 250 Mg ha⁻¹ for semi-evergreen tropical forest (SETF) and 0 to 140 Mg ha⁻¹ for semi-deciduous tropical forest (SDTF) are reported.

Use data from different sources: field measurements, LiDAR and spectral data, as well as, the measurement and calibration process are important components that add error to the estimates of above-ground biomass. This approach provides information on the quantification of associated uncertainty with above-ground biomass in two types of tropical forest. The results represent an opportunity to continue the evaluation of uncertainty at different levels of estimation of above-ground biomass. The results represent a first approach to the evaluation of uncertainty at different levels of estimation of above-ground biomass.

Key words: LiDAR, Prediction intervals, error, tropical forest, Quantum regression forest

3.1 INTRODUCCIÓN

Los bosques tropicales son de particular interés para la conservación debido a la alta biodiversidad que contienen, además del papel relevante que juegan dentro del ciclo global del carbono y recientemente en las políticas de mitigación al cambio climático (Meyer *et al.*, 2013).

Realizar estimaciones precisas de biomasa es fundamental para la reducción de las incertidumbres en el ciclo global del carbono (Hese *et al.*, 2005); sin embargo, el nivel de incertidumbre asociada a las estimaciones de los flujos y almacenes de carbono a escalas regionales y nacionales permanece relativamente alto (Ángeles-Pérez *et al.*, 2011).

Para reducir la incertidumbre en las estimaciones de biomasa, es necesario procurar la realización de observaciones espacialmente continuas que sean lo suficientemente finas en capturar la variabilidad de la superficie, por lo que apoyarse en las herramientas de percepción remota es esencial para alcanzar dichos objetivos (Hansen *et al.*, 2014; Saarela *et al.*, 2015).

La precisión de las estimaciones finales de biomasa aérea está en función de un conjunto de errores potenciales provenientes del tipo de variables analizadas, las mediciones, el procesamiento, así como, los enfoques estadísticos empleados (Hill *et al.*, 2013; Barbosa *et al.*, 2014). De manera particular, los diferentes componentes de error que afectan las estimaciones de biomasa aérea están relacionadas a las mediciones de los árboles, el uso incorrecto de las ecuaciones alométricas, la selección inadecuada de las variables para construir los modelos, el muestreo sesgado, la concordancia y tamaño de las parcelas, la calibración de los datos de percepción remota, así como la representatividad de sitios en el área de estudio (Lu *et al.*, 2012; J. Chave *et al.*, 2014).

Diversos autores reconocen los problemas que provocan las diferentes fuentes de error. Chave *et al.* (2004) reportaron 16 % de error en la medición de árboles con DN \geq 10 cm. Los autores

señalan que el error está presente en la medición del diámetro normal, la altura y el valor de la densidad de la madera de cada especie. Cada medición conlleva ciertas dificultades sobre todo en especies tropicales. Medir el diámetro normal puede verse afectado por las formas irregulares del fuste; obtener una medida confiable de la altura se dificulta por la condición cerrada de los estratos; y no siempre es posible contar con el valor de densidad de la madera debido a la gran cantidad de especies presentes en estos bosques (Couteron *et al.*, 2012). Aunque en este caso, un valor promedio por tipo de bosque, familia o género puede ser útil (Barbosa *et al.*, 2014).

En cuanto a los modelos alométricos es preferible usar ecuaciones construidas para las especies de la zona de estudio, que incluyan las variables altura y densidad de la madera para evitar agregar un porcentaje de error importante (Chave *et al.*, 2004). Por su parte, Svob *et al.* (2014) señalan que la elección del modelo alométrico debe basarse tanto en el objetivo de la investigación, como en las características mismas de la base de datos, siendo representativo del rango diamétrico y ecosistema estudiado. En este sentido, Chave *et al.* (2004) reportaron errores atribuibles a la ecuación alométrica, 10 a 15% para sitios de 0.25 ha y 5 a 10% para sitios de 1 ha. Mientras que, Hunter *et al.* (2013) reportaron 16% de error en la estimación de biomasa por árbol y disminuye (5 a 6%) cuando se incrementa el tamaño del sitio (1 ha), a consecuencia de la medición de altura.

Otra fuente de error importante, que señala Chave *et al.* (2004), está asociada a la representatividad a escala de paisaje. Esta situación debe ser evaluada a través del establecimiento de una red de parcelas aleatoriamente distribuidas en la zona para capturar la variabilidad de los tipos de bosque. Sin embargo, esto significaría un costo económico importante. El proceso de obtener información suficiente y de alta calidad en los ecosistemas tropicales es difícil y costosa (Goetz *et al.*, 2015). Además, las características y factores

estacionales adversos de los bosques tropicales, como inundaciones o huracanes durante la temporada de lluvias, pueden comprometer las mediciones recolectadas o la cantidad de sitios muestreados. Es por ello que para realizar estimaciones de biomasa aérea a nivel regional, el uso combinado de datos de percepción remota, parece ser lo más adecuado, sin invertir el tiempo y el dinero que implica realizar el inventario forestal sobre una gran superficie (Chi *et al.*, 2017).

En este sentido, Saarela *et al.* (2015) indican que el nivel de incertidumbre de las estimaciones, aunado a lo que se señaló anteriormente, está ligado al tipo de datos espectrales y a los algoritmos de procesamiento. Por ejemplo, cuando se usan imágenes multitemporales, las condiciones atmosféricas influyen en los datos ópticos, la vegetación presenta diferencias fenológicas, la cantidad de hojas en el dosel varía, así como, el crecimiento de la vegetación del sotobosque, por lo que los análisis de pre-procesamiento deberían corregir tales diferencias.

Es por ello que cuando se emplean datos provenientes de diversas fuentes (datos de campo, de sensores remotos y ecuaciones alométricas) para estimar la biomasa aérea en grandes extensiones es necesario evaluar la incertidumbre como parte esencial del proceso (Gonçalves *et al.*, 2017).

Con respecto a la discrepancia espacial entre datos de sitios medidos en campo y aquellos extraídos por sensores remotos, Asner *et al.* (2009) reportaron un error de 0 a 10 Mg ha⁻¹ (0 - 3.5% de la biomasa media) en bosques tropicales de Hawaii, a causa de la discordancia entre los datos de campo y LiDAR. La precisión de las estimaciones finales de biomasa aérea también se ve afectada por el tamaño mínimo de parcela. Si el tamaño de la parcela es pequeño no se capturaría la variabilidad adecuada de la variable de interés. El uso de parcelas de muestreo mayores que 0.1 ha son consideradas más confiables en la estimación de biomasa, a diferencia de las parcelas pequeñas que tienden a sobreestimar la biomasa aérea por hectárea. Cuanto mayor sea la escala espacial en los sitios de muestreo, mayor será la probabilidad de capturar la

heterogeneidad estructural del bosque (Meyer *et al.*, 2013; Barbosa *et al.*, 2014; Hernández-Stefanoni *et al.*, 2014).

