SPACEBORNE LIDAR FOR ESTIMATING FOREST BIOPHYSICAL PARAMETERS

A Dissertation

by

LANA LANDRA NARINE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Sorin C. Popescu
Committee Members,	Marian Eriksson
	Anthony Filippi
	X. Ben Wu
Head of Department,	Cliff Lamb

May 2019

Major Subject: Ecosystem Science and Management

ABSTRACT

The Ice, Cloud and land Elevation Satellite-2 (ICESat-2) was launched on September 15th, 2018 and while this mission primarily serves to capture ice topography measurements of the earth's surface, it also offers a phenomenal opportunity to estimate biophysical forest parameters at multiple spatial scales. This study served to develop approaches for utilizing ICESat-2 data over vegetated areas. The main objectives were to: (1) derive a simulated ICESat-2 photon-counting lidar (PCL) vegetation product using airborne lidar data and examine the use of simulated PCL metrics for modeling AGB and canopy cover, (2) create wall-to-wall AGB maps at 30-m spatial resolution and characterize AGB uncertainty by using simulated PCL-estimated AGB and predictor variables from Landsat data and derived products, and (3) investigate deep learning (DL) neural networks for producing an AGB product with ICESat-2, using simulated PCL-estimated AGB Landsat imagery, canopy cover and land cover maps. The study was carried out in Sam Houston National Forest located in south-east Texas, using existing airborne lidar data and known ICESat-2 track locations for the first two years of the mission. Three scenarios were analyzed; 1) simulated data without the addition of noise, 2) processed simulated data for nighttime and 3) daytime scenarios. AGB model testing with no noise, nighttime and daytime scenarios resulted in R^2 values of 0.79, 0.79 and 0.63 respectively, with root mean square error (RMSE) values of 19.16 Mg/ha, 19.23 Mg/ha, and 25.35 Mg/ha. Canopy cover (4.6 m) models achieved R^2 values of 0.93, 0.75 and 0.63 and RMSE values of 6.36%, 12.33% and 15.01% for the no noise, nighttime

and daytime scenarios respectively. Random Forest (RF) and deep neural network (DNN) models used with predicted AGB estimates and the mapped predictors exhibited moderate accuracies (0.42 to 0.51) with RMSE values between 19 Mg/ha to 20 Mg/ha. Overall, findings from this study suggest the potential of ICESat-2 for estimating AGB and canopy cover and generating a wall-to-wall AGB product by adopting a combinatory approach with spectral metrics derived from Landsat optical imagery, canopy cover and land cover.

DEDICATION

To my parents, Noomon and Dhanragie Narine, my husband Aditya Tadakaluru and my daughter, Ananya, for their unconditional love and support.

ACKNOWLEDGEMENTS

I am grateful to everyone who has inspired and supported me throughout this journey and made this dissertation possible. I would like to express my sincere appreciation for my advisor, Dr. Sorin Popescu, for the opportunity to work on this exciting study and for his mentorship. His guidance and support throughout this endeavor has inspired me to always strive to be the best that I can possibly be. Dr. Popescu's invaluable advising, understanding and trust every step of the way has enriched my life both personally and professionally. I have learnt more than I ever imagined possible over the past few years and I could not have asked for a better mentor. Thank you so much! I have also been privileged to have Dr. Marian Eriksson, Dr. Anthony Filippi and Dr. Ben Wu on my advisory committee. I am truly thankful for their advising, constructive feedback and discussions. They have enriched my experience as a student and I greatly appreciate their help and support with my dissertation and the chance to learn from each of them.

My sincerest thanks go to Dr. Amy Neuenschwander, Tan Zhou, Shruthi Srinivasan and Tan Zhou, for their invaluable contributions to the study. Special thanks to professors from University of Missouri; Dr. Francisco X. Aguilar, Dr. Stephen Shifley, Dr. Charlie Nilon and Dr. Hank Stelzer, for recommending me for graduate study at Texas A&M University. I sincerely appreciate their support and contributions in shaping my education.

I am where I am because of my parents. I am lucky to have been brought up with a solid foundation that continues to shape my life. My parents have sacrificed so much to raise us and I am beyond grateful for their love and support that never ceases. Thank you to my father, who instilled the value of hard work and determination and for always believing in me and my goals. Thank you to my mother for imparting her wisdom throughout my life. This would not have been possible without your help and support. I hope that I can be as great of a mom to my daughter as you are to me. I thank my sister, Lynsey and brother, Lendel for always being by my side and my pillars of support. I am blessed to have you both in my life and to share in our joys and I am so proud of the people you both have grown to be.

I would like to thank my husband and best friend Aditya, for his encouragement, love and support. He has been on this journey with me years before I started graduate school and continues to support me along the way. Thank you for always believing in me, even when I doubt my abilities and for your love at the end of day. Finally, my dear daughter Ananya, you have made me a stronger and better person. I am beyond lucky and eternally grateful to be your mama. I love you!

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Sorin C. Popescu, Dr. Marian Eriksson, and Dr. X. Ben Wu of the Department of Ecosystem Science and Management and Dr. Anthony Filippi of the Department of Geography.

Discrete return airborne lidar data for Sam Houston National Forest in Texas used as reference data in Chapter 2 and for simulating Ice, Cloud and land Elevation Satellite-2 (ICESat-2) data was provided by Dr. Sorin Popescu (Department of Ecosystem Science and Management). Reference aboveground biomass data was provided by Shruthi Srinivasan (Texas A&M Forest Service). The photon counting lidar (PCL) simulator used to simulate the Ice, Cloud and land Elevation Satellite-2 (ICESat-2) data was developed by Dr. Amy Neuenschwander (University of Texas and Vegetation Scientist on the ICESat-2 Science Definition Team (SDT)). Track locations used for simulating ICESat-2 PCL data, were provided by Dr. Sorin C. Popescu, (Department of Ecosystem Science and Management and Vegetation Scientist on the ICESat-2 SDT). Kaitlin Harbeck (NASA Goddard Space Flight Center) contributed to simulated ICESat-2 data, with the integration of noise photons. Data processing, depicted in Chapter 2 was conducted in part by Dr. Tan Zhou, formerly from the Department of Ecosystem Science and Management and was published in 2019 in *Remote Sensing of Environment*.

Funding Sources

Research and graduate study were supported by Dr. Sorin Popescu's College of Agriculture and Life Sciences graduate research assistant funding and NASA grant (grant number NNX15AD02G).

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

The biophysical assessment of forest resources provides critical information about the functioning of ecosystems and their capacity to provide goods and services to an increasing population (MacDicken 2015; Tewari 2016). Moreover, with increasing concerns about climate change, assessments of forest resources, specifically forest aboveground biomass (AGB), has been highlighted (Le Toan et al. 2011). For this study, AGB refers to the total dry weight of the biological material of trees found above the ground (e.g. stem, foliage, branches) and expressed on per unit area (per hectare) basis (Hu et al. 2016; Jenkins et al. 2003). An estimated 50% of plant biomass is carbon (Drake et al. 2003), so assessments of AGB facilitate the measurement of forest carbon and can contribute to an improved understanding of the carbon cycle (Hall et al. 2011). There is a crucial need to reduce uncertainties associated with quantification and spatial distribution of terrestrial carbon stocks and accurate estimates of AGB or other forest parameters that relate to AGB like forest canopy heights, can meet this need (Le Toan et al. 2011; Margolis et al. 2015).

Lidar systems are active remote sensing devices that transmit energy and record backscattered energy from features on the earth's surface (Campbell and Wynne 2011). By measuring the time taken for laser energy to reach the surface and also return to the sensor and converting those times to distance measurements, three-dimensional biophysical attributes become obtainable (Gwenzi and Lefsky 2014; Popescu 2007). The data acquired by the lidar sensor can then be analyzed to derive vegetation metrics (e.g.

Margolis et al. 2015; Neigh et al. 2013; Nelson et al. 2017; Popescu et al. 2011). Spaceborne light detection and ranging (lidar) technology developed in this decade will offer an innovative and promising platform for the characterization and monitoring of AGB and forest carbon (Markus et al. 2017). With the launch of NASA's Ice, Cloud and land Elevation Satellite-2 (ICESat-2) on September 15th, 2018 (NASA 2017), data from its laser altimeter will be available to the scientific community as early as the spring of 2019 (Neuenschwander and Pitts 2019). As a space-based remote sensing tool, ICESat-2 offers powerful capabilities of estimating forest attributes from local to global spatial scales. To exemplify, the precursor to ICESat-2, ICESat, operated from 2003 to 2010 and the data collected by its sensor, the Geoscience Laser Altimeter System (GLAS), was used in estimating forest heights (Baghdadi et al. 2014; Harding and Carabajal 2005; Lefsky et al. 2007) and AGB (Lefsky et al. 2005; Nelson et al. 2017) as well as used for the development of global maps of AGB (Hu et al. 2016) and forest canopy heights (Simard et al. 2011). GLAS produced 1064 nm and 532 nm laser pulses at a frequency of 40Hz and illuminated a spot on the ground measuring 60 m in diameter every 172 m in the along-track direction (Zwally et al. 2002). To examine the accuracy of GLAS for measuring vegetation parameters at the footprint level, data from GLAS footprints over Sam Houston National Forest in Texas were compared with tree measurements derived from discrete return airborne lidar (Popescu et al. 2011). In this study, GLAS-derived height variables explained 80% of the variance in airborne lidar derived AGB (RMSE = 37.7 Mg/ha) (Popescu et al. 2011), demonstrating strong potential for characterizing vegetation structure.

In terms of large-scale vegetation mapping, Simard et al. (2011) used Random Forest (RF) to model a relationship between GLAS canopy heights and spatially continuous, global ancillary variables that included climate maps, elevation from the Shuttle Radar Topography Mission (SRTM) and tree cover from the Moderate Resolution Imaging Spectroradiometer (MODIS). Validation of the 1-km wall-to-wall canopy height map with selected FLUXNET sites resulted in a RMSE of 6.1 m (R^2 = 0.50). In another study, Hu et al. (2016) utilized data from GLAS, ground inventory data, normalized difference vegetation index (NDVI) and land cover from MODIS, climatic variables, and topography from SRTM, to generate a wall-to-wall, global AGB map at 1 km spatial resolution. Height measurements from GLAS were first extrapolated to spatially continuous layers and then used with the other mapped predictor variables to build a relationship with ground-based estimates of AGB with RF (Hu et al. 2016). Evaluation of the final AGB product using the ground inventory data, gave a R^2 and RMSE of 0.56 and 87.5 Mg/ha respectively (Hu et al. 2016).

Like its predecessor, the Advanced Topographic Laser Altimeter System (ATLAS) instrument aboard ICESat-2 is designed to primarily measure ice sheet mass but will capture data that can potentially be used to measure structural characteristics of forests and contribute to AGB mapping (Markus et al. 2017) as previously demonstrated with GLAS data. The ATLAS instrument will collect measurements between 88° north and south latitudes during its three-year duration (Markus et al. 2017). There are substantial differences between GLAS and ATLAS where different processing methodologies for deriving and estimating forest attributes are required. Unlike ICESat's

full waveform lidar which emitted pulses in the near-infrared region, ATLAS is a photon counting system, operating in the visible wavelengths, at 532 nm (Glenn et al. 2016). GLAS produced a single beam and nominal 65 m-diameter footprints that were spaced about 172 m apart (Schutz et al. 2005) while ATLAS operates at 532 nm and emits three pairs of beams spaced 3.3 km apart with a 90-m distance within pairs (Markus et al. 2017). A pair consists of a strong beam and a weak beam based on a 4:1 transmit energy ratio (Neuenschwander and Magruder 2016). Each beam has a 14-m diameter footprint (Leigh et al. 2015), possibly up to 17 m (Markus et al., 2017) with an along-track sampling interval of 70 cm to facilitate dense sampling and serve to better capture elevation changes (NASA 2017). Thus, the improved spatial coverage associated with ATLAS will better capture changes in ice sheets than ICESat and potentially allow for vegetation mapping at a higher resolution (Abdalati et al. 2010).

The current literature has few vegetation studies focused on ICESat-2. For example, Montesano et al. (2015) conducted photon simulations representative of data that will be obtained from ICESat-2, calculated canopy height metrics, modeled AGB and computed uncertainty associated with AGB for a synthetic *Larix* forest gradient of a taiga-tundra ecotone. This study highlighted difficulty in differentiating AGB and characterizing vegetation structure under sparse forest cover, which is characteristic of *Larix* forest gradients in boreal forests (Montesanto et al. 2015). Lower precision in estimating heights in low-density savanna landscapes was also noted in a study by Gwenzi et al. (2016). Data from the Multiple Altimeter Beam Experimental Lidar (MABEL) instrument, which serves as test instrument for ICESat-2, was used to

investigate the capability of ICESat-2 in assessing savanna vegetation (Gwenzi et al. 2016). In another study, MABEL was used to assess low-height vegetation, mainly shrubs and grasslands as well as a denser vegetation cover consisting of tree species and sagebrush communities typical of dryland systems (Glenn et al. 2016). For the site dominated with low-height vegetation, MABEL data explained only 12% and 18% of the variance in shrub cover and vegetation height respectively compared to 49% of variance in the mean maximum height for the taller and more vegetation dense site (Glenn et al. 2016). A common approach in these studies is the use of unfiltered photon data similar to what will be derived by ATLAS which includes background or noise photons. Literature pertaining to ICESat-2's land and vegetation product or ATL08, is limited. The ATL08 product will provide terrain and canopy height estimates as well as canopy cover parameters from measurements, at a step-size of 100 m along the ground track (Neuenschwander and Magruder 2016).

This study highlights the use of data in a format similar to what will be provided by ICESat-2's ATL08 for estimating and mapping AGB and modeling canopy cover. It is essential to understand how the data can be utilized to assess forests and be prepared to have a methodology in place to use as soon as it becomes available. Using simulated ICESat-2 data along two years of planned ICESat-2 profiles, results from this study will demonstrate how data acquired by ATLAS onboard ICESat-2 can be used to estimate forest biophysical parameters, and provide insights about expected accuracies.

1.1. Project Significance

Lidar facilitates the mapping of the three-dimensional structure of forests, facilitating assessments of AGB over a range of scales (Popescu 2007) and an improved understanding of the amount and distribution of carbon stored by forest ecosystems (Hall et al. 2011). ICESat-2 offers an incredible opportunity to obtain large-scale coverage about vegetation and associated changes, through data collected along transects on the earth's surface. By deriving measurements within the laser's footprint and combining the results with other data sources, forest structure can be modeled beyond the extent of this footprint to provide spatially contiguous information at multiple scales. AGB and canopy cover estimation models and AGB maps, which are the primary outcomes of this study, will demonstrate approaches for the use of satellite data that can potentially support decision-making efforts to contribute to natural resource sustainability and climate change initiatives.

1.2. Research Objectives

The primary aim of this study was to develop approaches for utilizing data that will be provided by ICESat-2, for characterizing forest structure. It focused on Sam Houston National Forest (SHNF) in south-east Texas (Latitude 30° 42′ N, Longitude 95° 23′ W) using existing data for the area, to propose methodologies for estimating forest measurements and extrapolating data from lidar footprints to provide full wall-to-wall coverage for an entire area. The main objectives of this study were to:

1. Derive a simulated ICESat-2 PCL vegetation product along known ICESat-2 track locations over SHNF, using airborne lidar data and examine the use of simulated PCL metrics for modeling AGB and canopy cover.

2. Create wall-to-wall AGB maps at 30-m spatial resolution and characterize AGB uncertainty by using simulated photon-counting lidar (PCL)-estimated AGB and predictor variables from Landsat data and derived products.

3. Investigate deep learning (DL) neural networks for producing an AGB product with ICESat-2, using simulated ICESat-2 PCL-estimated AGB Landsat imagery, canopy cover and land cover maps.

Data for the following scenarios were analyzed for meeting the objectives of this study: (1) Simulated ICESat-2 PCL vegetation product from data with noise levels associated with daytime operation of ICESat-2 (daytime scenario), (2) Simulated ICESat-2 PCL vegetation product from data with noise levels associated with nighttime operation of ICESat-2 (nighttime scenario), and (3) Simulated ICESat-2 PCL vegetation product without the impact of noise (no noise). Each of the three main objectives above is addressed in Chapters 2, 3 and 4 of this dissertation. Chapter 2 presents the use of simulated ICESat-2 vegetation product from airborne lidar data for estimating AGB and canopy cover along known ICESat-2 track locations over SHNF. An approach for mapping AGB from simulated PCL-estimated AGB and predictor variables from Landsat data is provided in Chapter 3 and an investigation of DL neural networks for producing an AGB product is presented in Chapter 4. Each chapter was written as a standalone manuscript for submission in a peer-reviewed journal. Chapter 2 has been

published in *Remote Sensing of Environment* and Chapter 3 is currently under review in *Annals of Forest Research*. Final conclusions for this study are presented in Chapter 5 of this dissertation.

1.3. References

Baghdadi, N., le Maire, G., Fayad, I., Bailly, J.S., Nouvellon, Y., Lemos, C., Hakamada, R., & Ieee (2014). Estimation of Forest Height and Aboveground Biomass from ICESat-2/GLAS Data in Eucaluptus Plantations in Brazil. *2014 Ieee International Geoscience and Remote Sensing Symposium (Igarss)*, 725-728.

Campbell, J.B., & Wynne, R.H. (2011). *Introduction to Remote Sensing, Fifth Edition*. New York, New York: Guilford Publications.

Drake, J.B., Knox, R.G., Dubayah, R.O., Clark, D.B., Condit, R., Blair, J.B., & Hofton, M. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global Ecology and Biogeography, 12*, 147-159.