Ahora bien, la precisión de las estimaciones de biomasa se expresa en la mayoría de los casos reportando el valor RMSE (root mean square error – raíz del error cuadrático medio) (Chen *et al.*, 2016). Sin embargo, dicha precisión varía espacialmente conforme a la heterogeneidad del paisaje, la densidad de datos muestreados, así como, las características de los sensores remotos empleados (Zolkos *et al.*, 2013). Por lo que es necesario conocer la incertidumbre a nivel de píxeles, para tener en cuenta su variación espacial (Lu *et al.*, 2012).

Olofsson *et al.* (2013) apunta que la incertidumbre de una estimación puede ser expresada mediante el cálculo del error estándar o reportando un intervalo de confianza. Este último proporciona un rango de valores para un parámetro, teniendo en cuenta la incertidumbre de la estimación basada en la muestra. Barbosa *et al.* (2014) aclara que en modelamiento, la incertidumbre se refiere a aquellos valores reales de la variable de interés que no se conocen y puede ser estimada por el rango y probabilidad de valores posibles.

Fu *et al.* (2017) señalan que existen al menos tres metodologías para cuantificar incertidumbre.

1) Enfoque de propagación de errores que suma las varianzas de los componentes de error a través de las aproximaciones de primer orden de series de Taylor (Chave *et al.*, 2004; Gonçalves *et al.*, 2017). 2) Basado en el análisis del modelo, que estima de forma aislada los errores asociados al muestreo y al modelo mismo (Chen *et al.*, 2015), y 3) A través de técnicas de simulación como el método Monte Carlo, el cual usa miles de iteraciones para obtener la distribución de probabilidad y estimaciones finales de los parámetros del modelo (Coulston *et al.*, 2016). Este último enfoque puede estimar y evaluar las incertidumbres causadas por fuentes de error múltiple que se propagan a través de modelos complejos (Kauranne *et al.*, 2017).

Coulston *et al.* (2016) subraya que en el modelamiento a través del algoritmo RF no existe una cuantificación directa del error de predicción, como en el caso de los enfoques paramétricos tradicionales como regresión. En su trabajo, los autores realizan una evaluación de incertidumbre en torno a una predicción futura para una nueva observación a través de intervalos de predicción. Los intervalos de predicción son análogos a los intervalos de confianza, aunque los primeros son más amplios. La diferencia es que los intervalos de confianza dependen de la estimación a partir de una muestra (y difieren con cada muestra) que cubre el parámetro verdadero con una probabilidad de $1 - \alpha$; mientras que los intervalos de predicción cubren una observación futura de una variable aleatoria, teniendo dos fuentes de incertidumbre: debido a la estimación de parámetros desconocidos y debida a la aleatoriedad supuesta (Mayr *et al.*, 2012).

Con base en lo anterior se plantearon los objetivos de este trabajo: evaluar la incertidumbre a nivel pixel, como intervalos de predicción, de la estimación de biomasa aérea en dos tipos de selva mediana en la Península de Yucatán. Este procedimiento se realizó con el algoritmo Quantile Regression Forests (QRF). Cabe aclarar que esta evaluación fue posible debido a que se emplearon estimaciones de biomasa aérea sobre franjas, derivadas de datos LiDAR, con el propósito de aumentar la muestra de campo original y posteriormente empleando variables espectrales y climáticas se realizó la estimación de biomasa aérea sobre toda el área de estudio.

3.2 MATERIALES Y MÉTODOS

Área de estudio

Este estudio se llevó a cabo en dos tipos de selvas medianas: SMSP y SMSC de la Península de Yucatán, México, localizadas en la porción sureste de la República Mexicana ($17^{\circ}00'$ y $21^{\circ}45'$ N, $86^{\circ}30'$ y $9^{\circ}30'$ O). La superficie aproximada de la península es de 141 523 km², equivalente

al 7% del territorio nacional, de los cuales, 28 171 km² corresponden a SMSC y 48 699 km² a la SMSP (Instituto Nacional de Estadística y Geografía [INEGI], 2013).

Se utilizaron las estimaciones de biomasa aérea que se obtuvieron mediante el algoritmo Random Forest para los dos tipos de selva evaluados Figura 3.1.

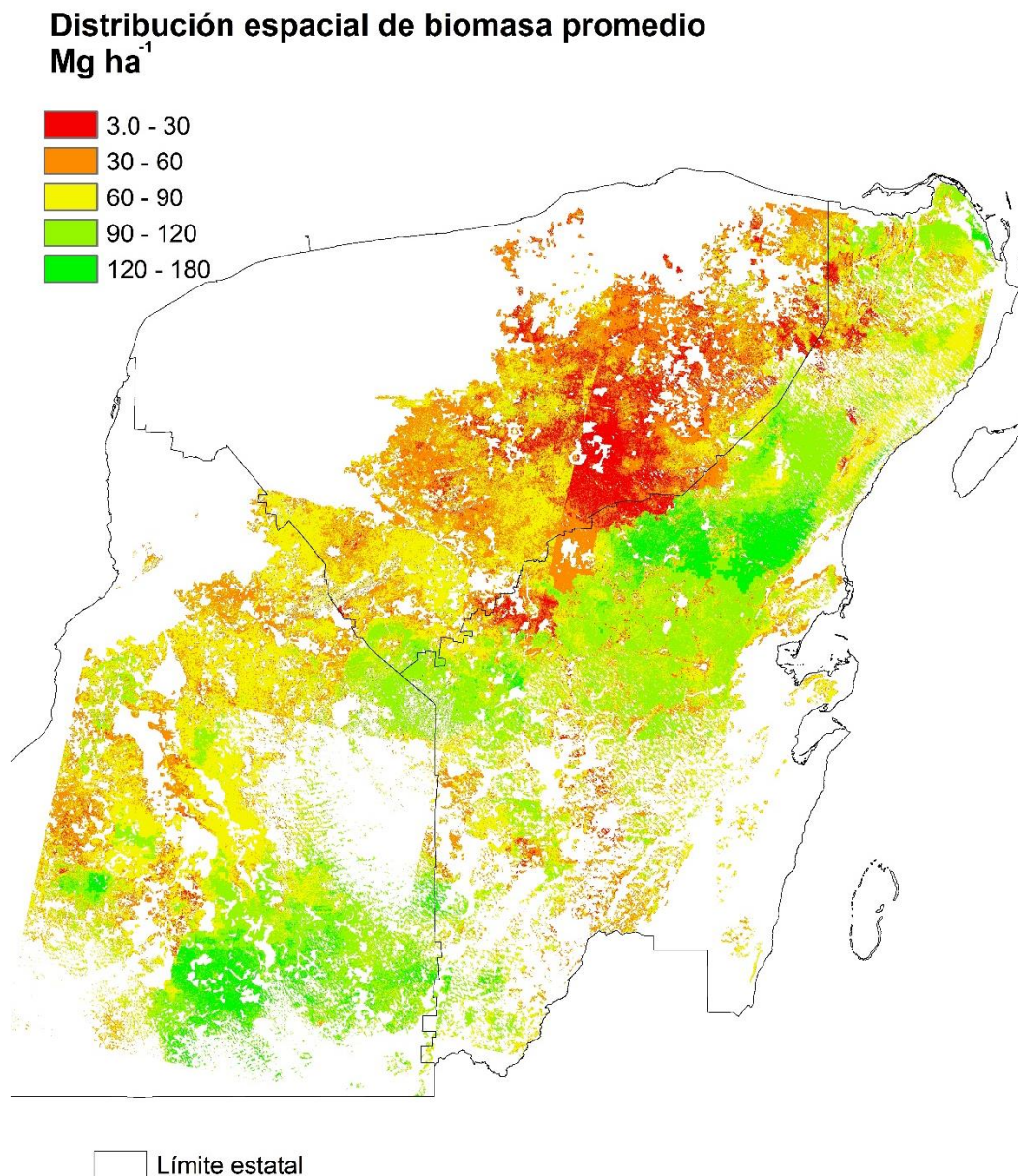


Figura 3.1 Distribución espacial de biomasa aérea en la selva mediana subcaducifolia y mediana subperennifolia de la Península de Yucatán, México.

La incertidumbre de los valores de biomasa aérea calculados se estimó a nivel pixel, como intervalos de predicción, mediante el algoritmo Quantile Regression Forests (Meinshausen, 2006), implementado en el paquete R (Meinshausen, 2017). Este procedimiento es capaz de inferir cuantiles condicionales de diversas variables predictoras. La regresión cuantil es un método que estima relaciones funcionales entre variables para todas las porciones de una distribución de probabilidad y es útil para construir intervalos de predicción. (Cade & Noon, 2003). Los cuantiles proporcionan información más completa sobre la distribución de la variable de respuesta (Y), como una función de la variable predictora (X). El ancho de esta predicción puede ser más amplio o más estrecho dependiendo de X, por lo que algunas nuevas observaciones pueden ser predichas con más precisión que otras. Los detalles matemáticos del procedimiento están descritos en (Meinshausen, 2006), y su implementación en R en (Meinshausen, 2017).

Con Quantile Regression Forests (QRF), es posible obtener dar un intervalo de predicción que cubrirá la nueva observación de la variable de respuesta, en este caso, la biomasa aérea calculada sobre en el área de estudio. Este enfoque se ha utilizado en estimación de biomasa aérea en bosques templados (Soriano-Luna *et al.*, 2018) y contenido de carbono en el suelo (Guevara *et al.*, 2018).

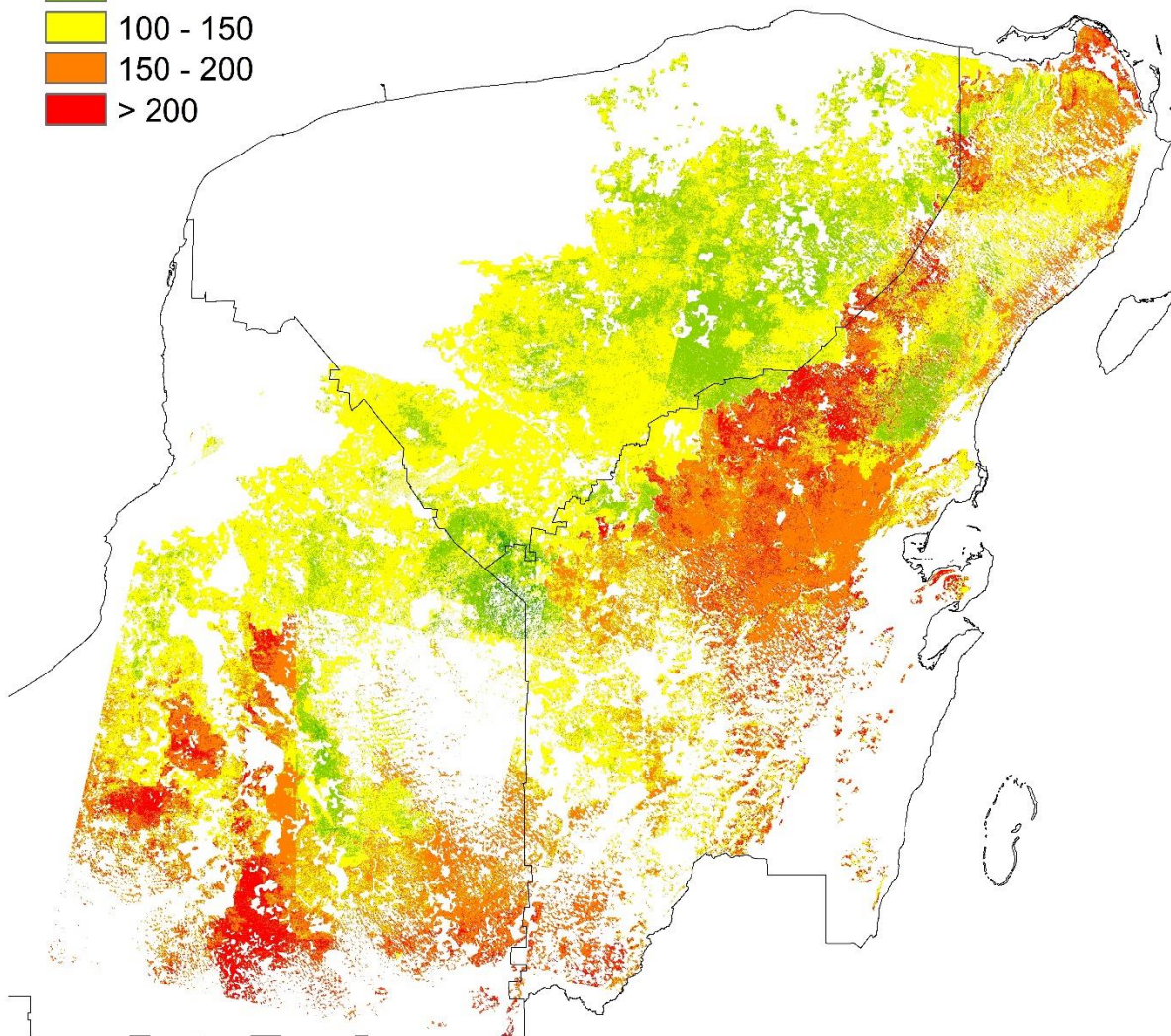
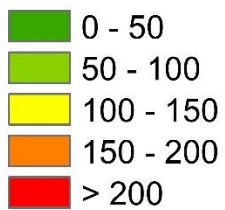
3.3 Resultados