Glenn, N.F., Neuenschwander, A., Vierling, L.A., Spaete, L., Li, A.H., Shinneman, D.J., Pilliod, D.S., Arkle, R.S., & McIlroy, S.K. (2016). Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. *Remote Sensing of Environment, 185*, 233-242.

Gwenzi, D., & Lefsky, M.A. (2014). Prospects of photon counting lidar for savanna ecosystem structural studies. *Isprs Technical Commission I Symposium, 40-1*, 141-147.

Gwenzi, D., Lefsky, M.A., Suchdeo, V.P., & Harding, D.J. (2016). Prospects of the ICESat-2 laser altimetry mission for savanna ecosystem structural studies based on airborne simulation data. *Isprs Journal of Photogrammetry and Remote Sensing, 118*, 68-82.

Hall, F.G., Bergen, K., Blair, J.B., Dubayah, R., Houghton, R., Hurtt, G., Kellndorfer, J., Lefsky, M., Ranson, J., Saatchi, S., Shugart, H.H., & Wickland, D. (2011). Characterizing 3D vegetation structure from space: Mission requirements. *Remote Sensing of Environment, 115*, 2753-2775.

Harding, D.J., & Carabajal, C.C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters, 32*, 4.

Hu, T.Y., Su, Y.J., Xue, B.L., Liu, J., Zhao, X.Q., Fang, J.Y., & Guo, Q.H. (2016). Mapping Global Forest Aboveground Biomass with Spaceborne LiDAR, Optical Imagery, and Forest Inventory Data. *Remote Sensing, 8*, 27.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., & Birdsey, R.A. (2003). National-scale biomass estimators for United States tree species. *Forest Science, 49*, 12-35.

Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., & Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment, 115*, 2850-2860.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O., & de Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters, 32*.

Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B., & Hunter, M.O. (2007). Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing, 1*, 18.

Leigh, H.W., Magruder, L.A., Carabajal, C.C., Saba, J.L., & McGarry, J.F. (2015). Development of Onboard Digital Elevation and Relief Databases for ICESat-2. *Ieee Transactions on Geoscience and Remote Sensing, 53*, 2011-2020.

MacDicken, K.G. (2015). Global Forest Resources Assessment 2015: What, why and how? *Forest Ecology and Management, 352*, 3-8.

Margolis, H.A., Nelson, R.F., Montesano, P.M., Beaudoin, A., Sun, G., Andersen, H.-E., & Wulder, M.A. (2015). Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America1. *Canadian Journal of Forest Research, 45*, 838-855.

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., Popescu, S., Shum, C.K., Schutz, B.E., Smith, B., Yang, Y., & Zwally, J. (2017). The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sensing of Environment, 190*, 260-273.

Marselis, S., Armston, J., & Dubayah, R. (2016). *Summary of the Second GEDI Science Team Meeting*, in: National Aeronautics and Space Administration, The Earth Observer, November - December 2016, Volume 28, Issue 6, pp: 31-36.

Montesano, P.M., Rosette, J., Sun, G., North, P., Nelson, R.F., Dubayah, R.O., Ranson, K.J., & Kharuk, V. (2015). The uncertainty of biomass estimates from modeled ICESat-2 returns across a boreal forest gradient. *Remote Sensing of Environment, 158*, 95-109.

National Aeronautics and Space Administration. (2018). Global Ecosystem Dynamics Investigation Lidar (GEDI on ISS). https://eospso.nasa.gov/missions/global-ecosystem-dynamics-investigation-lidar (accessed: 05.10.18).

National Aeronautics and Space Administration, 2017. ICESat & ICESat-2. http://icesat.gsfc.nasa.gov/ (accessed: 05.10.18).

Neigh, C.S.R., Nelson, R.F., Ranson, K.J., Margolis, H.A., Montesano, P.M., Sun, G., Kharuk, V., Næsset, E., Wulder, M.A., & Andersen, H.-E. (2013). Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne LiDAR. *Remote Sensing of Environment, 137*, 274-287.

Nelson, R., Margolis, H., Montesano, P., Sun, G., Cook, B., Corp, L., Andersen, H.-E., deJong, B., Pellat, F.P., Fickel, T., Kauffman, J., & 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.

Neuenschwander, A.L., & Magruder, L.A. (2016). The Potential Impact of Vertical Sampling Uncertainty on ICESat-2/ATLAS Terrain and Canopy Height Retrievals for Multiple Ecosystems. *Remote Sensing, 8*, 16.

Neuenschwander, A., & Pitts, K. (2019). The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sensing of Environment, 221*, 247-259.

Popescu, S.C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy, 31*, 646-655.

Popescu, S.C., Zhao, K., Neuenschwander, A., & Lin, C. (2011). Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environment, 115*, 2786-2797.

Schutz, B.E., Zwally, H.J., Shuman, C.A., Hancock, D., & DiMarzio, J.P. (2005). Overview of the ICESat Mission. *Geophysical Research Letters, 32*, 4.

Simard, M., Pinto, N., Fisher, J.B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research-Biogeosciences, 116*, 12

Stysley, P.R., Coyle, D.B., Clarke, G.B., Frese, E., Blalock, G., Morey, P., Kay, R.B., Poulios, D., & Hersh, M. (2016). Laser Production for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar. In, *Conference on Laser Radar Technology and Applications XXI*. Baltimore, MD: Spie-Int Soc Optical Engineering.

Tewari, V.P. (2016). Forest inventory, assessment, and monitoring, and long-term forest observational studies, with special reference to India. *Forest Science and Technology, 12*, 24-32.

Zwally, H.J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S.,

Spinhirne, J., & Thomas, R. (2002). ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics, 34*, 405-445.

2. ESTIMATING ABOVEGROUND BIOMASS AND FOREST CANOPY COVER WITH SIMULATED ICESAT-2 DATA*

2.1. Introduction

Forests are an integral component of the earth's carbon cycle and contribute to maintaining carbon balance and mitigating climate change (Lu 2006; McKinley et al. 2011; Sessa and Dolman 2008). Even though forests account for 70-90% of terrestrial biomass (Houghton et al. 2009), there are uncertainties associated with the amount and distribution of carbon stored in forests (Hall et al. 2011; Hese et al. 2005; Sessa and Dolman 2008). Knowledge of the quantitative changes in forest biomass is even more limited but crucial to global carbon balance (Houghton et al. 2009). An estimated 50% of plant biomass is carbon (Drake et al. 2003; Le Toan et al. 2011) where estimates of AGB have been used as surrogates for aboveground carbon (Nelson et al. 2012). As such, estimation of AGB can reduce uncertainty in quantifying terrestrial carbon (Hese et al. 2005; Le Toan et al. 2011). Remote sensing systems have been recognized as practical tools that, when integrated with ground-based forest inventory, can facilitate reliable, up-to-date AGB estimates (Sessa and Dolman 2008). Light detection and ranging (lidar) instruments are active remote sensing devices that transmit energy and record backscattered energy from features on the earth's surface (Campbell and Wynne

* Reprinted with permission from "Estimating aboveground biomass and forest canopy cover with simulated ICESat-2 data" by Lana L. Narine, Sorin Popescu, Amy Neuenschwander, Tan Zhou, Shruthi Srinivasan, and Kaitlin Harbeck, 2019. *Remote Sensing of Environment*, 224, 1-11, DOI: 0.1016/j.rse.2019.01.037, Copyright [2019] by Elsevier.

2011). By measuring the time taken for laser energy to reach the surface and also return to the sensor and converting those times to distance measurements, three-dimensional attributes become obtainable (Popescu 2007). The data acquired by the lidar sensor can then be analyzed to provide precise vegetation metrics. Measurements, including canopy height, basal area, timber volume and biomass have all been successfully derived from lidar at multiple spatial scales (e.g. Boudreau et al. 2008; Holmgren 2004; Neigh et al. 2013; Nelson et al. 2017; Popescu 2007; Rahlf et al. 2014). Thus, lidar can be used to monitor biomass over a large scale by providing information on three-dimensional vegetation structure.

Lidar sensors are mounted on terrestrial, airborne and spaceborne platforms, with each type offering unique advantages. For instance, terrestrial lidar scanning (TLS) has the capability of acquiring tree-level attributes like diameter at breast height (dbh) with high accuracy (RMSEs < 4 cm) and millimeter-level detail (Kankare et al. 2015; Liang et al. 2016). Tree diameters have also been accurately estimated with lidar sensors from airborne platforms (e.g. Popescu 2007) and AGB and gross volume have been estimated at the stand level (e.g. Sheridan et al. 2015). Spaceborne lidar offers a remarkable opportunity to assess forest resources at or near global scales and cover areas that are traditionally inaccessible to ground-based samples. The National Aeronautics and Space Administration (NASA) launched Ice, Cloud, and land Elevation Satellite (ICESat) in January 2003 with the primary goal of measuring ice sheets via the Geoscience Laser Altimeter System (GLAS) sensor, a spaceborne lidar. This mission served to capture changes in ice elevation and facilitate an assessment of their contributions to global sea-

level rise (Zwally et al. 2002). The ICESat mission was successfully completed in October 2009, providing invaluable global data about polar ice sheets as well as measurements that were analyzed to provide cloud information and vegetation metrics (e.g. Farrell et al. 2009; Khan et al. 2014; Lefsky et al. 2007; Pritchard et al. 2009; Spinhirne et al. 2005). Noteworthy is that the GLAS sensor onboard ICESat represented the first and only spaceborne lidar instrument to capture measurements at a near-global scale. GLAS captured 1.92 billion lidar waveforms globally and facilitated the estimation of forest resources, specifically forest biomass, over traditionally inaccessible forested areas like boreal forests and the tropics (Nelson et al. 2009). Wall-to-wall maps of forest canopy heights have also been developed using GLAS data (Lefsky 2010; Simard et al. 2011). Finding strong correlations between GLAS-derived canopy heights and other existing data, Simard et al. (2011) modeled global canopy heights using GLAS data and tree cover, elevation and climatology maps.

A follow-on mission to ICESat, ICESat-2, was launched on September 15th, 2018 (NASA 2017). The Advanced Topographic Laser Altimeter System (ATLAS) onboard ICESat-2 was designed to primarily measure ice sheet mass but captures data that can potentially be used to measure structural characteristics of forests at large spatial scales (Markus et al. 2017). Vegetation data products from ICESat-2 will complement other space-based vegetation missions (Markus et al. 2017), such as the Global Ecosystem Dynamics Investigation Lidar (GEDI) (Stysley et al. 2016), NASA-ISRO Synthetic Aperture Radar (NISAR) (Rosen et al. 2016) and the European Space Agency P-band radar BIOMASS mission (Carreiras et al. 2017). It is therefore necessary to

determine how data collected by ICESat-2 can be used to contribute to estimating and monitoring forest carbon.

There are substantial differences between GLAS and the ATLAS instrument (Table 1), where different processing methodologies for obtaining and extrapolating forest measurements are required. ATLAS operates at 532 nm and unlike a single beam produced by GLAS, emits three pairs of beams spaced 3.3 km apart with a 90-m distance within pairs (Markus et al. 2017). A pair consists of a strong beam and a weak beam based on a 4:1 transmit energy ratio (Neuenschwander and Magruder 2016). Each beam has a 14-m diameter footprint (Leigh et al. 2015), possibly up to 17 m (Markus et al. 2017) with an along-track sampling interval of 70 cm to facilitate dense sampling and serve to better capture elevation changes (NASA 2017). GLAS emitted pulses in the near-infrared region which is important for vegetation studies (Campbell and Wynne 2011) but the green wavelength used by ATLAS is associated with reduced leaf reflectance (Swatantran et al. 2016) and its use may also be limited during overcast conditions due to absorption by clouds (Lefsky et al. 2002, p. 20). However, the improved spatial coverage associated with ATLAS will better capture changes in ice sheets than ICESat and allow for vegetation mapping at a higher resolution (Abdalati et al. 2010). Improved performance, in terms of sampling rates and data collection with the single photon lidar (SPL) technique used for ICESat-2, has also been demonstrated (Degnan 2002; Swatantran et al. 2016).

Table 1. Comparison of GLAS on ICESat with ATLAS onboard ICESat-2 (from Markus et al. 2016)

Specification	System	
	Geoscience Laser Altimeter System (GLAS) on ICESat	Advanced Topographic Laser Altimeter System (ATLAS) on ICESat-2
Measurement approach	Energy waveform	Photon counting
Wavelength	1064 nm	532 nm
Repetition rate	40 Hz	10 kHz
Number of beams	1	6 (3 pairs with 90 m space within pairs and 3.3 km pair seperation)
Footprint size	70 m	14 m
Along-track sampling	172 m	0.7 m

One of ICESat-2's mission data products will be its land-vegetation along track product (ATL08) (Brown et al. 2013). ATL08 will provide terrain and canopy height estimates as well as canopy cover parameters from measurements, at a step-size of 100 m along the ground track (Neuenschwander et al. 2017). These data parameters will be used to obtain ATL18 products which consist of gridded terrain and canopy maps. ATL08 and ATL18 products will therefore facilitate forest assessments at global scales and facilitate biomass and carbon monitoring that would not otherwise be possible with only ground measurements.

The current literature has few PCL studies focused on ICESat-2. For instance, Popescu et al. (2018) describe algorithms to filter noise, classify photons and retrieve terrain and canopy heights from ICESat-2 data. The application of noise filtering and photon classification algorithms to simulated ICESat-2 data yielded average RMSE values of 2.70 m and 3.59 m for estimating canopy heights. Overall, results from testing the algorithms to process PCL data similar to expected ICESat-2 data indicate potential for characterizing forest structure with ICESat-2 (Popescu et al. 2018). Montesano et al.

(2015) conducted photon simulations representative of data that will be obtained from ICESat-2, calculated canopy height metrics, modeled AGB and computed uncertainty associated with AGB for a synthetic *Larix* forest gradient of a taiga-tundra ecotone. This study highlighted difficulty in differentiating AGB and characterizing vegetation structure under sparse forest cover, which is characteristic of *Larix* forest gradients in boreal forests (Montesano et al. 2015). Lower precision in estimating heights in low-density savanna landscapes was also noted in a study by Gwenzi et al. (2016). Data from the Multiple Altimeter Beam Experimental Lidar (MABEL) instrument, which serves as test instrument for ICESat-2, was used to investigate the capability of ICESat-2 in assessing savanna vegetation (Gwenzi et al. 2016). In another study, MABEL was used to assess low-height vegetation, mainly shrubs and grasslands as well as a denser vegetation cover consisting of tree species and sagebrush communities typical of dryland systems (Glenn et al. 2016). For the site dominated with low-height vegetation, MABEL data explained only 12% and 18% of the variance in shrub cover and vegetation height respectively compared to 49% of variance in the mean maximum height for the taller and more vegetation dense site (Glenn et al. 2016). Since more accurate estimates of vegetation characteristics are critical for vegetation management and monitoring activities, a synergistic approach between ATLAS metrics and Landsat 8 OLI data was emphasized as a more feasible approach. A common approach in these studies is the use of unfiltered photon data similar to what will be derived by ATLAS, which includes background or noise photons. Literature pertaining to ICESat-2's vegetation products, specifically ATL08, is limited. Thus, the investigation of a methodology involving the

utilization of ICESat-2 for assessing forest parameters, focusing on the mission's vegetation product, will contribute to existing studies on ICESat-2 PCL data.

This study highlights the use of data in a format similar to what will be provided by ICESat-2's ATL08 for modelling AGB and canopy cover. Results will demonstrate how data acquired by ATLAS onboard ICESat-2 can be used to measure forest biophysical parameters.

2.1.1. Objectives

The goal of this study was to develop approaches for utilizing data that will be collected by ICESat-2 for estimating AGB and canopy cover. As a representative type of forests of the southeastern US, Sam Houston National Forest (SHNF) served as the study area and existing data were used for developing a methodology to estimate forest attributes. The specific objectives of this study were to:

1. Derive a simulated ICESat-2 PCL vegetation product along known ICESat-2 track locations over SHNF, using airborne lidar data.
2. Examine the use of simulated PCL metrics for modelling AGB along ICESat-2 profiles using a simulated ICESat-2 PCL vegetation product and reference AGB estimated from airborne lidar data.
3. Estimate forest canopy cover using simulated PCL canopy product data and airborne lidar-derived canopy cover.

Data for the following scenarios were analyzed for meeting the objectives of this study: (1) Simulated ICESat-2 PCL vegetation product from data with noise levels

associated with nighttime operation of ICESat-2 (nighttime scenario), (2) Simulated ICESat-2 PCL vegetation product from data with noise levels associated with daytime operation of ICESat-2 (daytime scenario), and (3) Simulated ICESat-2 PCL vegetation product without the impact of noise.

2.2. Materials and methods

2.2.1. Study Area

The study area is located in south-east Texas (30° 42′ N, 95° 23′ W) and encompasses an estimated area of 48 km² in the SHNF. It is situated in Walker County, Texas within the Pineywoods ecoregion, which is characterized as being the wettest area in the state and consisting of predominantly pine forests (Texas A&M Forest Service, 2015). According to the 2011 National Land Cover Database (NLCD), approximately 58% of the region or 80% of its forested area (NLCD classes: deciduous forest, evergreen forest, mixed forest, and woody wetland) is classified as evergreen forests (Homer et al. 2015; MRLC 2017), which primarily include Loblolly pine (*Pinus taeda*) plantations and old growth Loblolly pine stands (Popescu 2007). The site also includes hardwoods that grow on floodplains and drier uplands. Popular upland and bottomland hardwoods are White Oak (*Quercus alba*), Southern Red Oak (*Quercus falca*), Water Oak (*Quercus nigra*) and Sweetgum. Based on a three-dimensional canopy height model (CHM) generated from discrete return lidar data for the study area, the average tree height is estimated to be 13 m and maximum, 65 m. Average tree canopy cover is estimated to be 86% within vegetated areas and 97% within pine forests only (MRLC

2017). The area consists of gentle slopes with elevations ranging from 62 m to 105 m and an average elevation of 85 m (Popescu 2007).

The study site was also used for examining GLAS data from the ICESat mission for AGB estimates (Popescu et al. 2011), which allows for comparisons of AGB estimates from forthcoming ICESat-2 data and those derived from ICESat. ICESat-2 tracks for the first two years of the mission were made available by its Science Definition Team (SDT) (Neuenschwander et al. 2017). These track locations over the study site, as represented by 1 m National Agriculture Imagery Program (NAIP) imagery (USDA Farm Service Agency 2015) from the same year of discrete return lidar acquisition, are illustrated in Figure 1.

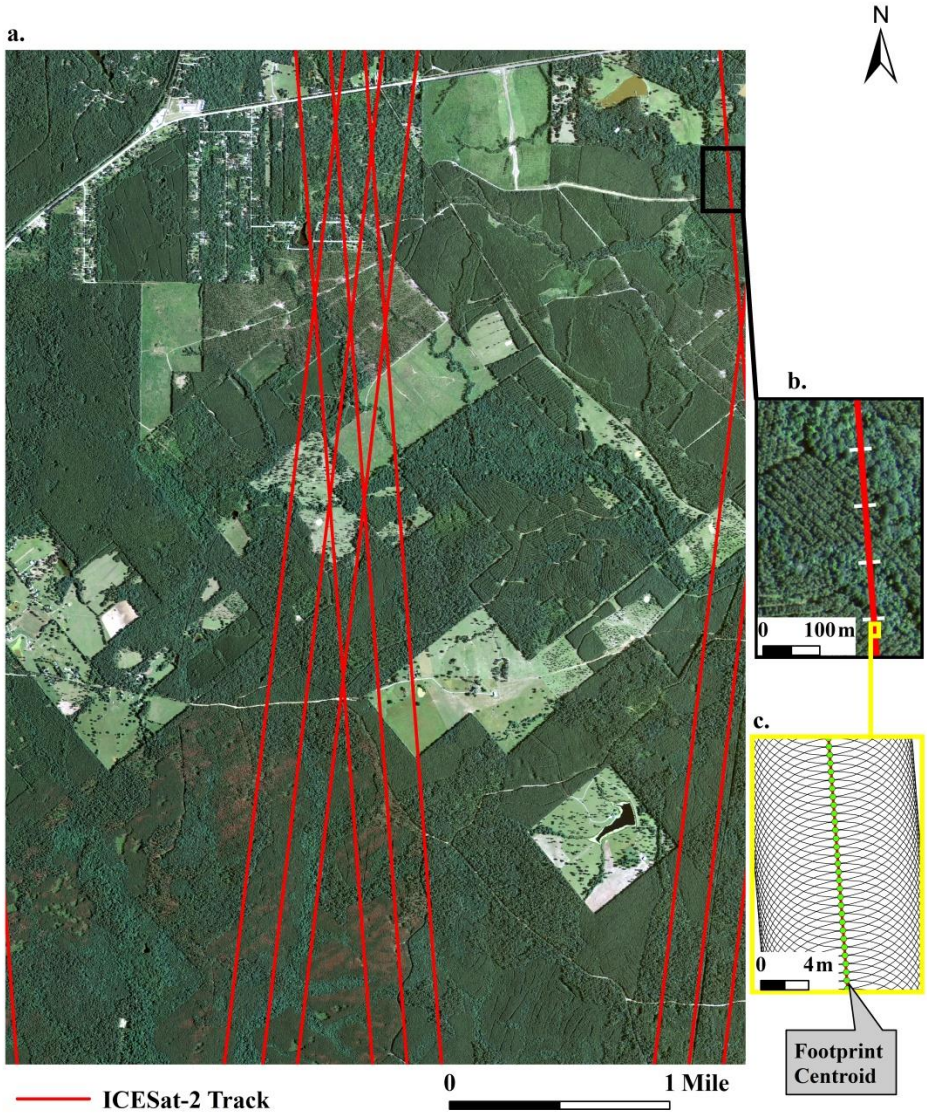


Figure 1. a. Map of ICESat-2 tracks overlaid on 2010 NAIP aerial imagery for the study area within SHNF with demarcation of 100 m segments along-track shown on inset map b. and an illustration of ICESat-2 along-track footprints spaced 70 cm apart, with 7 m radii on inset map c.

2.2.2. Airborne lidar data and simulated PCL data

Discrete return lidar data were acquired for SHNF in November 2010 using an Optech 3100EA sensor and were used to simulate the PCL data structures similar to what are collected by the ATLAS system. Airborne lidar data were acquired along 12 flight lines in the north-south direction and 19 lines flown from east to west, at around 600 m above-ground-level (AGL) with a point density of 4 points per m^2. PCL data were generated using an ICESat-2 footprint generator and PCL simulator (Neuenschwander and Magruder 2016). The footprint generator and PCL simulator were developed in MATLAB/Simulink (The MathWorks, Inc. 2015). Footprints were created using the start and end coordinates for each proposed ICESat-2 track within the study area (Figure 1) and a 70-cm center-to-center spacing along the track.

The PCL simulation process was previously used to evaluate vertical sampling error associated with ICESat-2 terrain and canopy height retrievals (Neuenschwander and Magruder 2016). The data used were simulated from full-waveform and discrete return airborne lidar data and represented expected height retrievals without the impact of solar noise (Neuenschwander and Magruder 2016). In this study, airborne lidar data processed to obtain AGL heights were used as the input data for simulating photons. PCL data were generated by first identifying lidar points falling within a 7 m radius, representative of an ICESat-2 footprint size (Figure 1c) (Leigh et al. 2015). A histogram of the height values was then constructed to create a 14-m diameter pseudo-waveform, which characterized the vegetation's vertical structure within the footprint (Blair and Hofton 1999; Neuenschwander and Magruder 2016). This was normalized to create a

probabilistic distribution function (PDF) and a height vector was generated for each footprint as described in Neuenschwander and Magruder (2016). A Poisson random distribution was then used to randomly sample the number of return photons. The mean number of signal photons per shot for the study area was modeled at 1.9, based on the ATLAS performance model for temperate forests (Martino 2010). This signal level represents the best case scenario for vegetated areas, compared to 1.0 over boreal forest and 0.6 for tropical forest (Neuenschwander and Magruder 2016). A weighted random sample from the height vector was carried out with the normalized PDF as the weighing function. The algorithm returned photon heights and height metrics and a mean of 2 photons were generated per footprint.

In addition to signal photons, other photons (noise) are detected by the ATLAS instrument. These photon events are associated with solar background noise and atmospheric scattering (Moussavi et al. 2014) and since they are not discernable from signal photons, represent a challenge with analyzing PCL data (Glenn et al. 2016; Moussavi et al. 2014). With ICESat-2 operating under both day and night conditions, the effect of solar background noise is expected to be the major source of noise during daytime operations (Degnan 2002). To account for the effects of noise, another simulation was carried using the 2010 discrete return lidar dataset for the site. After signal photons were generated using the PCL simulator, noise levels representative of the impact of the atmosphere and solar background noise, were added to the dataset. In total, data for three scenarios were analyzed; 1) simulated data without the addition of noise, 2) daytime scenario, and 3) nighttime scenario. Noise photons were added and

data processed using noise filtering and photon classification algorithms for the latter two scenarios (Popescu et al. 2018).

2.2.3. Reference AGB Estimates

A three-dimensional 1 m canopy height model (CHM) was generated from the airborne lidar data collected for the study area in 2010 and used as input in TreeVaW, a lidar processing software that can automatically retrieve tree-level parameters like tree locations and crown widths (Popescu et al. 2003; Popescu and Wynne 2004). The identification of individual trees with TreeVaW relies on the relationship between tree heights and crown widths, which can be manually specified by adjusting the default regression model based on loblolly pines. Linear regression models were used for pines and deciduous species separately, to develop equations between crown size as the independent variable and tree height. Ground measured tree heights and corresponding crown diameters from a total of 705 pine trees and 603 deciduous trees from field inventories carried out in 2004 and 2009 were used to obtain TreeVaW coefficients. Output consisting of estimated dbh measurements, were then used to estimate AGB for individual trees following a generalized biomass regression equation (Eq. (1)):

$$\text{AGB (kg)} = \exp(\beta_0 + \beta_1 \ln \text{dbh}), \quad (1)$$

where dbh is the diameter at breast height (cm) and β_0 and β_1, are the parameters for the species group (Jenkins et al. 2003, p. 20). AGB for pines were calculated using the parameters for the “pine” species groups, with β_0 = -2.5356 and β_1= 2.4349. To

calculate AGB for deciduous trees, parameters for the "hard maple/oak/hickory/beech" group were used, with β_0 = -2.0127 and β_1= 2.4342 (Jenkins et al. 2003, p. 20).

A similar methodology for estimating AGB of pine trees within the same study area was previously introduced by Popescu (2007). In this study, aboveground biomass of field-measured pine trees was estimated using the general biomass equation for pines from Jenkins et al. (2003), a CHM from airborne lidar data was processed with TreeVaW and AGB was estimated using lidar-derived dbh. The accuracy of lidar-derived dbh was assessed and the relationship between lidar and ground estimated biomass was also evaluated with linear regression. Results indicated a good fit, where the lidar-estimated biomass explained 87% of the variance of field-estimated AGB, with an RMSE of 169 kg, representing 47% of the dependent mean (Popescu 2007, p. 652). Evidently, lidar data reliably estimated dbh (R^2 = 0.87, RMSE = 4.9 cm), which were used to estimate biomass of trees within SHNF.

2.2.4. Airborne Lidar Canopy Cover

Canopy cover calculated from airborne lidar data was used as a dependent variable for developing regression models with simulated PCL-derived metrics. To retrieve canopy cover parameters, airborne lidar data for SHNF was clipped using LAStools (Isenburg 2015) with the spatial coverage of simulated PCL data within the study area used as the clipping extent. Canopy cover, defined as the proportion of all returns above a specified height threshold to the total returns (USDA Forest Service 2014) within a 100 m segment, was considered at 2 m and 4.6 m height thresholds.

Consistent with Li et al. (2015), a 2 m cut-off height implied that canopy heights under this value were not considered forest vegetation, while the latter height cut-off of 4.6 m represented the US Forest Service's minimum height of a tree (USDA Forest Service 2016). Airborne lidar canopy cover metrics were computed with FUSION (McGaughey 2016) (Version 3.50) for matching segments of simulated PCL and reference airborne lidar-estimated AGB data as described in Section 2.5.

2.2.5. Data processing

The distinction of signal photons from a noise background is critical for accurate vegetation height retrievals. Processing algorithms described in Popescu et al. (2018) were used to derive forest canopy heights from the simulated PCL data associated with daytime and nighttime scenarios. Noise filtering consisted of three main steps. To summarize, first, elevation information from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) GDEM (version 9) and the Global Canopy Height Map (GCHM) were used as an initial filter to retain photons that were likely from the ground and canopy. Second, cluster analysis was used to remove outlier photons and third, 95% confidence interval (CI) filters within each along-track distance interval were applied. To classify photons, an overlapping moving window was applied and cubic spline function used to obtain top of canopy (TOC) and ground surfaces. The detailed steps of the noise filtering and photon classification algorithms, and an assessment of algorithm performance under several vegetation conditions and noise levels associated with different operation times, are presented in Popescu et al. (2018).

After processing the simulated PCL data scenarios (daytime and nighttime), canopy heights were derived by subtracting the ground elevation from estimated top of canopy heights.

The processed simulated PCL datasets for the daytime and nighttime scenarios and the simulation without the addition of noise were each clipped to create 100 m along-track segments using FUSION (McGaughey 2016). A 100 m step-size along the proposed ground tracks was chosen since canopy and terrain parameters for ATL08 will be provided at the same scale, with each 100 m described as a segment (Neuenschwander et al. 2017). In this study, data for segments less than 100 m were excluded, resulting in a total of 606 100 m segments for each scenario, within the study site.

Lidar studies have demonstrated significant relationships between mean height, height percentiles and AGB (e.g. Ku et al. 2012; Sheridan et al. 2015). Height percentiles, canopy cover and canopy density metrics for the simulated PCL segments were calculated using FUSION (McGaughey 2016). A list of the variables derived for each of the three scenarios, is presented in Table 2. Canopy cover was calculated at two height thresholds; 2 m and 4.6 m. A height-bin approach was previously implemented by Popescu and Zhao (2008) within the study area to examine the vertical structure of the forest canopy. Similarly, canopy density, defined as the proportion of returns in each height stratum to the total returns within each cell (USDA Forest Service 2014, p.3) was calculated for 0 - 25 m height bins, at 5 m intervals. The last height bin included photon returns above 25 m.

Spatially coincident reference AGB values were retrieved in ArcMap and a similar approach was undertaken with the airborne lidar data in LAStools to obtain canopy cover metrics within the same extent. AGB values for 100 m segments were up-scaled to represent per hectare estimates for regression analysis.

2.2.6. Statistical analysis

Examination of the reference airborne lidar-estimated AGB using variogram analysis and calculation of Moran's I suggested the presence of spatial autocorrelation. The range was estimated to be 500 m so systematic sampling was implemented to select every 5th 100 m segment from a randomly selected starting point, resulting in a total of 121 segments for each scenario that were analyzed. The simulated PCL segments were randomly assigned to training and test sets, where one-third of the segments were used for model testing (n = 36). Three regression models were developed with the training datasets (n = 85) to relate airborne lidar data to simulated PCL parameters. Utilizing simulated PCL data for estimating AGB involved the use of linear regression models that incorporated simulated PCL-derived metrics from 100 m segments and AGB acquired from corresponding spatial extents. Canopy cover calculated from airborne lidar data within spatially coincident segments and PCL height, canopy cover and canopy density metrics were also used for developing a relationship with airborne lidar-derived canopy cover. A total of 16 variables (Table 2) were considered for building linear regression models to relate simulated PCL parameters with the dependent variables. Stepwise regression was used to fit possible models for predicting AGB and

canopy cover and Corrected AIC (AIC_C) and Mallow's C_p criteria were applied for selecting the best model. Results were also compared with those obtained from the mixed stepwise regression algorithm to select variables for the final multiple regression models. Multicollinearity among the selected independent variables was then investigated using variance inflation factors (VIFs) where variables with a VIF > 10 were considered indicators of multicollinearity and were subsequently removed. This indicator has been applied in other studies involving biomass estimation with remote sensing data using regression analysis (e.g. Margolis et al. 2015; Nelson et al. 2017; Sheridan et al. 2015; Srinivasan et al. 2014).

To evaluate model performance with the separate test set, R^2 and RMSE were calculated (Eq. (2) and Eq. (3)):

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n}(P_i - O_i)^2/n} \quad (2)$$

$$R^2 = 1 - \sum_i(P_i - O_i)^2 \; / \; \sum_i(O_i - \bar{O_i})^2 \quad (3)$$

Where n is the number of observations (n = 36), P_i is the model-predicted AGB, O_i is the reference airborne lidar-estimated AGB, and $\bar{O_i}$ is the mean reference airborne lidar-estimated AGB.

Table 2. Regression variables

Independent Variables (Simulated PCL data)	Description	Response Variables (AGB and canopy cover from airborne lidar data)
Mean, maximum and percentile heights (meters):		- AGB (Mg/ha)
max	Maximum height	- Canopy cover (percentage of returns above 2 m / 4.6 m)
mean	Mean height	
P10	10th percentile height	
P25	25th percentile height	
P50	50th percentile height	
P75	75th percentile height	
P90	90th percentile height	
P95	95th percentile height	
Canopy cover:		
C1_2m	Percentage of all returns above 2 m	
C2_4_6m	Percentage of all returns above 4.6 m	
Canopy density:		
D1_0_5	Proportion of returns in the stratum- 0-5 m	
D2_5_10	Proportion of returns in the stratum - 5-10 m	
D3_10_15	Proportion of returns in the stratum - 10-15 m	
D4_15_20	Proportion of returns in the stratum - 15-20 m	
D5_20_25	Proportion of returns in the stratum - 20-25 m	
D6_25	Proportion of returns in the stratum - above 25 m	

2.3. Results

Linear regression results for estimating AGB and canopy cover with simulated PCL data are highlighted in Table 3. Variables in the final models were all statistically significant. The PCL simulation with no noise represented the best case scenario for the study site. With this simulation, the expected photon detection rate for the forest type in the study area was applied, but background noise, which represents a challenge for photon counting systems (Glenn at al. 2016), was not included. The resulting AGB model with this data used mean height, and the 15-20 m height bin, explained 80% of the variance in reference airborne lidar-estimated AGB for the training set and yielded a

RMSE of 22.14 Mg/ha or 28% of the dependent mean of 79 Mg/ha. Both variables were statistically significant at 0.001 level. When the model was fitted on the test set, the R^2 and RMSE were 0.79 and 19.16 Mg/ ha respectively (Figure 2c). Canopy cover models exhibited greater predictive abilities than the AGB model. The final model for the 2 m canopy cover used 25th and 50th height percentiles as well as the canopy cover metric (2 m), which were all statistically significant at 0.05 level. These variables explained 91% variance of the airborne lidar-derived canopy cover calculated at the 2 m height threshold for the training dataset and test set. RMSE values were 6.30% (10% of mean airborne lidar-derived canopy cover of 62%) and 7.58% with the training data and test data respectively. Canopy cover (4.6 m) calculated with the simulated PCL data was the single best predictor of airborne lidar-derived canopy cover at this threshold. The simulated PCL canopy cover at the 4.6 m threshold was significant at 0.001 level. This metric explained 91% of the variance in airborne lidar-derived canopy cover and had a RMSE value of 6.48% (or 11% of a mean canopy cover of 57%). With the test dataset, the R^2 and RMSE values were 0.93 and 6.36% respectively.