Mapas de incertidumbre de biomasa aérea en los dos tipos de selva mediana

La incertidumbre predicha, expresada en valores absolutos de biomasa aérea (Mg ha^{-1}), se observa en la Figura 3.2. La incertidumbre predicha expresada en porcentaje fue calculada como la semiamplitud de intervalo de predicción dividido por su media (Urbazaev *et al.*, 2018), Figura

3.3. Los valores más altos de incertidumbre están asociados con valores altos de biomasa aérea tanto en la SMSP como la SMSC (Figuras 3.2, 3.3).

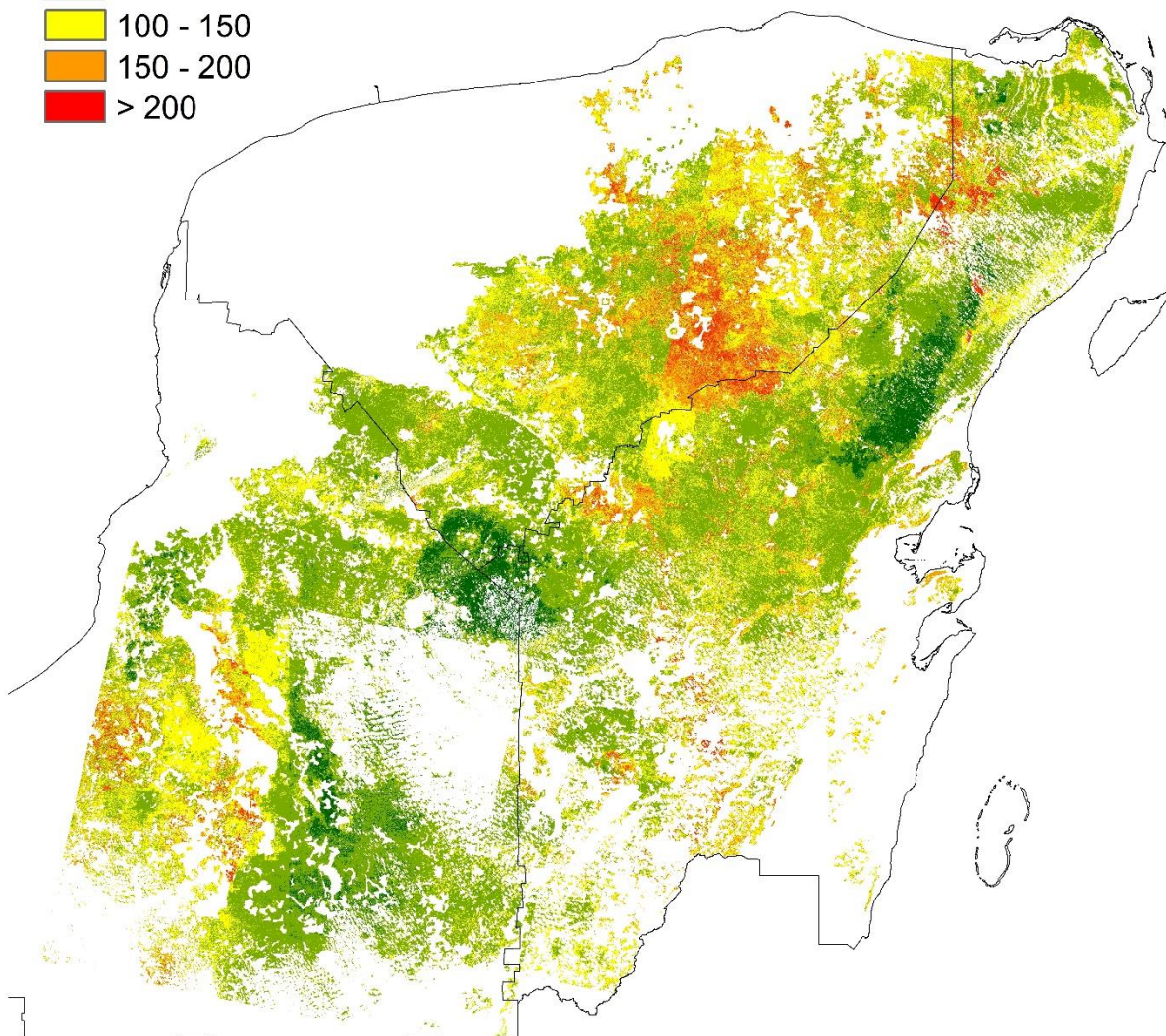
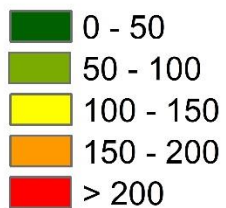
Distribución de incertidumbre de biomasa aérea (Mg ha⁻¹)



□ Límite estatal

Figura 3.2 Distribución de la incertidumbre de biomasa aérea (Mg ha⁻¹) en los dos tipos de selva mediana evaluados. La incertidumbre de las estimaciones se expresó como intervalos de predicción al 95%.

Distribución de incertidumbre de biomasa aérea en porcentaje (%)



□ Límite estatal

Figura 3.3 Distribución de la incertidumbre de biomasa aérea (%) expresado como la semiamplitud de los intervalos de predicción al 95% dividido entre su media en los dos tipos de selva mediana evaluados.

3.4 DISCUSIÓN

Como se observa en la Figura 3.2, la incertidumbre promedio fue de 162 Mg ha⁻¹ para la SMSP y de 107 Mg ha⁻¹ para la SMSC. Mientras que en la Figura 3.3 se presenta en términos de porcentaje. Los resultados representan valores altos de incertidumbre a diferencia de lo reportados para este tipo de vegetación en otros estudios (Asner *et al.*, 2009; Cartus *et al.*, 2014; Urbazaev *et al.*, 2018).