Estimation models with the nighttime dataset did not perform as well as those using the simulation without the addition of noise but results are highly indicative of potential for AGB and canopy cover estimation with ICESat-2 data (Figures 2a, 3a, and 4a). With the nighttime scenario, simulated PCL-derived 10th and 90th height percentiles and canopy cover (4.6 m) explained 74% of the variance in reference airborne lidar-estimated AGB with a RMSE of 25.50 Mg/ha, representing 32% of the dependent mean. The three variables in the model were found to be statistically

significant at 0.05 level. When the model was fitted on the test set, the R^2 increased to 0.79 and the RMSE was 19.23 Mg/ha. As with the no noise scenario, canopy cover models outperformed the AGB model. In terms of canopy cover at the 2 m height threshold, R^2 for model training and model evaluation were 0.65 and 0.83 respectively. This linear regression model used the 10th height percentile (p-value < 0.001) and canopy cover (4.6 m) (p-value < 0.05) as independent variables and indicated RMSE values of 12.49% (20% of the dependent mean) and 10.39% with the training data and test set respectively. The canopy cover model for the 4.6 m height threshold used three variables that were significant at 0.05 level. Maximum height, 10th height percentile and canopy density for the 15-20 m height bin explained 68% and 75% of the variance associated with the airborne lidar-derived canopy cover (4.6 m) and provided RMSE values of 12.31% (22% of dependent mean) and 12.33% with the model-building dataset and test dataset respectively.

AGB and canopy cover models generated with the daytime data did not perform as well as the two other scenarios (Figures 2b, 3b, and 4b). The 10th and 90th height percentiles and canopy cover (4.6 m) explained 66% and 63% of the variance associated the reference airborne lidar-estimated AGB and RMSE values were 28.90 Mg/ha or 37% of a mean value of 79 Mg/ha with the training dataset and 25.35 Mg/ha with the test data. Similar R^2 values were observed for the canopy cover models. The regression model for canopy cover at the 2 m used the 10th height percentile which was significant at 0.001 level, and provided a R^2 value 0.50 and a RMSE of 14.50% (24% of the dependent mean) with the training dataset. The 10th height percentile as well as canopy

density for the 15-20 m height bin explained 57% variance of airborne lidar-derived canopy cover model at the 4.6 m height threshold with a RMSE of 14.02% (25% of the dependent mean). These models explained 56% and 63% of the variability associated with canopy cover estimated for the 2 m and 4.6 m height thresholds from the test set respectively. With the test data, RMSE values were larger for the daytime scenario; 16.69% and 15.01% for the 2 m canopy cover and 4.6 m canopy cover respectively.

Table 3. Regression results for estimating AGB and canopy cover

Scenario	Dependent Variable	RMSE		R^2		Model
		Training	Test	Training	Test	
Nighttime	**AGB**	25.50 Mg/ha	19.23 Mg/ha	0.74	0.79	-10.00 + 4.62P10 + 3.29P90 - 0.54C2_4_6m
	Canopy Cover (2 m)	12.49%	10.39%	0.65	0.83	-0.98 + 1.67P10 + 0.39C2_4_6m
	Canopy Cover (4.6 m)	12.31%	12.33%	0.68	0.75	2.44 + 0.88max + 1.90P10 + 14.63D4_15_20
Daytime	**AGB**	28.90	25.35	0.66	0.63	14.39 + 5.50P10 + 2.35P90 - 0.67C2_4_6m
	Canopy Cover (2 m)	14.50%	16.69%	0.50	0.56	28.36 + 2.26P10
	Canopy Cover (4.6 m)	14.02%	15.01%	0.57	0.63	18.55 + 2.29P10 + 10.86D4_15_20
Simulation with no noise	**AGB**	22.14 Mg/ha	19.16 Mg/ha	0.80	0.79	-27.21 + 11.63mean - 101.62D4_15_20
	Canopy Cover (2 m)	6.30%	7.58%	0.91	0.91	-8.51 + 0.86P25 - 0.91P50 + 0.95C1_2m
	Canopy Cover (4.6 m)	6.48%	6.36%	0.91	0.93	-6.14 + 0.87C2_4_6m

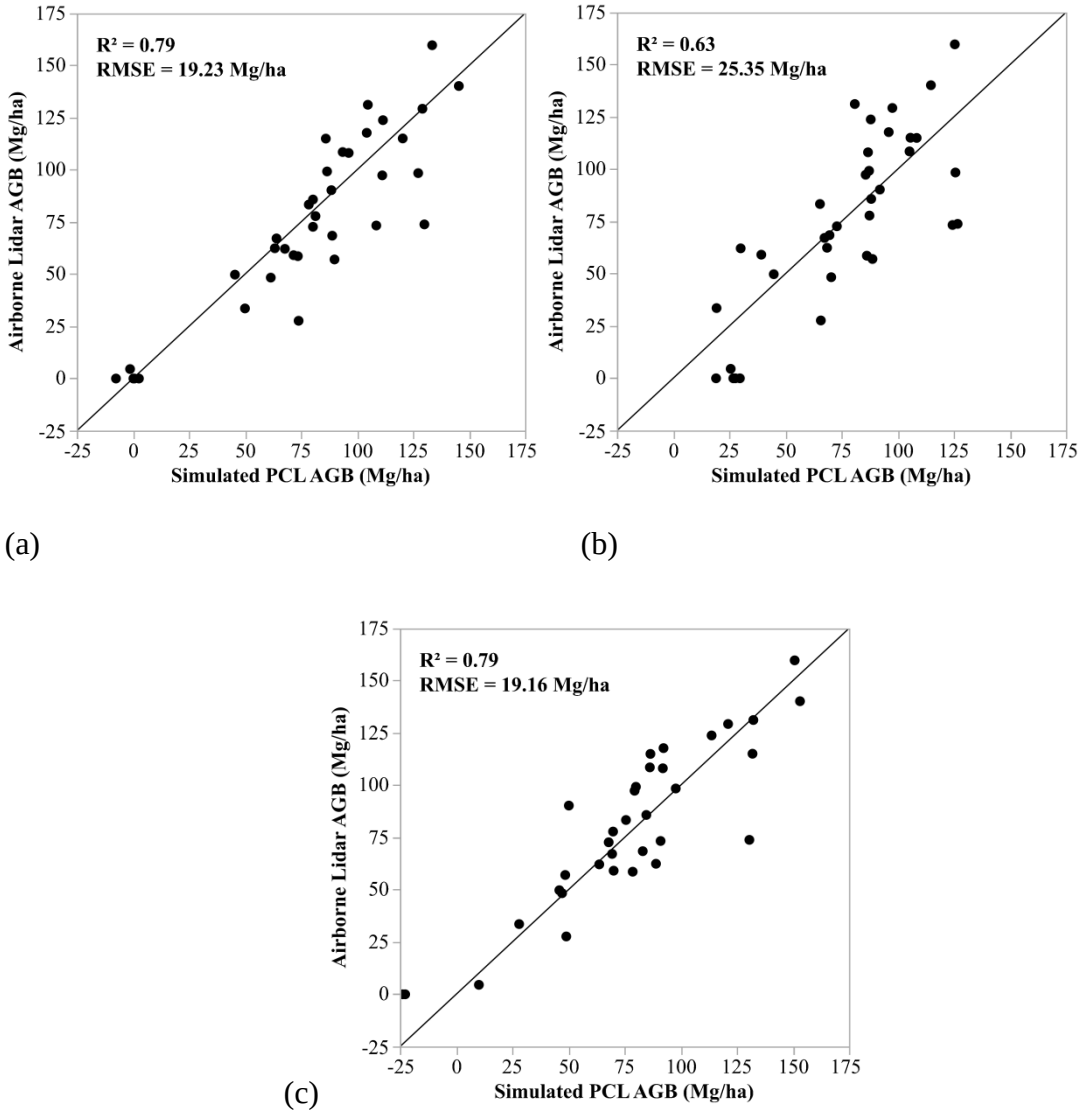
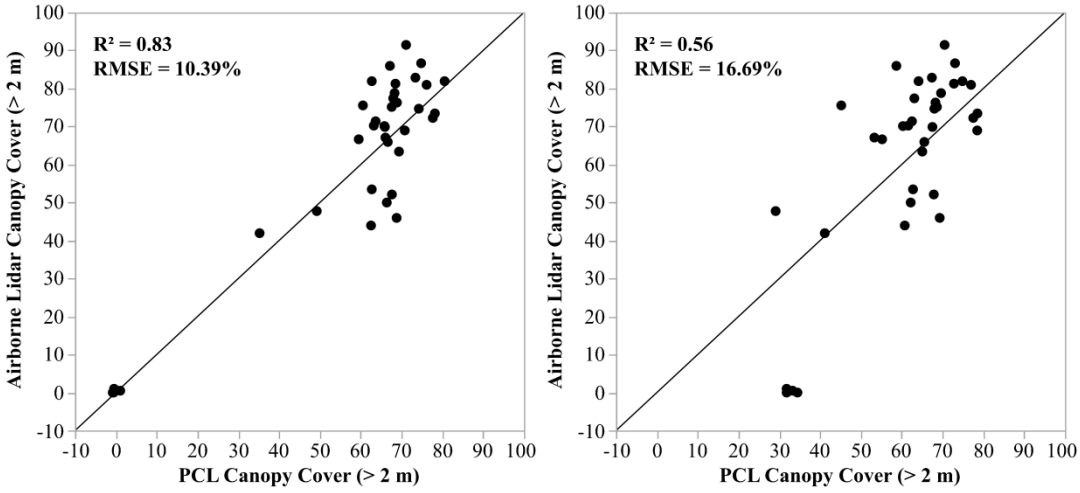
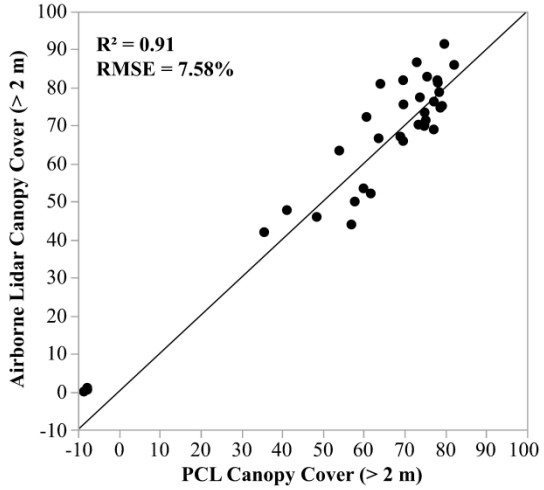


Figure 2. Reference airborne lidar-estimated AGB versus predicted AGB (Mg/ha) with the test data, shown with the 1:1 line for the (a) Nighttime scenario, (b) Daytime scenario, and (c) PCL simulation with no noise

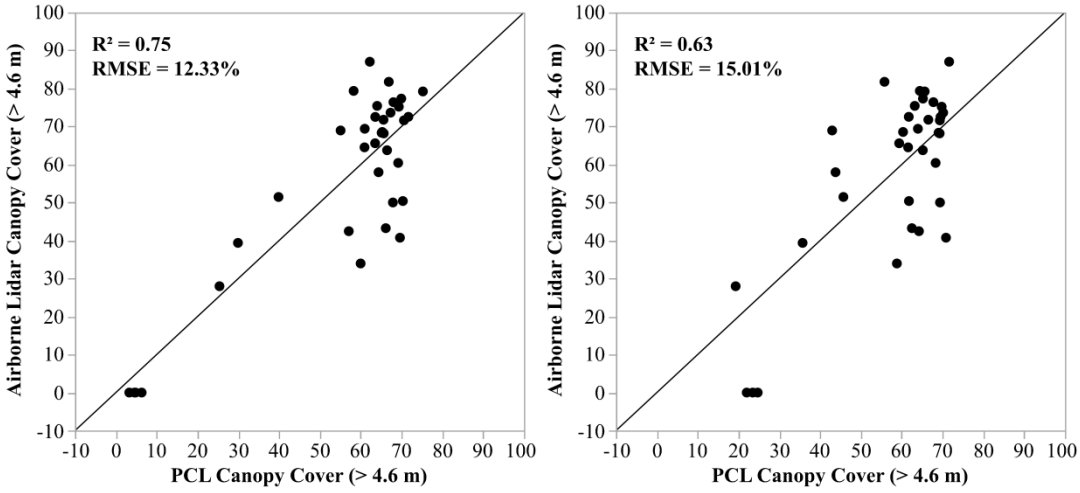


(a) (b)

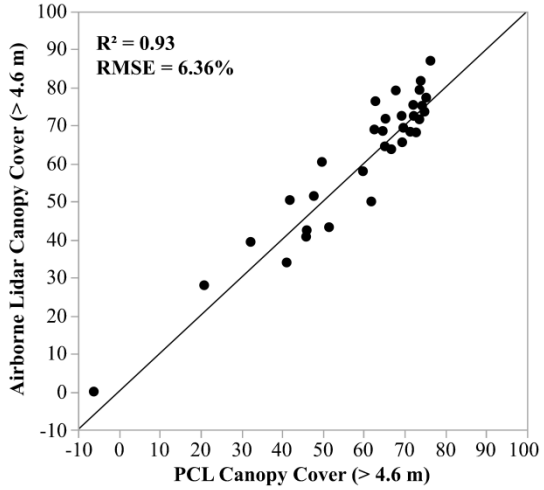


(c)

Figure 3. Airborne lidar-derived canopy cover versus canopy cover (> 2 m) estimated from simulated PCL data using the test dataset, shown with the 1:1 line for the (a) Nighttime scenario, (b) Daytime scenario, and (c) PCL simulation with no noise



(a) (b)



(c)

Figure 4. Airborne lidar-derived canopy cover versus canopy cover (> 4.6 m) estimated from simulated PCL data using the test dataset, shown with the 1:1 line for the (a) Nighttime scenario, (b) Daytime scenario, and (c) PCL simulation with no noise

2.4. Discussion

This study aims to contribute to an understanding of the data that will be acquired from ICESat-2 for characterizing vegetation structure using simulated PCL data over forests located in southeastern Texas. Model testing with separate datasets for the simulated ICESat-2 data scenarios provided promising results with R^2 values between 0.56 and 0.93 for predicting reference airborne lidar-estimated AGB and canopy cover. Given the estimated signal level expected over temperate forests by the ATLAS instrument and then parsing of the data into 100 m segments for the PCL simulation without noise, models outperformed those developed with the daytime and nighttime scenarios. With this dataset, R^2 values ranged from 0.79 to 0.93 with low RMSE values; between 6% and 8% for predicting canopy cover and 19.16 Mg/ha to 22.14 Mg/ha for estimating AGB. However, the presence of background noise represents a challenge for extracting signal photons (Glenn et al. 2016; Neuenschwander and Magruder 2016) which, among other factors, can impact the accuracy of the ATL08 product. This renders an examination of simulated PCL data with expected noise as well as the photon detection rate essential. The use of data with noise levels for daytime and nighttime operation times that were processed to report metrics for 100 m segments provided greater insights about ICESat-2's ATL08 for assessing forest parameters under temperate forest settings.

Solar background noise is greatly minimized with nighttime operation of a PCL system such as ATLAS (Popescu et al. 2018) so more accurate results were expected with the nighttime scenario than with the daytime scenario. In an evaluation of the noise

filtering and photon classification algorithm to process simulated PCL data, Popescu et al. (2018) indicated difficulty in differentiating the additional noise photons assigned to the daytime scenario from the signal photons. Popescu et al. (2018) demonstrate higher signal-to-noise ratios with nighttime scenarios, indicated a substantially smaller proportion of photons filtered due to fewer noise photons in these scenarios and reported improved accuracy of ground and TOC measurements derived from filtered and classified nighttime data. As a result, nighttime scenarios resulted in more accurate canopy height estimates than corresponding daytime scenarios. Canopy heights and associated height metrics such as height percentiles are instrumental in predicting forest biophysical parameters, including AGB (e.g. Popescu 2007; Sheridan et al. 2015), thus emphasizing the importance of obtaining accurate canopy heights from lidar systems. In this study, the ability of the nighttime scenario to predict AGB was remarkable, with results comparable to those generated from simulated data without the addition of noise and the challenge with estimating AGB with the daytime scenario was emphasized. For instance, with test data, the R^2 value was 0.79 for both the nighttime and the no noise scenarios, but was considerably lower for the daytime scenario, at 0.63 and a similar trend was exhibited for RMSEs. Findings from this study suggest that data obtained from nighttime operation of ICESat-2 may facilitate better characterization of forest structure than from its day counterpart.

Models developed for the daytime and nighttime scenarios both used the 10th and 90th height percentiles as well as canopy cover to predict AGB. Lidar studies (e.g. Ku et al. 2012; Sheridan et al. 2015), demonstrate the predictive power of height

percentiles in regression models and is therefore consistent with findings from this study. For instance, Sheridan et al. (2015) indicated the 90th height percentile calculated from discrete return airborne lidar data as one of the best predictors of gross volume (gV) of Pacific Northwest Forest Inventory and Analysis (FIA) plots.

ICESat-2's ATL08 product will provide canopy heights and canopy cover estimates (Neuenschwander et al. 2017) which may be used to estimate AGB, as demonstrated in this study with processed simulated ICESat-2 data. Also, even though canopy cover will be provided in the ATL08 product, other available metrics should be investigated for improving estimates. Between the two canopy cover models, height percentiles, canopy cover, maximum height and canopy density were important predictors. Simulated PCL-derived canopy cover was only used for predicting canopy cover at the 2 m threshold for the nighttime scenario, while height and density metrics were used in other canopy cover models for both the daytime and nighttime scenarios. Additionally, canopy cover models for the nighttime scenario consistently outperformed those for the daytime scenario, with higher R^2 values and lower RMSEs reported using the nighttime data.

Results presented will be used to facilitate the examination of approaches for extrapolating the data to provide wall-to-wall coverage. The performance of AGB and forest cover models, particularly with the nighttime scenario, suggests potential for AGB and forest cover mapping respectively and a chance to simultaneously explore synergistic approaches for upscaling ICESat-2 data. For instance, Glenn et al. (2016) demonstrated potential synergies between Landsat-2 and ICESat-2 for spatially

contiguous vegetation cover mapping. The combination of lidar metrics with spectral metrics from Landsat 8 Operational Land Imager (OLI) resulted in better vegetation cover predictions (Glenn et al. 2016). Similarly, biomass predictions were found to be improved by combining lidar and optical sensors (Glenn et al. 2016).

This study investigated the prediction of AGB and canopy cover over a study site where terrain effects are negligible and potentially serve as a reference (Popescu et al. 2011) of ICESat-2's predictive ability without errors associated with contrasting topographical conditions. The study site consists of primarily pine stands and mixed hardwoods growing on gentle slopes and results are expected to vary considerably under different forest ecosystems and with greater topographic variation. The predicted number of signal photons applicable to forests in study area is highest compared to signal predictions for boreal and tropical forests from the ATLAS instrument (Neuenschwander and Magruder 2016). Based on design cases developed by ICESat-2 SDT, the ATLAS instrument will detect up to 2 returns per outgoing laser pulse over boreal forest and either 0 or 1 return over tropical forest, while 3 returns are possible for the forest used in this simulation. Neuenschwander and Magruder (2016) evaluated simulated ICESat-2 data against airborne lidar data and found that the precision of terrain heights decrease with an increase in vegetation due to the limited number of ground returns from dense forest ecosystems. Consequently, larger errors in terrain and canopy height retrievals were reported for deciduous hardwoods and tropical forests, than for wooded savanna and boreal forest ecosystems. Popescu et al. (2018) evaluated performance of their noise filtering and photon classification algorithm using simulated PCL and MABEL data to

retrieve canopy heights under different vegetation conditions and reported that results were least accurate for tropical forests where returned photon density was lowest. These findings are expected to have implications, in terms of accuracy, of the ATL08 product and subsequent AGB modelling.

2.5. Conclusions

This study used simulated ICESat-2 data for assessing forest parameters and determining AGB. Metrics calculated from processed simulated ICESat-2 data for 100 m segments along proposed ICESat-2 tracks were evaluated for predicting AGB and canopy cover. The R-squared values obtained from multiple regression models using daytime and nighttime scenarios, ranging from 0.56 to 0.83, highlight the predictive ability of simulated PCL-derived metrics to characterize forest vegetation. These findings also suggest the potential of ICESat-2 for estimating AGB and canopy cover under temperate forest settings, given the expected photon detection rate of ATLAS, processing of simulated data scenarios with noise levels for daytime and nighttime operation, and especially in the data format provided by prospective standard ICESat-2 data vegetation products, e.g., with metrics reported over 100 m of vegetation profiles.

Forthcoming research will focus on further validation of these results and methods for upscaling estimates. While this study focused on Sam Houston National Forest, results can be extrapolated and scaled-up to regional and continental spatial extents. In doing so, variations in forest types and topography will need to be considered as these factors influence the accuracy of estimates.

2.6. References

Abdalati, W., Zwally, H.J., Bindschadler, R., Csatho, B., Farrell, S.L., Fricker, H.A., Harding, D., Kwok, R., Lefsky, M., Markus, T., Marshak, A., Neumann, T., Palm, S., Schutz, B., Smith, B., Spinhirne, J., & Webb, C. (2010). The ICESat-2 Laser Altimetry Mission. *Proceedings of the Ieee, 98*, 735-751.

Blair, J.B., & Hofton, M.A. (1999). Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters, 26*, 2509-2512.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., & Kimes, D.S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment, 112*, 3876-3890.

Brown, M., Escobar, V., Carroll, M., Neumann, T. , & Jasinski, M. 2013. Ice, Cloud, and Land Elevation Satellite 2 (ICESat-2) applications plan. http://icesat.gsfc.nasa.gov/icesat2/applications/ICESat-2_Applications_Plan_v8.pdf (accessed: 06.04.18).

Campbell, J.B., & Wynne, R.H. (2011). *Introduction to Remote Sensing, Fifth Edition*. New York, New York: Guilford Publications.

Carreiras, J.M.B., Shaun, Q.G., Toan, T.L., Minh, D.H.T., Saatchi, S.S., Carvalhais, N., Reichstein, M., & Scipal, K. (2017). Coverage of high biomass forests by the ESA BIOMASS mission under defense restrictions. *Remote Sensing of Environment, 196*, 154-162.

Degnan, J.J. (2002). Photon-counting multikilohertz microlaser altimeters for airborne and spaceborne topographic measurements. *Journal of Geodynamics, 34*, 503-549.

Drake, J.B., Knox, R.G., Dubayah, R.O., Clark, D.B., Condit, R., Blair, J.B., & Hofton, M. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global Ecology and Biogeography, 12*, 147-159.

Farrell, S.L., Laxon, S.W., McAdoo, D.C., Yi, D.H., & Zwally, H.J. (2009). Five years of Arctic sea ice freeboard measurements from the Ice, Cloud and land Elevation Satellite. *Journal of Geophysical Research-Oceans, 114*, 17.

Glenn, N.F., Neuenschwander, A., Vierling, L.A., Spaete, L., Li, A.H., Shinneman, D.J., Pilliod, D.S., Arkle, R.S., & McIlroy, S.K. (2016). Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. *Remote Sensing of Environment, 185*, 233-242.

Gwenzi, D., Lefsky, M.A., Suchdeo, V.P., & Harding, D.J. (2016). Prospects of the ICESat-2 laser altimetry mission for savanna ecosystem structural studies based on airborne simulation data. *Isprs Journal of Photogrammetry and Remote Sensing, 118*, 68-82.

Hall, F.G., Bergen, K., Blair, J.B., Dubayah, R., Houghton, R., Hurtt, G., Kellndorfer, J., Lefsky, M., Ranson, J., Saatchi, S., Shugart, H.H., & Wickland, D. (2011). Characterizing 3D vegetation structure from space: Mission requirements. *Remote Sensing of Environment, 115*, 2753-2775.

Hese, S., Lucht, W., Schmullius, C., Barnsley, M., Dubayah, R., Knorr, D., Neumann, K., Riedel, T., & Schroter, K. (2005). Global biomass mapping for an improved understanding of the CO2 balance - the Earth observation mission Carbon-3D. *Remote Sensing of Environment, 94*, 94-104.

Holmgren, J. (2004). Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research, 19*, 543-553.

Houghton, R.A., Hall, F., & Goetz, S.J. (2009). Importance of biomass in the global carbon cycle. *Journal of Geophysical Research-Biogeosciences, 114*, 13.

Homer, C., Dewitz, J., Yang, L.M., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the Conterminous United States - Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering and Remote Sensing, 81*, 345-354.

Isenburg, M. (2015). LASTools: award-winning software for rapid LiDAR processing. http://lastools.org (accessed: 05.11.18).

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., & Birdsey, R.A. (2003). National-scale biomass estimators for United States tree species. *Forest Science, 49*, 12-35.

Kankare, V., Liang, X., Vastaranta, M., Yu, X., Holopainen, M., & Hyyppä, J. (2015). Diameter distribution estimation with laser scanning based multisource single tree inventory. *Isprs Journal of Photogrammetry and Remote Sensing, 108*, 161-171.

Khan, S.A., Kjaer, K.H., Bevis, M., Bamber, J.L., Wahr, J., Kjeldsen, K.K., Bjork, A.A., Korsgaard, N.J., Stearns, L.A., van den Broeke, M.R., Liu, L., Larsen, N.K., & Muresan,

I.S. (2014). Sustained mass loss of the northeast Greenland ice sheet triggered by regional warming. *Nature Climate Change, 4*, 292-299.

Ku, N.W., Popescu, S.C., Ansley, R.J., Perotto-Baldivieso, H.L., & Filippi, A.M. (2012). Assessment of Available Rangeland Woody Plant Biomass with a Terrestrial Lidar System. *Photogrammetric Engineering and Remote Sensing, 78*, 349-361.

Lefsky, M.A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters, 37*, 5.

Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002). Lidar remote sensing for ecosystem studies. *Bioscience, 52*, 19-30.

Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B., & Hunter, M.O. (2007). Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing, 1*, 18.

Leigh, H.W., Magruder, L.A., Carabajal, C.C., Saba, J.L., & McGarry, J.F. (2015). Development of Onboard Digital Elevation and Relief Databases for ICESat-2. *Ieee Transactions on Geoscience and Remote Sensing, 53*, 2011-2020.

Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., & Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment, 115*, 2850-2860.

Li, L., Guo, Q.H., Tao, S.L., Kelly, M.G., & Xu, G.C. (2015). Lidar with multi-temporal MODIS provide a means to upscale predictions of forest biomass. *Isprs Journal of Photogrammetry and Remote Sensing, 102*, 198-208.

Liang, X.L., Kankare, V., Hyyppa, J., Wang, Y.S., Kukko, A., Haggren, H., Yu, X.W., Kaartinen, H., Jaakkola, A., Guan, F.Y., Holopainen, M., & Vastaranta, M. (2016). Terrestrial laser scanning in forest inventories. *Isprs Journal of Photogrammetry and Remote Sensing, 115*, 63-77.

Lu, D.S. (2006). The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing, 27*, 1297-1328.

Margolis, H.A., Nelson, R.F., Montesano, P.M., Beaudoin, A., Sun, G., Andersen, H.-E., & Wulder, M.A. (2015). Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America1. *Canadian Journal of Forest Research, 45*, 838-855.

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., Popescu, S., Shum, C.K., Schutz, B.E., Smith, B., Yang, Y., & Zwally, J. (2017). The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sensing of Environment, 190*, 260-273.

Martino, A. (2010). ATLAS Performance Spreadsheet. http://icesat.gsfc.nasa.gov/icesat2/data/sigma/sigma_data.php (accessed: 06.04.18).

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.

McKinley, D.C., Ryan, M.G., Birdsey, R.A., Giardina, C.P., Harmon, M.E., Heath, L.S., Houghton, R.A., Jackson, R.B., Morrison, J.F., Murray, B.C., Pataki, D.E., & Skog, K.E. (2011). A synthesis of current knowledge on forests and carbon storage in the United States. *Ecological Applications, 21*, 1902-1924

Montesano, P.M., Rosette, J., Sun, G., North, P., Nelson, R.F., Dubayah, R.O., Ranson, K.J., & Kharuk, V. (2015). The uncertainty of biomass estimates from modeled ICESat-2 returns across a boreal forest gradient. *Remote Sensing of Environment, 158*, 95-109.

Moussavi, M.S., Abdalati, W., Scambos, T., & Neuenschwander, A. (2014). Applicability of an automatic surface detection approach to micropulse photon-counting lidar altimetry data: implications for canopy height retrieval from future ICESat-2 data. *International Journal of Remote Sensing, 35*, 5263-5279.

MRLC (2017). Multi-resolution land characteristics consortium (MRLC): national land cover database. http://www.mrlc.gov/ (accessed: 02.09.17).

National Aeronautics and Space Administration, 2017. ICESat & ICESat-2. http://icesat.gsfc.nasa.gov/ (accessed: 06.04.18).

Neigh, C.S.R., Nelson, R.F., Ranson, K.J., Margolis, H.A., Montesano, P.M., Sun, G.Q., Kharuk, V., Naesset, E., Wulder, M.A., & Andersen, H.E. (2013). Taking stock of

circumboreal forest carbon with ground measurements, airborne and spaceborne LiDAR. *Remote Sensing of Environment, 137*, 274-287.

Nelson, R., Boudreau, J., Gregoire, T.G., Margolis, H., Naesset, E., Gobakken, T., & Stahl, G. (2009). Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere, 39*, 862-881.

Nelson, R., Gobakken, T., Næsset, E., Gregoire, T.G., Ståhl, G., Holm, S., & Flewelling, J. (2012). Lidar sampling — Using an airborne profiler to estimate forest biomass in Hedmark County, Norway. *Remote Sensing of Environment, 123*, 563-578.

Nelson, R., Margolis, H., Montesano, P., Sun, G., Cook, B., Corp, L., Andersen, H.-E., deJong, B., Pellat, F.P., Fickel, T., Kauffman, J., & 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.

Neuenschwander, A.L., & Magruder, L.A. (2016). The Potential Impact of Vertical Sampling Uncertainty on ICESat-2/ATLAS Terrain and Canopy Height Retrievals for Multiple Ecosystems. *Remote Sensing, 8*, 16.

Neuenschwander, A., Popescu, S., Nelson, R., Harding, D., Pitts, K., Pederson, D., & Sheridan, R. (2017) *ICE, CLOUD, and Land Elevation Satellite (ICESat-2) Algorithm Theoretical Basis Document (ATBD) for land-vegetation along-track products (ATL08)*, 78p.

Popescu, S.C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy, 31*, 646-655.

Popescu, S.C., & Wynne, R.H. (2004). Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering and Remote Sensing, 70,* 589-604.

Popescu, S.C., Wynne, R.H., & Nelson, R.F. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing, 29*, 564-577.

Popescu, S.C., & Zhao, K. (2008). A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment, 112*, 767-781.

Popescu, S.C., Zhao, K., Neuenschwander, A., & Lin, C. (2011). Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environment, 115*, 2786-2797.

Popescu, S.C., Zhou, T., Nelson, R., Neuenschwander, A., Sheridan, R., Narine, L., & Walsh, K.M. (2018). Photon counting LiDAR: An adaptive ground and canopy height retrieval algorithm for ICESat-2 data. *Remote Sensing of Environment, 208*, 154-170.

Pritchard, H.D., Arthern, R.J., Vaughan, D.G., & Edwards, L.A. (2009). Extensive dynamic thinning on the margins of the Greenland and Antarctic ice sheets. *Nature, 461*, 971-975.

Rahlf, J., Breidenbach, J., Solberg, S., Naesset, E., & Astrup, R. (2014). Comparison of four types of 3D data for timber volume estimation. *Remote Sensing of Environment, 155*, 325-333.

Rosen, P., Hensley, S., Shaffer, S., Edelstein, W., Kim, Y., Kumar, R., Misra, T., Bhan, R., Satish, R., Sagi, R., & Ieee (2016). An update on the NASA-ISRO dual-frequency DBF SAR (NISAR) mission. *2016 Ieee International Geoscience and Remote Sensing Symposium (Igarss)*, 2106-2108.

Sessa, R., & Dolman, H. (Eds.). (2008). *Terrestrial essential climate variables for climate change assessment, mitigation and adaptation*. Rome: FAO GTOS-52.

Sheridan, R., Popescu, S., Gatziolis, D., Morgan, C., & Ku, N.-W. (2015). Modeling Forest Aboveground Biomass and Volume Using Airborne LiDAR Metrics and Forest Inventory and Analysis Data in the Pacific Northwest. *Remote Sensing, 7*, 229-255.

Simard, M., Pinto, N., Fisher, J.B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research-Biogeosciences, 116*, 12.

Spinhirne, J.D., Palm, S.P., Hart, W.D., Hlavka, D.L., & Welton, E.J. (2005). Cloud and aerosol measurements from GLAS: Overview and initial results. *Geophysical Research Letters, 32*, 5.

Srinivasan, S., Popescu, S.C., Eriksson, M., Sheridan, R.D., & Ku, N.-W. (2014). Multi-temporal terrestrial laser scanning for modeling tree biomass change. *Forest Ecology and Management, 318*, 304-317.

Stysley, P.R., Coyle, D.B., Clarke, G.B., Frese, E., Blalock, G., Morey, P., Kay, R.B., Poulios, D., & Hersh, M. (2016). Laser Production for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar. In, *Conference on Laser Radar Technology and Applications XXI*. Baltimore, MD: Spie-Int Soc Optical Engineering.

Swatantran, A., Tang, H., Barrett, T., DeCola, P., & Dubayah, R. (2016). Rapid, High-Resolution Forest Structure and Terrain Mapping over Large Areas using Single Photon Lidar. *Scientific Reports, 6*.

Texas A&M Forest Service (2015). East Texas Forestlands, 2012. http://texasforestservice.tamu.edu/fia/publications/ (accessed: 05.10.18).

The MathWorks, Inc. (2015). MATLAB and Statistics Toolbox, Release R2015b. Natick, Massachusetts, United States.

USDA Farm Service Agency. (2015). NAIP Imagery. http://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/ (accessed: 02.09.17).

USDA Forest Service. (2014). First order lidar metrics: A supporting document for lidar deliverables. http://www.fs.fed.us/eng/rsac/lidar_training/pdf/LidarMetricsDescriptionOfDeliverables_Generic_12_15_14.pdf (accessed: 06.04.18).

USDA Forest Service. (2016). Forest inventory and analysis. http://www.nrs.fs.fed.us/fia/data-tools/state-reports/glossary/default.asp (accessed: 06.04.18).

Zwally, H.J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S., Spinhirne, J., & Thomas, R. (2002). ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics, 34*, 405-445.