Gonçalves *et al.* (2017) señalan que es difícil realizar comparaciones directas entre los estudios debido a las diferencias en composición florística, estructura de la vegetación o el tamaño de la parcela, así como, el método de análisis para reportar la incertidumbre de estimaciones de biomasa aérea. Lo ideal sería cuantificar los errores asociados a las diferentes etapas de la estimación y escalamiento de la biomasa aérea como lo señala (Chave *et al.*, 2014).

La mayoría de estudios que realizan un análisis de incertidumbre en la estimación de biomasa aérea, lo hacen a través del enfoque de propagación de errores; ya sea analizando uno o dos componentes de error como el trabajo de Su *et al.* (2016), que evaluaron la incertidumbre a nivel de pixel considerando el error de localización del sitio. Calculado como la desviación estándar de 100 iteraciones de estimaciones de biomasa aérea, lograron incrementar la precisión del modelo de 0.64 a 0.75 y disminuir el RMSE de 50 a 42 Mg ha⁻¹. El valor absoluto de la incertidumbre incrementó con la densidad de la biomasa forestal. El error evaluado contribuyó en 10% a la estimación final de biomasa aérea a nivel nacional en China.

Aunque en este estudio no se cuantificó la incertidumbre por componente de error, la discordancia, tanto espacial como temporal, entre parcelas en campo y los datos registrados de LiDAR son fuentes importantes de error que influyeron en los resultados presentados. Gonçalves *et al.* (2017) señalan que si estos componentes de error fueran cero, el error total de su estudio

(20% – 34%) disminuiría a casi la mitad, mientras que, si se redujera los errores debido a la medición y modelos alométricos, su error total disminuiría solo en 20%.

En este trabajo, se mencionó que la calibración de los modelos para estimar la biomasa aérea en franjas de datos LiDAR se realizó con datos de sitios del INFyS, los cuales tenían una forma rectangular y con base en ello, se realizó el recorte de la nube de puntos de LiDAR. En principio, se ha sugerido utilizar sitios circulares, para correlacionar la información debido a que disminuye el error a consecuencia del efecto de borde (Frazer *et al.*, 2011). Además, la geolocalización de los sitios de muestreo del INFyS tienen un error promedio de 5.9 m y se encontraron superposiciones entre algunos sitios, por lo que se reconstruyeron los polígonos tomando en cuenta las medidas de diseño para el establecimiento de las parcelas en campo (CONAFOR, 2011). Lo anterior afectó los resultados a lo largo del proceso de estimación de biomasa aérea en las dos selvas medianas evaluadas.

Así mismo, aunque los píxeles de las franjas (con la estimación de biomasa aérea) y los píxeles de datos auxiliares (Landsat y climáticas) correspondían en forma y tamaño (30 m); en la primera parte del proceso de calibración, con datos de sitios del INFyS, las formas (rectangulares) no correspondían a los píxeles de LiDAR. Por lo tanto, la incertidumbre asociada a la forma y tamaño del píxel, son componentes de error que también deber ser evaluados (Zhang *et al.*, 2014).

Considerando el tamaño del píxel, autores como Mascaro *et al.* (2011) y Cartus *et al.* (2014), han encontrado mejoras en la evaluación de incertidumbre a medida que el tamaño de píxel es mayor. Asner *et al.* (2010) señalaron que el error de carbono predicho debería disminuir a medida que disminuye la resolución espacial, porque los errores totales aumentan a una tasa más lenta que el carbono total; los errores relativos se propagan en función de (área de la parcela)^{-1/2}. Asner *et*

al. (2009) reportaron errores de 0-10 Mg ha⁻¹ de biomasa aérea cuando evaluaron la discordancia espacial entre datos de campo y LiDAR. Obtuvieron errores menores (78 - 90 Mg ha⁻¹) en pixeles con 1 ha de resolución y con menor concentración de biomasa media. Lo anterior sugiere que a menor concentración de biomasa, menor incertidumbre (Köhler & Huth, 2010; Zhang *et al.*, 2014). Por su parte, Mascaro *et al.* (2011) señalan que es posible lograr un error de 10% al mapear carbono aéreo, en bosques tropicales, a partir de datos LiDAR si se emplea 1 ha de resolución espacial. El porcentaje de error de su estudio es similar al reportado cuando solo se emplea datos de sitios en campo.

Sin embargo, otros autores sugieren que es necesario tener en cuenta la correlación de errores entre ubicaciones vecinas, puesto que, cuando se escalan errores pequeños (calculadas a nivel de parcela, como en el caso de los estudios arriba mencionados), a escalas nacionales o regionales pueden llegar a ser poco realistas (Saatchi *et al.*, 2011). Por ejemplo, González *et al.* (2010) calcularon un error medio a nivel de pixel de 24% y 18% cuando estimaron densidades de carbono forestal en bosques secundarios y maduros de *Sequoia sempervirens* en dos áreas de investigación en California. Tales porcentajes se redujeron hasta en 1% cuando la incertidumbre se calculó al nivel de paisaje (59 km²). Sin embargo, este grado de incertidumbre se logró debido al promediado espacial, es decir, errores espacialmente no correlacionados se cancelan entre sí y errores correlacionados no se cancelan, cuando se conjunta para toda la escala de paisaje (Saatchi *et al.*, 2011).

Lo mismo sucede cuando se evalúa la incertidumbre a escala global (tamaño del pixel 1km). Se ha convenido una precisión de ± 20 Mg ha⁻¹ o 20% para estimaciones de biomasa aérea a este nivel (Hall *et al.*, 2011). El estudio de Montesano *et al.* (2013) cumple con dichos requerimientos

al reducir los errores cuando combinaron SAR y datos ópticos pasivos ($RMSE = 11 \text{ Mg ha}^{-1}$) y donde más del 80% de los píxeles registraron dicha precisión.

En este trabajo la incertidumbre se evaluó a nivel de pixel, que varía desde 0 hasta 250 Mg ha^{-1} para la SMSP y de $0 - 140 \text{ Mg ha}^{-1}$ para la SMSC, sin embargo, no se realizó el análisis mediante un enfoque de propagación de errores como en los estudios ya revisados. Es complicado hacer una comparación directa con los resultados de otros trabajos porque las metodologías de evaluación de incertidumbre son diferentes. Lo que se reporta en este trabajo es la incertidumbre como intervalos de predicción (análogos a los intervalos de confianza), que permite ver de forma espacialmente explícita la incertidumbre asociada a las estimaciones de biomasa aérea para dos selvas medianas de la península de Yucatán a una resolución de 30 m.

No obstante, existen reportes de incertidumbre para esta zona, aunque a una menor resolución espacial (1 ha). Urbazaev *et al.* (2018) evalúan la biomasa aérea en México integrando fuentes de datos similares a este trabajo (INFyS, LiDAR y Landsat) y agregan datos de SAR. Es un enfoque similar de estimación de biomasa aérea a partir de datos LiDAR en franjas. Reportaron incertidumbres $<200\%$ (media = 44.8%) cuando estiman la biomasa aérea para todo el país, sin embargo, seleccionaron solo los píxeles que tienen menos del 50% de incertidumbre en la primera fase del modelamiento. También estiman la incertidumbre como intervalos de confianza, aunque los autores utilizan la simulación Monte Carlo. En este estudio, se seleccionaron aleatoriamente una muestra de píxeles, sin ninguna restricción, lo que incluye una mayor variación que se reflejó en los resultados de incertidumbre reportados (Figura 3.2).

Por lo tanto, el enfoque presentado en este trabajo supone que las variables seleccionadas, los datos muestreados y el algoritmo seleccionado brinda información sobre la biomasa aérea que se distribuye en los dos tipos de vegetación evaluados y la cuantificación de incertidumbre.

En el trabajo de Soriano-Luna *et al.* (2018) reportaron valores de incertidumbre amplios (0-200 Mg ha⁻¹) en biomasa aérea de bosques templados bajo manejo, empleando el mismo método que en este estudio. Los autores indican que la incertidumbre es afectada por el método de predicción utilizado y lograron mejores resultados cuando emplearon un método semi paramétrico, seguido del algoritmo QRF y regresión lineal.

Por otro lado, es evidente la diferencia entre la incertidumbre reportada por tipo de vegetación. Al igual que el trabajo de Zhang *et al.* (2014), las estimaciones de incertidumbre están restringidas por el tipo de vegetación que se evalúa. En este estudio, los valores absolutos más altos de incertidumbre se presentaron en la SMSP (0 - 250 Mg ha⁻¹). Las diferentes características entre tipos de vegetación como el clima y la precipitación, afectan los resultados de estimación de biomasa aérea y por ende, la incertidumbre reportada para la SMSC y SMSP.

Como se ha mencionado anteriormente, existen una serie de componentes que agregan cierto porcentaje de error al momento de estimar biomasa aérea a través de diferentes escalas. La discordancia espacial y temporal, entre los datos obtenidos en campo y de sensores remotos, afectan en gran proporción las predicciones finales de biomasa aérea (Brososke *et al.*, 2014). El primero, a consecuencia de los errores de geoposicionamiento, se intensifica en áreas densas y de estructura compleja, como los bosques tropicales, en donde la señal del GPS es débil, provocando incompatibilidad entre las fuentes de datos (imágenes satelitales, datos de LiDAR y parcelas en campo). Para el segundo, referente a la discordancia temporal, Gonçalves *et al.* (2017) sugieren utilizar modelos de crecimiento para biomasa y cuantificar el error asociado con el modelo. Además los autores recomiendan reducir los errores debido a estos componentes, más que al error asociado por mediciones en campo y alométricos, como una alternativa factible cuando los recursos son limitados.

Referente a la elección del algoritmo Random Forest para la estimación de biomasa aérea. Este realiza un proceso de promediado lo que crea resultados sesgados hacia la media de la muestra, por lo que valores pequeños o grandes de biomasa aérea comúnmente son sub o sobre estimados (Xu *et al.*, 2016). Además el valor de referencia fue calculado solo considerando los árboles medidos dentro del sitio INFyS (400 m²) ≥ 7.5 cm de diámetro normal, eliminando árboles pequeños y vegetación menor. Así mismo, aunque RF puede tratar con muchas variables y no es afectado por la inclusión de variables redundantes en el mismo modelo de predicción, elegir adecuadamente las variables útiles determinará el rendimiento del modelo, las predicciones realizadas y por lo tanto, la incertidumbre estimada (Brososke *et al.*, 2014).

Por otro lado, emplear datos espectrales y productos derivados de imágenes satelitales, también conlleva un porcentaje de incertidumbre debido a los efectos de saturación espectral y defectos de la imagen. Cartus *et al.* (2014) señalan las limitaciones asociadas con el mapeo de biomasa aérea empleando un gran número de imágenes satelitales, las cuales varía entre ellas mismas y a través de ecosistemas forestales, a pesar de los diversos procedimientos de calibración radiométrica y correcciones geométricas que se emplean.

Finalmente, cuando se estima y mapea la biomasa aérea de forma espacialmente continua en áreas extensas, como este estudio, la opción factible es emplear datos de sensores remotos que necesariamente tienen que ser calibrados con mediciones en campo. A lo largo de este proceso, las estimaciones en todos los niveles son afectados por diversas fuentes de error. Estas fuentes que componen la incertidumbre general deberían explicar la distribución de las estimaciones finales de biomasa aérea (Hill *et al.*, 2013).

Reportar la incertidumbre asociada a las estimaciones de biomasa aérea a nivel regional, a través del enfoque de propagación de errores, es una oportunidad de mejora en este estudio. Se requiere

cubrir todos los niveles agregados en las que se reporten las estimaciones de biomasa aérea, en este caso: a nivel sitio, franjas de LiDAR y a nivel paisaje (tipo de vegetación).

3.5 CONCLUSIONES

La incertidumbre de las estimaciones de biomasa aérea es un aspecto importante a considerar cuando se realiza estimaciones desde un nivel local a regional, con datos de diversas fuentes y a través de diferentes técnicas de modelamiento. En este trabajo se evaluó la incertidumbre, a nivel pixel, del mapa de biomasa aérea de dos tipos de selva mediana en la Península de Yucatán mediante el uso del algoritmo QRF, que estima intervalos de predicción con 95% de confianza. Los resultados varían desde 0 hasta 250 Mg ha⁻¹ para la SMSP y de 0 – 140 Mg ha⁻¹ para la SMSC. Este trabajo ofrece una perspectiva sobre la importancia de evaluar la incertidumbre relacionada a estimaciones de biomasa aérea en bosques tropicales a nivel regional. No obstante, es necesario realizar un trabajo integral de propagación de errores para evaluar concretamente cuanto porcentaje de error agrega cada componente reconocido en los diferentes niveles de la estimación de biomasa aérea en las selvas medianas de la Península de Yucatán.