3. MAPPING FOREST ABOVEGROUND BIOMASS WITH A SIMULATED ICESAT-2 VEGETATION CANOPY PRODUCT AND LANDSAT DATA[*]

3.1. Introduction

Accurate and spatially complete assessments of AGB are indicative of the extent to which forests contribute to the global carbon budget and can reduce uncertainties with the quantity and distribution of terrestrial carbon stocks (Houghton 2007; Houghton et al. 2007; Goetz & Dubayah 2011; Montesano et al. 2015). AGB carbon is primarily estimated using field inventory data where direct measurements such as tree diameters, are converted to AGB using biomass expansion factors or allometric regression equations (Brown 2002). These biomass regression equations are developed using allometric data collected from a sample of trees that have been destructively harvested and also usually over limited areas (Brown 2002; Popescu 2007). Combined with field measurements of AGB, remote sensing data can be used to reliably estimate AGB over multiple spatial scales (Popescu 2007; Nelson et al. 2017). Spaceborne light detection and ranging (lidar) instruments, in particular, are capable of providing measurements on a global scale and across areas that are otherwise difficult to access or where cost is prohibitive (Nelson et al. 2017). Reliable estimations of canopy height, basal area, AGB and aboveground carbon have been derived with spaceborne lidar (Nelson et al. 2017).

---

[*] Reprinted with permission from “Mapping forest aboveground biomass with a simulated ICESat-2 vegetation canopy product and Landsat data” by Lana L. Narine, Sorin Popescu, Tan Zhou, Shruthi Srinivasan, and Kaitlin Harbeck, 2019. *Annals of Forest Research*, 62(1), 1-17, DOI: 10.15287/afr.2018.1163, Copyright [2019] by "Marin Drăcea" National Research-Development Institute in Forestry.

While spatially continuous, global measurements from lidar are not yet possible, optical remotely sensed data can be integrated to generate spatially complete coverage of AGB and forest carbon at such scales (Hu et al. 2016).

The Ice, Cloud, and land Elevation Satellite (ICESat) operated from 2003 to 2010 and carried the Geoscience Laser Altimeter System (GLAS), a waveform lidar system, which provided global elevation data during the course of the mission (Zwally et al. 2002). Its primary objective was to capture elevation changes of the earth's polar ice sheets and reduce uncertainties with the ice sheet mass balance (Zwally et al. 2002). GLAS produced 1064 nm and 532 nm laser pulses at a frequency of 40 Hz and illuminated a spot on the ground measuring approximately 60 m in diameter every 172 m in the along-track direction (Zwally et al. 2002). The spaceborne data collected by GLAS offers multidisciplinary benefits, which includes global topography and vegetation canopy heights (Zwally et al. 2002; Schutz et al. 2005). The literature demonstrates the utility of GLAS data for estimating forest heights (Baghdadi et al. 2014; Harding & Carabajal 2005; Lefsky et al. 2007; Simard et al. 2011;) and AGB (Lefsky et al. 2005; Nelson et al. 2017). Furthermore, to overcome the spatial discontinuity of the lidar measurements, several studies have demonstrated approaches that integrate data from spaceborne multispectral sensors to generate wall-to-wall maps of AGB (Chi et al. 2015; Duncanson et al. 2010; Hu et al. 2016; Hudak et al. 2002). Essentially, the use of satellite imagery, such as data from Landsat sensors and from the moderate-resolution imaging spectroradiometer (MODIS) has been widely demonstrated across large-scale mapping studies.

With the recent launch of ICESat-2 on September 15th, 2018 (NASA 2017), up-to-date vegetation data at near-global scales and new prospects for vegetation mapping, will be available. This follow-on 3-year mission has been designed to overcome some of the limitations associated with ICESat to provide improved sampling and increased spatial coverage, as described in Markus et al. (2017). Specifically, ICESat-2's photon-counting lidar (PCL) instrument, the Advanced Laser Altimeter System (ATLAS), splits a 532 nm laser pulse into three pairs of beams spaced 3.3 km apart with a 90 m pair spacing (Gwenzi and Lefsky 2014; Markus et al. 2017). It operates at an increased repetition rate of 10 kHz, facilitating denser sampling and producing smaller footprints than GLAS, measuring 14 m and up to 17 m every 0.7 m along the ground track (Leigh et al. 2015; Markus et al. 2017). Since vegetation has lower reflectance than ice surfaces at the 532 nm wavelength, approximately 1/3 to 1/9 of the photon returns from ice and snow surfaces is expected for terrestrial surfaces (Neuenschwander and Magruder 2016). The number of signal photons per transmitted pulse will range between 0 and 3 over vegetated surfaces (Neuenschwander and Magruder 2016). However, differentiating signal photons from noise photons represents a challenge with photon counting systems like ATLAS, so effective filtering algorithms are crucial to deriving accurate estimates (Glenn et al. 2016). Noise levels also vary with operation times of ICESat-2, with less background noise associated with nighttime operation of ATLAS and increased potential for more accurate canopy height retrievals than from daytime retrievals (Popescu et al. 2018).

One of the ICESat-2's data products is the Land Water Vegetation Elevation or ATL08 product, which will provide terrain and canopy heights for non-polar regions (Neuenschwander et al. 2017). Terrain and height estimates will be provided at a fixed segment size of 100 m along the ground track. The ATL08 product will complement data from other space-based missions, like the Global Ecosystem Dynamics Investigation (GEDI) which was launched on December 5th, 2018 (NASA 2019) as well as data from optical sensors. The examination of approaches for AGB estimation and mapping could allow for development of an appropriate methodological framework for use as soon the ATL08 product becomes available. However, studies pertaining to the use of ICESat-2 data, especially its vegetation product for characterizing vegetation, are limited. For instance, Montesano et al. (2015) simulated ICESat-2 data and indicated difficulty in capturing vegetation structure where vegetation is sparse. An analysis of data from the Multiple Altimeter Beam Experimental Lidar (MABEL) instrument, ICESat-2's demonstrator instrument (Glenn et al. 2016), highlighted a synergistic approach between ICESat-2 and Landsat for improving AGB estimates. Gwenzi et al. (2016) indicated similar performance between ICESat and ICESat-2 using MABEL data, in estimating heights for savanna vegetation.

The aim of this study was to examine the use of ICESat-2 data, for generating wall-to-wall AGB coverage. The conceptual approach to AGB mapping has been highlighted in previous studies with GLAS data (Chi et al. 2015; Duncanson et al. 2010). However, considerable differences between the two instruments, such as measurement concept, spatial coverage and data products (Markus et al. 2017) warrant insights into

approaches for AGB mapping with expected data from ICESat-2. In this study, a synergistic approach with data similar to what will be provided by ICESat-2 and Landsat products was explored. With expected differences in noise levels associated with daytime and nighttime operation of ATLAS and associated impacts on canopy height estimation (Popescu et al. 2018), an examination of AGB mapping under different data scenarios are also included in this study. The primary goal of this study was to create wall-to-wall AGB maps at 30-m spatial resolution using simulated PCL-estimated AGB from known ICESat-2 track locations over Sam Houston National Forest in Texas and predictor variables from Landsat data. Three AGB maps were created from the following data scenarios: (i) Simulated ICESat-2 PCL vegetation product without the impact of noise (no noise scenario), (ii) Simulated ICESat-2 PCL vegetation product from data with noise levels associated with daytime operation of ICESat-2 (daytime scenario), and (iii) Simulated ICESat-2 PCL vegetation product from data with noise levels associated with nighttime operation of ICESat-2 (nighttime scenario). The same modeling technique was applied to generate AGB maps for the three scenarios with spatially explicit maps of model uncertainty produced for each corresponding AGB density map. ICESat-2 will provide global-scale sampling of forest resources and could potentially be leveraged to provide rich insight about the earth's forests. At the end of its mission, ICESat-2 data will be used to generate ATL18 products which consist of gridded terrain and canopy maps (Neuenschwander et al. 2017). This study ultimately serves to support upcoming AGB mapping with ICESat-2 and advance our understanding of the data for vegetation studies.

3.2. Materials and Methods

3.2.1. Study Area

The study area is located in SHNF in south-east Texas, USA (Latitude 30° 42′ N, Longitude 95° 23′ W) within the Pineywoods ecoregion and consists of approximately 48 km² of land. Summer temperatures average 28°C and the average winter temperature is around 12°C (USDA Forest Service 2018). The site consists of gentle slopes and elevations from 62 m and 105 m with a mean elevation of 85 m (Popescu 2007). Approximately 80% of the forests are classified as evergreen forest (Homer et al. 2015; MRLC 2017), which include Loblolly pine (*Pinus taeda*) plantations and old growth Loblolly pine stands (Popescu 2007) while common hardwoods found on floodplains and drier uplands include White Oak (*Quercus alba*) and Southern Red Oak (*Quercus falca*). Planned ICESat-2 track locations for the study site, provided by the mission's Science Definition Team (SDT) (Neuenschwander et al. 2017), are shown in Figure 5.

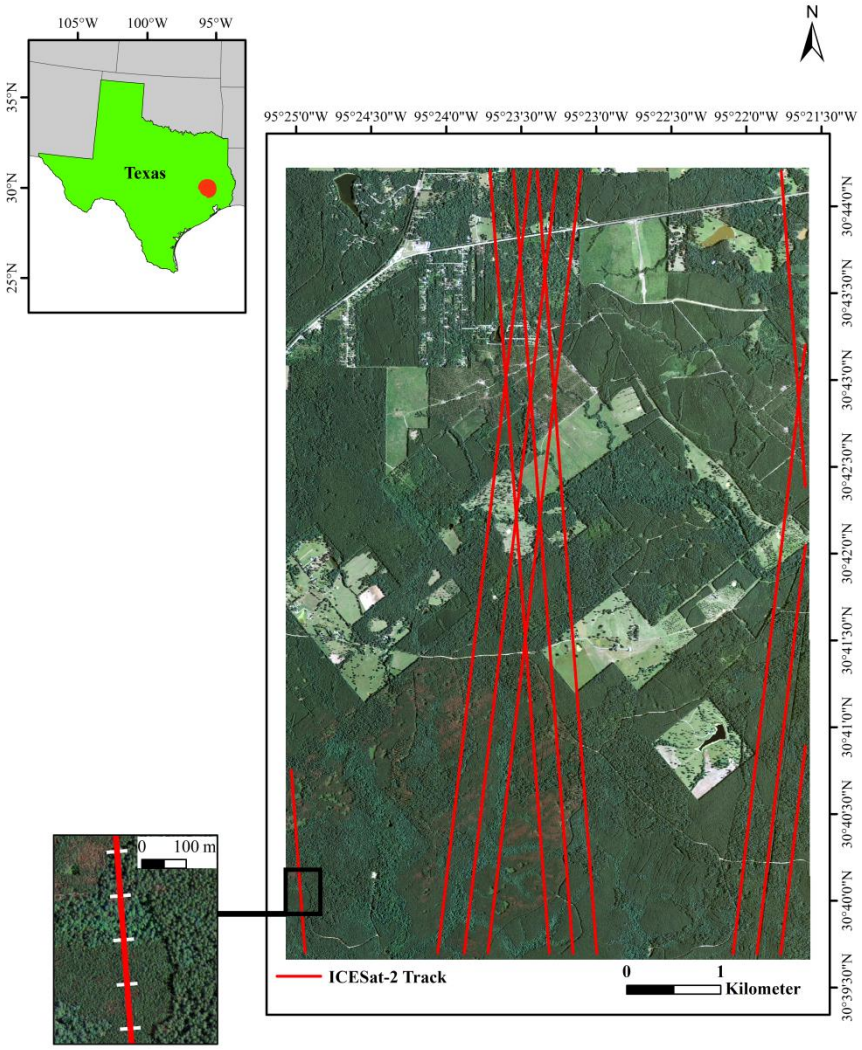


Figure 5. ICESat-2 tract locations overlaid on 2010 National Agriculture Imagery Program (NAIP) aerial imagery for the study area within SHNF, Texas (inset map, upper left corner) with demarcation of 100 m segments along-track on inset map, lower left corner

3.2.2. Simulated ICESat-2 Data Scenarios

Airborne lidar data collected in November 2010 were used to simulate ICESat-2 data along known ICESat-2 tracks over SHNF. The data were acquired along 12 flight

lines in the north-south direction and 19 lines flown from east to west, at around 600 m above-ground-level (AGL) and has a point density of 4 points per m^2.The airborne lidar data and simulated ICESat-2 footprints with a center-to-center spacing of 70 cm along each ground track were used as input for the simulation for the data scenarios. The simulation algorithm, described in detail in Neuenschwander and Magruder (2016), was developed by the ICESat-2 SDT. To summarize, four main steps were implemented [22]: (1) assemble heights from discrete return points within a 7 m radius of footprint centers to generate a pseudowaveform (Blair & Hofton 1999), (2) construct a height vector for each footprint, (3) randomly determine the number of photons (x) to sample per shot based on design cases developed by the mission's instrument team, and (4) randomly sample the vector of heights x times, weighted by the pseudo-waveform. The design cases developed by the mission's instrument team indicate the expected number of photons per footprint vary by vegetation type and can range from 0 to 3 returns per pulse (Neuenschwander & Magruder 2016). The mean number of signal photons per footprint for SHNF was modeled at 1.9, based on the ATLAS performance model for temperate forests (Martino 2010). To complete the simulation, background noise, representing anticipated solar background noise or from the atmosphere was added to the data. Since solar background noise is expected to be pronounced during daytime operation of ICESat-2 (Degnan 2002), the datasets were generated to reflect the different expected noise levels: (1) daytime scenario with expected noise levels for daytime operation of ICESat-2, and (2) nighttime scenario with noise levels based on night operation of ICESat-2. Noise photons are not discernable from signal photons so the

implementation of effective processing algorithms is critical in order to derive accurate forest measurements from PCL data. Noise filtering and photon classification algorithms (Popescu et al. 2018) were applied to the daytime and nighttime scenarios to derive top of canopy and ground estimates. As presented in Popescu et al. (2018), noise filtering entailed a multi-level approach for minimizing noise photons and moving overlapping windows and cubic spline interpolation were used to classify the remaining photons. In terms of accuracy of the algorithms, Popescu et al. (2018) reported average RMSE values of 1.83 m and 2.80 m for estimated ground elevation and 2.70 m and 3.59 m for canopy height estimations with the nighttime and daytime scenarios respectively. Following the application of the noise filtering and photon classification algorithms for daytime and nighttime scenarios investigated in this study, estimated canopy heights were then retrieved by subtracting estimated ground from top of canopy values.

The airborne lidar data for the study area was also processed to obtain above-ground-level heights and used for PCL simulation without the addition of background noise (no noise scenario). In total, three simulated PCL datasets (data scenarios) were generated for each planned ICESat-2 track over SHNF.

3.2.3. Reference AGB

Reference AGB estimates were calculated from a canopy height model (CHM) derived from airborne lidar data for SHNF. The CHM was used as input in a lidar software application called TreeVaW, to extract forest inventory parameters for individual trees (Popescu et al. 2003, Popescu & Wynne 2004). TreeVaW uses a

variable window technique with local maximum focal filtering to extract tree heights, where the window size is based on the relationships between tree heights and crown widths from field inventory data (Popescu 2007). Popescu & Wynne (2004) indicated that lidar measurements processed with this technique explained 97% of the variance associated with mean height of dominant pine trees. For SHNF, ground measured tree heights and corresponding crown diameters from a total of 705 pine trees and 603 deciduous trees from field inventories carried out in 2004 and 2009 were used to obtain TreeVaW coefficients to subsequently derive tree locations and measurements. The crown diameter of an individually located tree in the CHM was also estimated in TreeVaW, which involves averaging two values measured along perpendicular profiles from the center of a tree top by fitting a fourth-degree polynomial on each profile (Popescu et al. 2003, p. 564). Details of the processing approach implemented in TreeVaW are presented in Popescu et al. (2003) and Popescu & Wynne (2004). Output consisting of estimated diameter at breast height (dbh) measurements, were used to estimate tree-level AGB using a generalized biomass regression equation (Eq. (1)):

$$\text{AGB (kg)} = \exp(\beta_0 + \beta_1 \ln(\text{dbh})), \quad (1)$$

where dbh is the diameter at breast height (cm) and β_0 and β_1, are the parameters for the species group (Jenkins et al. 2003, p. 20). AGB for pines were calculated using the parameters for the "pine" species groups, with β_0 = -2.5356 and β_1= 2.4349. To calculate AGB for deciduous trees, parameters for the "hard maple/oak/hickory/beech" group were used, with β_0 = -2.0127 and β_1= 2.4342 (Jenkins et al. 2003, p. 20).

3.2.4. AGB Estimation from Simulated PCL-Derived Metrics

For each data scenario, 100 m segments along ICESat-2 profiles were extracted and maximum height, mean height, height percentiles, canopy cover and canopy density were calculated for each segment. A segment measuring 100 m in the along-track direction was chosen in order to be consistent with the format of the planned ATL08 product from ICESat-2 (Neuenschwander et al. 2017). A total of 121 segments over the study site were used, where one-third was randomly assigned for model testing and the remaining 85 segments were used for developing the models for estimating AGB. Linear regression models were used to relate the simulated PCL metrics for a subset of the segments to spatially coincident, airborne-lidar estimated AGB and the resulting models were evaluated with a separate test set. Details of the methodology used to estimate AGB and canopy cover from simulated PCL metrics are provided in Narine et al. (2019).

Using resulting AGB models, AGB density (Mg/ha) was estimated for the segments over SHNF and then applied to 30 m pixels to match Landsat TM pixels. Since different segment lengths traverse pixels, AGB was calculated based on the portion of a 100 m segment across a pixel and the value was extrapolated to represent the pixel size using the estimated area of a segment across a pixel. Pixels with segments measuring less than 7 m in a pixel were excluded from analysis and average pixel AGB was calculated in instances where there were two segments (parts) of equal lengths. The steps for assigning AGB to pixels were repeated for each scenario using the corresponding prediction equation developed from linear regression analysis.

3.2.5. Mapped Predictors

Landsat data is freely available and offers global coverage (Avitabile et al. 2012). Landsat data, specifically, Landsat 5 Thematic Mapper (TM) was chosen for use in this study to avoid temporal mismatch with the simulated data and reference AGB from airborne lidar data acquired in November, 2010. The data is provided at spatial resolution of 30 m and has 7 spectral bands, which include blue, green, red and near infra-red (NIR). A Landsat cloud-free scene encompassing the study site, from path/row 26/39, acquired in November, 2010 was downloaded from U.S. Geological Survey (USGS) Earth Explorer in GeoTIFF format and re-projected to UTM, WGS84, Zone 15N. The image was processed in ENVI to top-of-atmosphere (TOA) reflectance and vegetation indices were computed at the 30 m pixel size (Table 4). Land cover and canopy cover maps from the 2011 National Land Cover Database (NLCD) (Homer et al. 2015) were also downloaded and clipped to the extent of the study site for use as predictor variables. Land cover and canopy cover maps were georeferenced in the same projection as the Landsat TM image.

Table 4. Mapped predictor variables for RF regression

Variable	Description
Spectral Metrics from Landsat 5 TM	
Normalized Difference Vegetation Index (NDVI)	(NIR - Red) / (NIR + Red)
Enhanced Vegetation Index (EVI)	2.5 * ((NIR - Red) / (NIR + 6 * Red - 7.5 * Blue + 1))
Soil Adjusted Vegetation Index (SAVI)	((NIR - Red) / (NIR + Red + 0.5)) * (1.5)
Modified Soil Adjusted Vegetation Index (MSAVI)	(2 * NIR + 1 - sqrt ((2 * NIR + 1)² - 8 * (NIR - Red))) / 2
Land Cover	National Land Cover Database 2011 (NLCD 2011)
Canopy Cover	NLCD 2011 US Forest Service Tree Canopy Cover

3.2.6. AGB Mapping

The regression tree method, Random Forest (RF) (Breiman 2001), was used to model AGB for the three data scenarios using spectral metrics from Landsat 5 TM, landcover and canopy cover and estimate AGB for areas not overlaid by ICESat-2 tracks. RF is a nonparametric modeling technique and an established approach for mapping AGB using multisource data (Baccini et al. 2004; Chi et al. 2015; Houghton et al. 2007). Using a bootstrap sample of the training data, the best split is made at the root node using a random sample of the predictors (Liaw & Wiener 2002). Another sample of the variables is taken at the other node and the process is repeated until the regression tree grows as large as possible. The process is repeated with a new bootstrap sample and the data is predicted by averaging the predictions of all trees (Liaw & Wiener 2002). The un-sampled training data at each bootstrap iteration, called out-of-bag or OOB data, is used to generate OOB model predictions which are aggregated to calculate the error rate (Liaw & Wiener 2002). RF was carried out with the ModelMap R package (Freeman et al. 2018) which calls randomForest R package (Freeman et al. 2018). ModelMap was used to validate the model with OOB predictions on the training data and also with a separate test set and apply the model to create the final AGB map for each scenario (Freeman et al. 2018). About 70% of the data was randomly assigned to the training dataset and remaining 30% was allotted to the test set and used for model evaluation. As a result, the training and test set for each scenario consisted of 1448 and 620 30-m pixels respectively, equating to approximately 3% and 1% of the pixels which constitute the study area. A total of 1000 regression trees were built with the training data for each

scenario and results were used to assign AGB values to each 30 m pixel. The final AGB maps were masked using the land cover map, where AGB density in non-forest areas (NLCD classes: water, developed, barren, grassland/herbaceous, pasture and emergent herbaceous wetlands) was set to 0 Mg/ha.

Since trees in a RF model are independent, the standard deviation of individual-tree predictions can be calculated at the pixel-level to produce a measure of prediction variability (Freeman et al. 2016). According to Freeman et al. (2016), standard deviation of predictions for trees can be used as a measure of uncertainty in resulting maps. High uncertainty values indicate a lack of agreement among trees (Freeman et al. 2016, p. 15), with some trees predicting high AGB and others low AGB values while low uncertainty values translate to agreement in predictions from individual trees for a pixel. RF uncertainty maps corresponding to final AGB maps were generated.

Results were also compared with AGB estimated from airborne lidar data over the study. The reference AGB estimates derived from airborne lidar data were aggregated by summation at a 30 m resolution and pixel AGB values were converted to AGB density (Mg/ha). Cells spatially coincident with the original test datasets were then selected and predicted AGB were compared with the reference AGB pixels. AGB predictions for the three data scenarios were compared using RMSE and R^2 metrics.

3.3. Results

3.3.1. Estimated AGB from Simulated PCL Metrics

Details of results from regression analysis used to relate simulated PCL height metrics, canopy cover and canopy density for 100 m segments to reference airborne lidar-derived AGB are presented in Narine et al. (2019). To summarize, models had high accuracies and yielded RMSE values of 19.16 Mg/ha, 25.35 Mg/ha, and 19.23 Mg/ha for the no noise scenario, daytime scenario and nighttime scenario respectively. The prediction equation for the simulated dataset without noise used mean height and a canopy density variable for 15 m to 20 m height bin and explained 79% of the variance in airborne lidar-estimated AGB. The absence of noise photons and the applied photon detection rate for the forest being studied rendered this dataset the best case scenario. Given that background noise represents a challenge for photon counting systems (Glenn et al. 2016) estimation models associated with daytime and nighttime operation times offer more insight about the potential AGB estimation from ICESat-2 data. AGB models with the daytime and nighttime scenarios used the same variables; the 10th and 90th height percentiles and canopy cover calculated as the proportion of returns above 4.6 m. With the daytime scenario, the variables explained 63% of the variance in airborne lidar-derived AGB while the same variables calculated with the simulated nighttime dataset yielded a R^2 of 0.79.

3.3.2. RF Predicted AGB

Using the simulated PCL-estimated AGB as the dependent variable and spectral metrics, land cover and canopy cover, the RF regression tree models explained 49%, 39% and 47% of the variance with OOB data for the no noise, daytime and nighttime scenarios respectively. RMSEs estimated using OOB testing were 20.59 Mg/ha, 19.98 Mg/ha and 19.25 Mg/ha. With the test data, the predictive abilities of models, represented by the R^2 values, were 0.51, 0.42 and 0.49. Overall, the models tended to underestimate cells with high biomass (Figure 6). With test data, the RMSE values for the no noise and daytime scenarios were 19.69 Mg/ha and 19.67 Mg/ha, and slightly improved for the nighttime scenario, with a RMSE of 19.31 Mg/ha. The relationship between RF predicted AGB and AGB estimated from airborne lidar data over the study site yielded lower R^2 values and higher RMSEs and similar results for the three scenarios. When rounded to the nearest whole number, the RMSE was 35 Mg/ha with the no noise, daytime and nighttime scenarios while the variance explained was 46% with the no noise and nighttime scenarios and 45% with the daytime scenario (Table 5). The most important predictor of AGB for the three data scenarios was canopy cover followed by NDVI and land cover. Variable importance was almost split between NDVI and land cover while SAVI and MSAVI were the least important predictors of AGB.

Average AGB values from final AGB maps (Figures 7-9) generated with the RF regression models were 53.05 Mg/ha, 50.73 Mg/ha and 52.39 Mg/ha for the no noise, daytime and nighttime scenarios respectively. In comparison, average AGB of forests in the training data were 52.09 Mg/ha, 51.21 Mg/ha, and 52.24 Mg/ha. With this data, AGB

ranged from 0 Mg/ha to 150.64 Mg/ha for the no noise scenario, 0 Mg/ha to 139.61 Mg/ha, and 0 Mg/ha to 119.75 Mg/ha for the daytime and nighttime scenarios respectively. AGB predictions for 30 m pixels were within the range of the data but tended to underestimate AGB. AGB density predictions of forests in the study site ranged from 0.72 Mg/ha to 105.41 Mg/ha for the no noise scenario, 6.62 Mg/ha to 83 Mg/ha and 1.97 Mg/ha to 83.55 Mg/ha for the daytime and nighttime scenarios respectively.

Visual assessments of AGB maps revealed similarities in the spatial distribution pattern of predicted AGB for the three scenarios (Figures 7-9). However, the daytime and nighttime scenarios appeared to have more areas with AGB densities ranging from 20-40 Mg/ha than the no noise scenario while the map for the daytime scenario exhibited the greatest predominance of relatively low AGB values. To demonstrate, predicted AGB estimates with the three scenarios for a highlighted portion of the study area, is shown on Figure 10. Within the highlighted extent, predicted AGB for the daytime and nighttime scenarios exhibited lower AGB estimates than the no noise scenario. Compared to the other scenarios, lower AGB estimates were most prevalent with the daytime scenario, represented by the 20-40 Mg/ha range. Overall, AGB maps correspond to vegetation trends present with higher AGB in the southern portion of the study site primarily occupied by mature pines and lower biomass in the northern parts, including areas covered by young pine stands. At the pixel level, predictions of AGB from individual trees that varied greatly led to a high standard deviation of the predictions. Corresponding RF uncertainty maps (Figures 7-9) for the scenarios highlight

pixels with particularly large differences in individual tree predictions, especially for the daytime scenario. RF uncertainties highlighted in Figure 10 emphasize the predominance of high standard deviation of AGB predictions for the daytime scenario, compared to the other data scenarios. In this area, the 10-20 Mg/ha range was most prevalent with the no noise scenario followed by the nighttime scenario. In comparison, there were substantially more pixels within the 20-30 Mg/ha range, indicative of greater uncertainty, for the daytime scenario (Figure 10).

Table 5. Test set error statistics from RF models predicting AGB and relationships with aggregated airborne lidar-derived estimates of AGB under three scenarios; no noise, daytime and nighttime scenarios

	RF Model Performance with Test Set		Relationship with Reference AGB	
Scenario	**R^2**	**RMSE**	**R^2**	**RMSE**
No Noise	0.51	19.69 Mg/ha	0.46	34.99 Mg/ha
Daytime	0.42	19.67 Mg/ha	0.45	35.27 Mg/ha
Nighttime	0.49	19.31 Mg/ha	0.46	34.93 Mg/ha

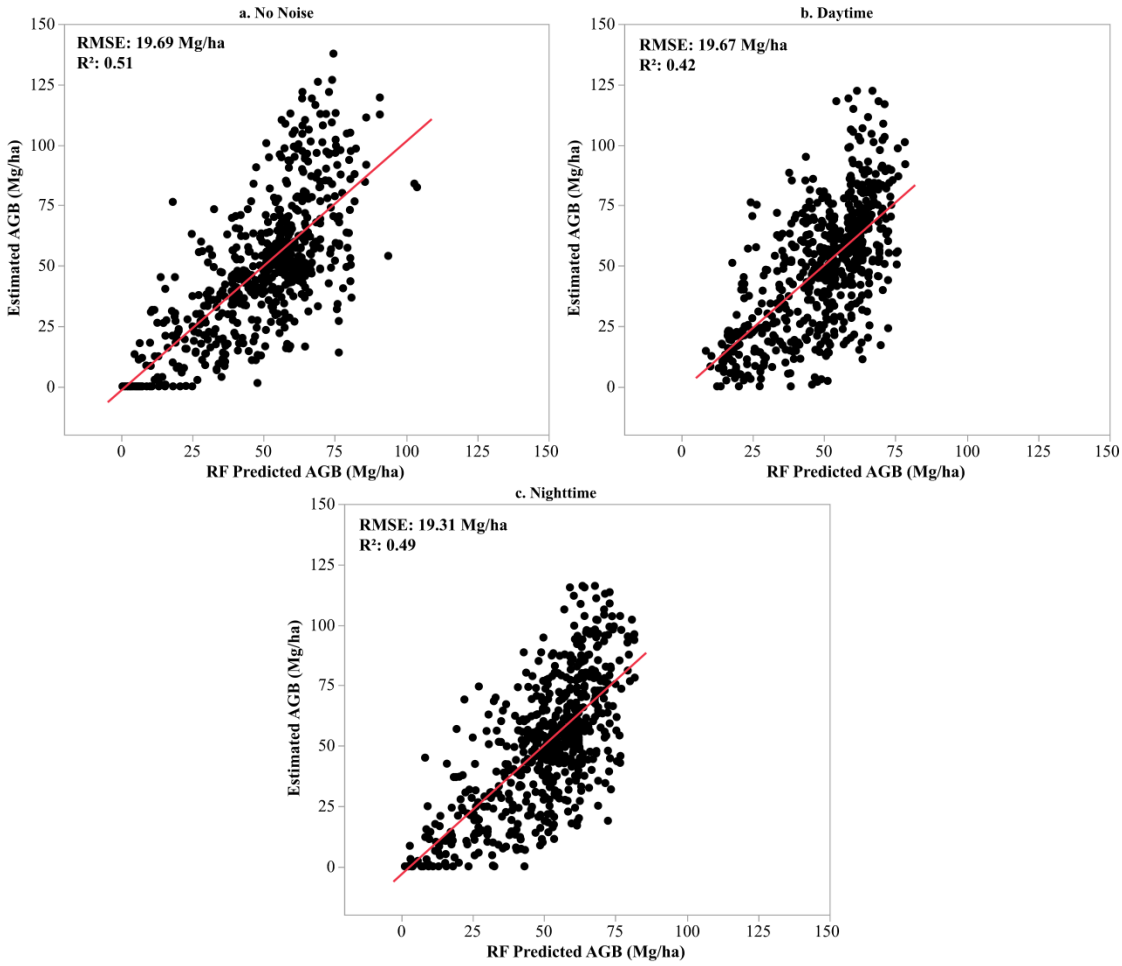


Figure 6. (a) Simulated PCL AGB estimated from linear regression vs RF predicted AGB with test data for the no noise scenario; (b) Simulated PCL AGB estimated from linear regression vs RF predicted AGB with test data for the daytime scenario (c) Simulated PCL AGB estimated from linear regression vs RF predicted AGB with test data for the nighttime scenario

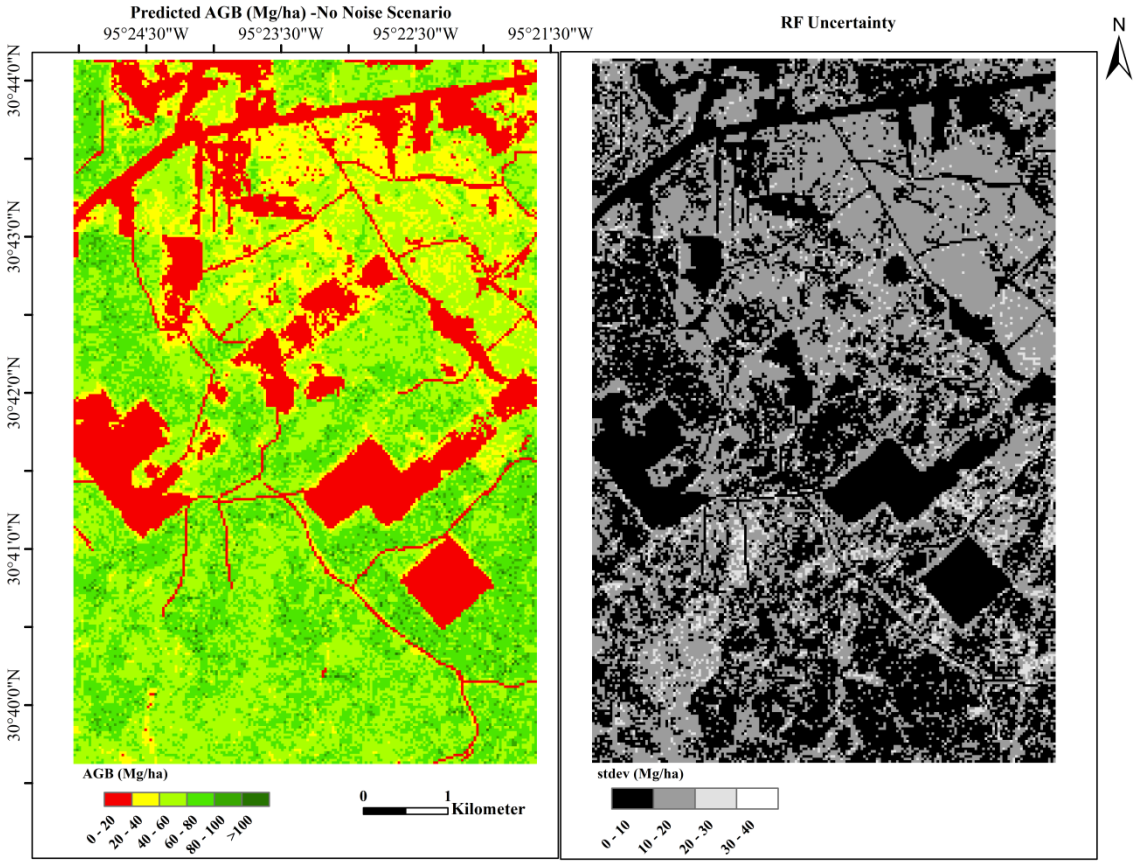


Figure 7. (a) RF predictions of AGB density at a 30 m spatial resolution for the no noise scenario; (b) RF uncertainty at a 30 m spatial resolution for the no noise scenario