3.6 LITERATURA CITADA

- Ángeles-Pérez, G., Wayson, C., Birdsey, R., Valdez-Lazalde, J. R., De los Santos-Posadas, H. M., & Plascencia-Escalante, F. O. (2011). *Sitios de monitoreo de flujo de CO₂ a largo plazo en bosques bajo manejo en el centro de México*. Paper presented at the Memorias III Simposio Internacional del Carbono en México.
- Asner, G. P., Hughes, R. F., Varga, T. A., Knapp, D. E., & Kennedy-Bowdoin, T. (2009). Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. *Ecosystems*, *12*(2), 261-278. doi:10.1007/s10021-008-9221-5
- Asner, G. P., Powell, G. V., Mascaró, J., Knapp, D. E., Clark, J. K., Jacobson, J., . . . Hughes, R. F. (2010). High-resolution forest carbon stocks and emissions in the Amazon. *Proc Natl Acad Sci U S A*, *107*(38), 16738-16742. doi:10.1073/pnas.1004875107

- Barbosa, J. M., Broadbent, E. N., & Bitencourt, M. D. (2014). Remote Sensing of Aboveground Biomass in Tropical Secondary Forests: A Review. *International Journal of Forestry Research*, 2014, 1-14. doi:10.1155/2014/715796
- Broszofske, K. D., Froese, R. E., Falkowski, M. J., & Banskota, A. (2014). A Review of Methods for Mapping and Prediction of Inventory Attributes for Operational Forest Management. *Forest Science*, 60(4), 733-756. doi:10.5849/forsci.12-134
- Cade, B. S., & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1(8), 412-420.
- Cartus, O., Kellndorfer, J., Walker, W., Franco, C., Bishop, J., Santos, L., & Fuentes, J. (2014). A National, Detailed Map of Forest Aboveground Carbon Stocks in Mexico. *Remote Sensing*, 6(6), 5559-5588. doi:10.3390/rs6065559
- Chave, Condit, R., Aguilar, S., Hernandez, A., Lao, S., & Perez, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1443), 409-420. doi:10.1098/rstb.2003.1425
- Chave, J., Rejou-Mechain, M., Burquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., . . . Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20(10), 3177-3190. doi:10.1111/gcb.12629
- Chen, Q., McRoberts, R. E., Wang, C., & Radtke, P. J. (2016). Forest aboveground biomass mapping and estimation across multiple spatial scales using model-based inference. *Remote Sensing of Environment*, 184, 350-360. doi:10.1016/j.rse.2016.07.023
- Chen, Q., Vaglio Laurin, G., & Valentini, R. (2015). Uncertainty of remotely sensed aboveground biomass over an African tropical forest: Propagating errors from trees to plots to pixels. *Remote Sensing of Environment*, 160, 134-143. doi:10.1016/j.rse.2015.01.009
- CONAFOR. (2011). *Manual y procedimientos para el muestreo de campo. Remuestreo 2012. Inventario Nacional Forestal y de Suelos*. Zapopan, Jalisco, México.: Comisión Nacional Forestal. Secretaría del Medio Ambiente y Recursos Naturales.
- Coulston, J. W., Blinn, C. E., Thomas, V. A., & Wynne, R. H. (2016). Approximating Prediction Uncertainty for Random Forest Regression Models. *Photogrammetric Engineering & Remote Sensing*, 82(3), 189-197. doi:10.14358/pers.82.3.189
- Couteron, P., Barbier, N., Proisy, C., Péliissier, R., & Vincent, G. (2012). Linking remote-sensing information to tropical forest structure: The crucial role of modelling. *Earthzine*, 4, 1-4.
- Frazer, G. W., Magnussen, S., Wulder, M. A., & Niemann, K. O. (2011). Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. *Remote Sensing of Environment*, 115(2), 636-649. doi:10.1016/j.rse.2010.10.008

- Fu, Y., Lei, Y., Zeng, W., Hao, R., Zhang, G., Zhong, Q., & Xu, M. (2017). Uncertainty assessment in aboveground biomass estimation at the regional scale using a new method considering both sampling error and model error. *Canadian Journal of Forest Research*, 47(8), 1095-1103. doi:10.1139/cjfr-2016-0436
- Goetz, S. J., Hansen, M., Houghton, R. A., Walker, W., Laporte, N., & Busch, J. (2015). Measurement and monitoring needs, capabilities and potential for addressing reduced emissions from deforestation and forest degradation under REDD+. *Environmental Research Letters*, 10(12). doi:10.1088/1748-9326/10/12/123001
- Gonçalves, F., Treuhaft, R., Law, B., Almeida, A., Walker, W., Baccini, A., . . . Graça, P. (2017). Estimating Aboveground Biomass in Tropical Forests: Field Methods and Error Analysis for the Calibration of Remote Sensing Observations. *Remote Sensing*, 9(12). doi:10.3390/rs9010047
- Gonzalez, P., Asner, G. P., Battles, J. J., Lefsky, M. A., Waring, K. M., & Palace, M. (2010). Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California. *Remote Sensing of Environment*, 114(7), 1561-1575. doi:10.1016/j.rse.2010.02.011
- Guevara, M., Olmedo, G. F., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., . . . Vargas, R. (2018). No silver bullet for digital soil mapping: country-specific soil organic carbon estimates across Latin America. *Soil*, 4(3), 173-193. doi:10.5194/soil-4-173-2018
- Hall, F. G., Bergen, K., Blair, J. B., Dubayah, R., Houghton, R., Hurtt, G., . . . Wickland, D. (2011). Characterizing 3D vegetation structure from space: Mission requirements. *Remote Sensing of Environment*, 115(11), 2753-2775. doi:10.1016/j.rse.2011.01.024
- Hernández-Stefanoni, J., Dupuy, J., Johnson, K., Birdsey, R., Tun-Dzul, F., Peduzzi, A., . . . López-Merlín, D. (2014). Improving Species Diversity and Biomass Estimates of Tropical Dry Forests Using Airborne LiDAR. *Remote Sensing*, 6(12), 4741-4763. doi:10.3390/rs6064741
- Hese, S., Lucht, W., Schmillius, C., Barnsley, M., Dubayah, R., Knorr, D., . . . Schröter, K. (2005). Global biomass mapping for an improved understanding of the CO₂ balance—the Earth observation mission Carbon-3D. *Remote Sensing of Environment*, 94(1), 94-104. doi:10.1016/j.rse.2004.09.006
- Hill, T. C., Williams, M., Bloom, A. A., Mitchard, E. T., & Ryan, C. M. (2013). Are inventory based and remotely sensed above-ground biomass estimates consistent? *PLoS One*, 8(9), e74170. doi:10.1371/journal.pone.0074170
- Hunter, M. O., Keller, M., Victoria, D., & Morton, D. C. (2013). Tree height and tropical forest biomass estimation. *Biogeosciences*, 10(12), 8385-8399. doi:10.5194/bg-10-8385-2013

- Kauranne, T., Joshi, A., Gautam, B., Manandhar, U., Nepal, S., Peuhkurinen, J., . . . Leppänen, V. (2017). LiDAR-Assisted Multi-Source Program (LAMP) for Measuring Above Ground Biomass and Forest Carbon. *Remote Sensing*, 9(2). doi:10.3390/rs9020154
- Köhler, P., & Huth, A. (2010). Towards ground-truthing of spaceborne estimates of above-ground life biomass and leaf area index in tropical rain forests. *Biogeosciences*, 7(8), 2531-2543. doi:10.5194/bg-7-2531-2010
- Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., . . . Saah, D. (2012). Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *International Journal of Forestry Research*, 2012, 1-16. doi:10.1155/2012/436537
- Mascaro, Detto, M., Asner, G. P., & Muller-Landau, H. C. (2011). Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sensing of Environment*, 115(12), 3770-3774. doi:10.1016/j.rse.2011.07.019
- Mayr, A., Hothorn, T., & Fenske, N. (2012). Prediction intervals for future BMI values of individual children-a non-parametric approach by quantile boosting. *BMC Medical Research Methodology*, 12(1), 6.
- Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7, 983-999.
- Meinshausen, N. (2017). quantregForest: Quantile Regression Forests. Retrieved from <http://github.com/lorismichel/quantregForest>
- Meyer, V., Saatchi, S. S., Chave, J., Dalling, J. W., Bohlman, S., Fricker, G. A., . . . Hubbell, S. (2013). Detecting tropical forest biomass dynamics from repeated airborne lidar measurements. *Biogeosciences*, 10(8), 5421-5438. doi:10.5194/bg-10-5421-2013
- Montesano, P. M., Cook, B. D., Sun, G., Simard, M., Nelson, R. F., Ranson, K. J., . . . Luthcke, S. (2013). Achieving accuracy requirements for forest biomass mapping: A spaceborne data fusion method for estimating forest biomass and LiDAR sampling error. *Remote Sensing of Environment*, 130, 153-170. doi:10.1016/j.rse.2012.11.016
- Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122-131. doi:10.1016/j.rse.2012.10.031
- Saarela, S., Grafström, A., Ståhl, G., Kangas, A., Holopainen, M., Tuominen, S., . . . Hyyppä, J. (2015). Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information. *Remote Sensing of Environment*, 158, 431-440. doi:10.1016/j.rse.2014.11.020
- Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., . . . Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three

- continents. *Proc Natl Acad Sci U S A*, 108(24), 9899-9904. doi:10.1073/pnas.1019576108
- Soriano-Luna, M., Ángeles-Pérez, G., Guevara, M., Birdsey, R., Pan, Y., Vaquera-Huerta, H., . . . Vargas, R. (2018). Determinants of Above-Ground Biomass and Its Spatial Variability in a Temperate Forest Managed for Timber Production. *Forests*, 9(8). doi:10.3390/f9080490
- Su, Y., Guo, Q., Xue, B., Hu, T., Alvarez, O., Tao, S., & Fang, J. (2016). Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne lidar, optical imagery, and forest inventory data. *Remote Sensing of Environment*, 173, 187-199. doi:10.1016/j.rse.2015.12.002
- Svob, S., Arroyo-Mora, J. P., & Kalacska, M. (2014). A wood density and aboveground biomass variability assessment using pre-felling inventory data in Costa Rica. *Carbon Balance and Management*, 9(1), 9. doi:10.1186/s13021-014-0009-y
- Urbazaev, M., Thiel, C., Cremer, F., Dubayah, R., Migliavacca, M., Reichstein, M., & Schmillius, C. (2018). Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. *Carbon Balance Manag*, 13(1), 5. doi:10.1186/s13021-018-0093-5
- Xu, L., Saatchi, S. S., Yang, Y., Yu, Y., & White, L. (2016). Performance of non-parametric algorithms for spatial mapping of tropical forest structure. *Carbon Balance Manag*, 11(1), 18. doi:10.1186/s13021-016-0062-9
- Zhang, G., Ganguly, S., Nemani, R. R., White, M. A., Milesi, C., Hashimoto, H., . . . Myneni, R. B. (2014). Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. *Remote Sensing of Environment*, 151, 44-56. doi:10.1016/j.rse.2014.01.025
- Zolkos, S. G., Goetz, S. J., & Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment*, 128, 289-298. doi:10.1016/j.rse.2012.10.017

CONCLUSIONES GENERALES Y COMENTARIOS FINALES

Se estimó la biomasa aérea en la selva mediana subcaducifolia y subperennifolia de la Península de Yucatán, empleando un enfoque de escalamiento local a regional. La sinergia de mediciones en campo, coincidentes con información capturada por LiDAR e imágenes Landsat, así como, los métodos estadísticos empleados, permitieron generar mapas de la distribución espacial e incertidumbre asociada de la variable de interés.

La relación entre datos de sitios de inventario y métricas derivadas de LiDAR permitió estimar y obtener un mapa de la distribución espacial de biomasa aérea en franjas. Sin embargo, ninguno de los métodos estadísticos empleados mostró superioridad con respecto al otro, tanto en la varianza explicada como en el RMSE reportado.

Se generó el mapa de distribución de biomasa aérea a nivel regional mediante al análisis que permite relacionar la biomasa aérea estimada en franjas con variables auxiliares, tanto espectrales de imágenes Landsat como climáticas, que cubren el área de estudio. Dicha distribución fue explicada en mayor proporción por las variables climáticas dada su relación a esta escala.

La evaluación de incertidumbre producto del escalamiento de las estimaciones de biomasa aérea se generó mediante una metodología que calcula intervalos de predicción. Esta evaluación permitió mostrar una aproximación inicial cuando la estimación de biomasa se realiza a diferentes niveles.

La estimación de biomasa aérea sobre áreas extensas aún presenta desafíos que enfrentar, dado que se deben optimizar las metodologías existentes o crear nuevas que permitan mejorar las estimaciones, reducir la incertidumbre asociada, a su vez que, los costos relacionados a mediciones (en campo y de sensores remotos) sea mínima.

La información derivada de LiDAR puede ser la herramienta operativa que ayude a mejorar la precisión de las estimaciones de biomasa aérea, minimizando costos, siempre que se empleen enfoques de muestreo adecuados y se apoyen en datos auxiliares registrados en superficies grandes.

Es necesario evaluar los errores potenciales que afectan las estimaciones de biomasa aérea para mejorar los protocolos y metodologías empleadas que permita minimizar el nivel de incertidumbre asociada.

La evaluación precisa de biomasa aérea en regiones extensas y de difícil acceso como los bosques tropicales será posible, siempre y cuando, la información obtenida en los diferentes niveles (local y/o regional) permita la comparación directa. Para ello se deben subsanar una serie de errores tanto espaciales como temporales en la adquisición de datos, que permita la calibración y validación adecuada de los modelos desarrollados.