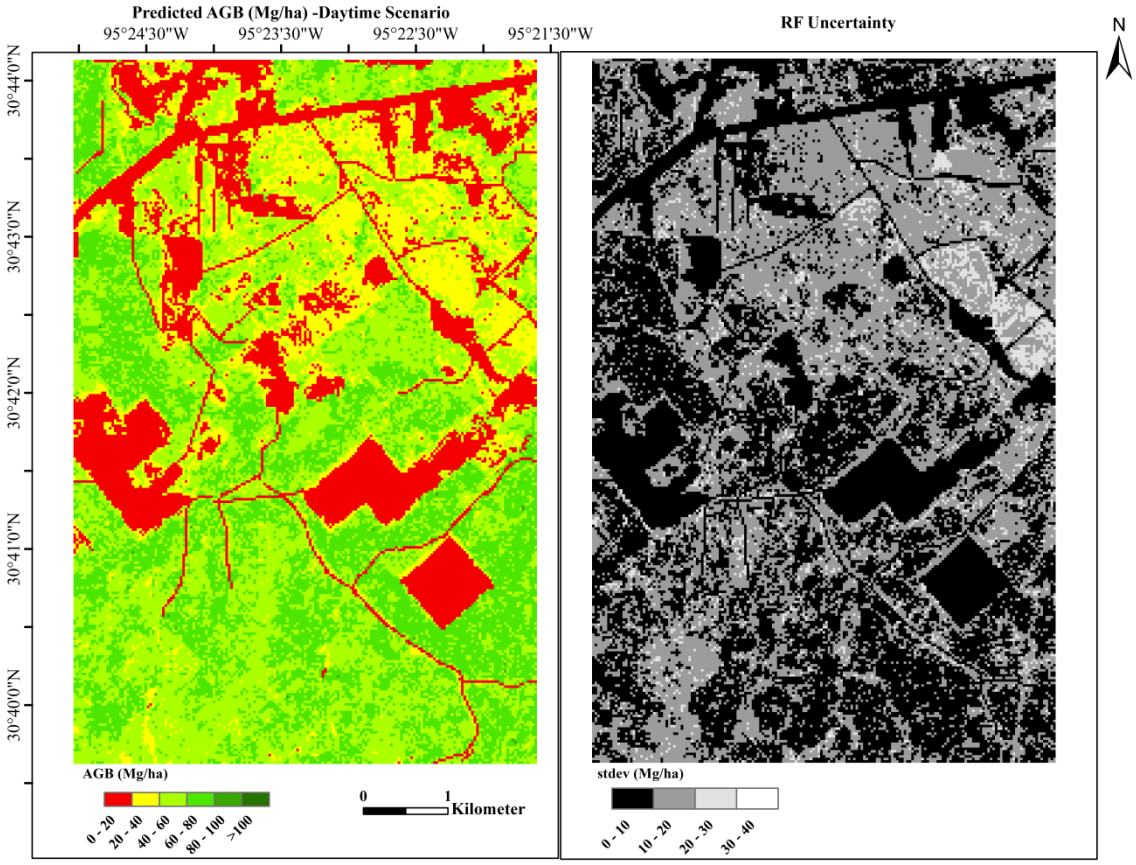


Figure 8. (a) RF predictions of AGB density at a 30 m spatial resolution for the daytime scenario; (b) RF uncertainty at a 30 m spatial resolution for the daytime scenario

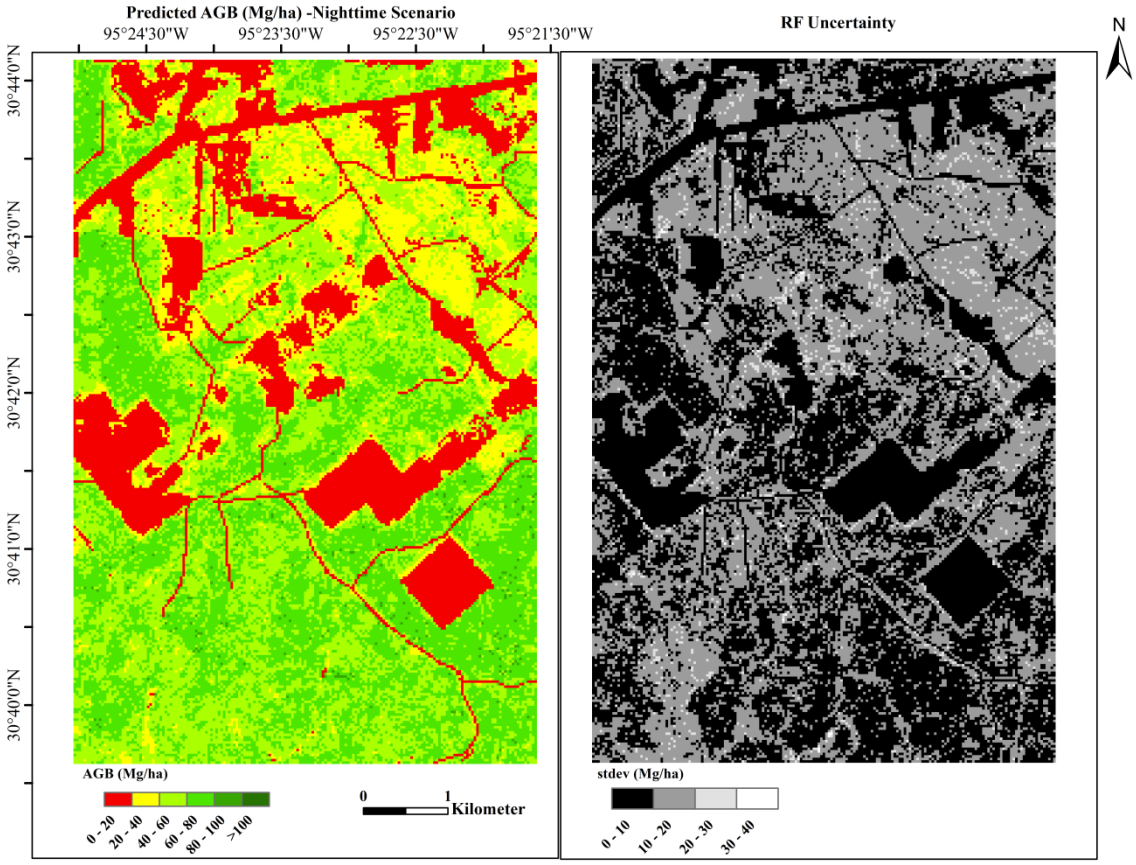


Figure 9. (a) RF predictions of AGB density at a 30 m spatial resolution for the nighttime scenario; (b) RF uncertainty at a 30 m spatial resolution for the nighttime scenario

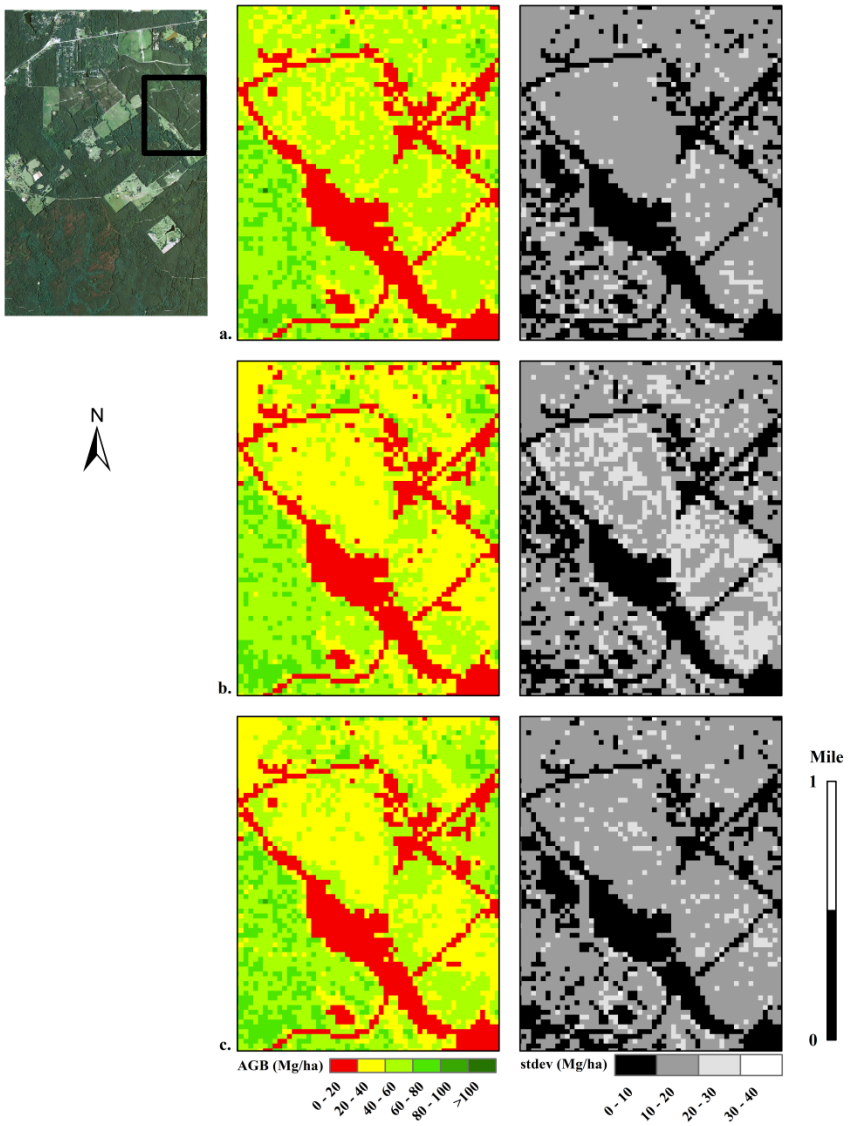


Figure 10. Spatial distribution of RF predicted AGB and corresponding AGB uncertainty for highlighted section in the study area (inset map, upper left corner) for the a. no noise scenario, b. daytime scenario, and c. nighttime scenario

3.4. Discussion

The mapping of AGB has been identified as an approach to advance the knowledge about the terrestrial aspect of the carbon cycle (Le Toan et al. 2011).

Consequently, is it necessary to develop approaches for characterizing the spatial distribution of AGB and forest carbon. ICESat-2 presents an opportunity to obtain large-scale coverage about vegetation through the collection of data along transects on the earth's surface. Despite the lack of spatially complete data that will be measured from this source, complete coverage offered by Landsat or MODIS sensors may be leveraged to achieve comprehensive estimates of forest attributes. This study presents one approach for deriving wall-to-wall coverage of AGB from ICESat-2. Two sets of models were used; one for modeling the relationship between simulated PCL metrics for 100 m segments and reference airborne lidar-derived AGB and another set for relating mapped predictor variables to simulated PCL-estimated AGB. Using nonparametric regression models, AGB density was extrapolated from 3% of the 30 m pixels which comprise the study site to achieve spatially explicit AGB values. While the study area was focused on the SHNF, this approach is applicable to scaling up to larger spatial extents. The use of 30 m map resolution is particularly useful for upscaling purposes with the availability of spatially comprehensive data, such as Landsat imagery and NLCD products. In doing so, other variables may also be investigated for improving AGB estimates. For example, information from optical imagery such as Tasseled cap indices, and ancillary data such as elevation and slope may be incorporated for regional scale mapping (Zald et al. 2016). Additionally, vegetation indices from multi-date satellite imagery could be investigated when upscaling to larger extents. To demonstrate, Li et al. (2015) highlighted the use of vegetation indices from multi-date imagery during the growing season for achieving more reliable results for regional AGB estimation. It is also important to note that

coarser map resolutions may provide good accuracies. To exemplify, Deo et al. (2018) investigated the impact of spatial resolution on the accuracy of regional scale AGB maps encompassing the eastern USA from Landsat-derived parameters and concluded that grid sizes up to 1 km provided reasonable accuracies. In this study, validation statistics for the AGB models indicated RMSEs ranging from 69.34 Mg/ha for the 30 m grid size, up to 80.78 Mg/ha with the 1000 km grid size but with similar correlation (r) results between predictions and the reference data (Deo et al. 2018).

Canopy cover has proven to be a valuable predictor of AGB in this study. Canopy cover calculated from simulated PCL segments was a significant predictor of airborne lidar-derived AGB and results from the RF models produced in this study indicated NLCD canopy cover as the most important predictor of AGB, regardless of the data scenario. Alternative approaches to generating AGB maps may include extrapolating canopy cover estimated in the ATL08 product and then using this as a predictor variable in RF models. A similar approach with GLAS parameters is presented in Hu et al. (2016), where RF was used to generate spatially continuous GLAS parameters and then combined with mapped predictors including NDVI and landcover, to estimate global AGB at a 1-km resolution. In the case of extrapolating canopy cover and then predicting AGB, the possibility for simultaneously producing maps of these parameters will be beneficial as both represent fundamental attributes of forest vegetation structure.

An advantage of using RF for modeling is the possibility of using OOB as part of model evaluation (Freeman et al. 2016). An evaluation of RF model performance for the

no noise, daytime and nighttime scenarios with a separate test set yielded similar results to those provided with OOB data; R^2 values were 0.51, 0.42, and 0.49, compared to 0.49, 0.39, and 0.47 with OOB testing. RMSE values also ranged between 19 Mg/ha and 21 Mg/ha with the OOB and independent test set. Overall, results from both sources emphasized the ability of the nighttime scenario to out-perform the daytime scenario. RF also provided an indicator of uncertainty in the resulting maps, estimated as the variability of predictions from 1000 independent trees used to estimate pixel AGB. A comparison of the AGB and AGB uncertainty maps for the three scenarios highlighted areas with lower AGB predictions and corresponding higher uncertainty ranges for the daytime versus the nighttime and no noise scenarios. Factors that may contribute to final AGB uncertainty include allometric models for reference AGB estimation, input data errors (e.g. NLCD, Landsat reflectance) and methods used for upscaling AGB to the pixel size for modeling. Prospective studies could involve a comprehensive uncertainty analysis of AGB estimates predicted from actual ICESat-2 data to ascertain error originating from different sources.

GEDI commenced its two-year mission in November, 2018 and was designed specifically to provide data for characterizing three-dimensional vegetation structure (NASA 2019). While ICESat-2 will collect measurements between 88° north and south latitudes during its three-year duration (Markus et al. 2017), GEDI will provide waveform observations between 52° north and 52° south latitudes (Marselis et al. 2016). Hence, the adoption of a synergistic approach between ICESat-2 and GEDI may support a more accurate quantification of forest attributes, and specifically forest AGB. Future

research should investigate potential synergies between ICESat-2 and GEDI for producing an AGB product.

3.5. Conclusions

Up-to-date and spatially explicit assessments of AGB density can support the monitoring of forest carbon, contributing to reduced uncertainties with the carbon budget and an improved understanding of changes in terrestrial carbon storage. In this study, we present an approach for mapping AGB with ICESat-2 which can be summarized as follows: (i) simulating and processing of ICESat-2 data to generate three data scenarios in a format similar to the ATL08 product, (ii) developing relationships between the simulated vegetation product data and airborne-lidar derived AGB, and then between predicted AGB and mapped predictors consisting of spectral metrics from Landsat 5 TM, landcover and canopy cover using RF, and (iii) mapping AGB at 30-m spatial resolution and producing a corresponding measure of uncertainty using the resulting RF models. Findings highlight canopy cover as the most important predictor of AGB and indicate similarities in the predictive capabilities of RF models for the three settings analyzed, and especially for the nighttime scenario which outperformed the daytime scenario. Overall, the methodology is conducive to achieving wall-to-wall AGB coverage at much larger extents and in doing so, other variables such as topographic, climatic, and spectral indices may be considered for improving AGB estimates. Efforts focused on the contributions of error from different sources would also be needed to ascertain final AGB uncertainties. In addition, the recent launch of GEDI will advance

the capability of assessing AGB, thus highlighting the need for investigations that support a synergistic approach between ICESat-2 and GEDI to characterize AGB and other forest attributes. Furthermore, while spatial coverage from both ICESat-2 and GEDI are not possible, the integration of spatially comprehensive data, like Landsat optical imagery, can potentially facilitate a more comprehensive and accurate AGB product.

3.6. References

Avitabile V., Baccini A., Friedl M.A., & Schmullius C. (2012). Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sensing of Environment*, 117: 366-380.

Baccini A., Friedl M.A., Woodcock C.E., Warbington R. (2004). Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters*, 31: 1-4.

Baghdadi N., le Maire G., Fayad I., Bailly J.S., Nouvellon Y., Lemos C., & Hakamada, R. (2014). Estimation of forest height and aboveground biomass from ICESat-2/GLAS data in eucaluptus plantations in Brazil. In 2014 *IEEE International Geoscience and Remote Sensing Symposium*, 13-18th July 2014, Quebec City, QC, Canada. IEEE, pp. 725-728.

Blair J.B. & Hofton M.A. (1999). Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters*, 26: 2509-2512.

Breiman L. (2001). Random forests. *Machine Learning*, 45: 5-32.

Brown S. (2002). Measuring carbon in forests: current status and future challenges. *Environmental Pollution*, 116: 363-372.

Chi H., Sun G.Q., Huang J.L., Guo Z.F., Ni W.J. & Fu A.M. (2015). National forest aboveground biomass mapping from ICESat/GLAS data and MODIS imagery in China. *Remote Sensing*, 7: 5534-5564.

Degnan J.J. (2002). Photon-counting multikilohertz microlaser altimeters for airborne and spaceborne topographic measurements. *Journal of Geodynamics*, 34: 503-549.

Deo R.K., Domke G.M., Russell M.B., Woodall C.W. & Andersen H.E. (2018). Evaluating the influence of spatial resolution of Landsat predictors on the accuracy of biomass models for large-area estimation across the eastern USA. *Environmental Research Letters*, 13(5): 1-9.

Duncanson L.I., Niemann K.O. & Wulder M.A., 2010. Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sensing of Environment*, 114: 138-154.

Freeman E.A., Frescino T.S. & Moisen G.G. (2018). ModelMap: an R Package for Model Creation and Map Production. Available online: https://cran.r-project.org/web/packages/ModelMap/vignettes/VModelMap.pdf (accessed: 10.05.18).

Freeman E.A., Moisen G.G., Coulston J.W. & Wilson B.T. (2016). Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Canadian Journal of Forest Research*, 46: 323-339.

Glenn N.F., Neuenschwander A., Vierling L.A., Spaete L., Li A.H., Shinneman D.J., Pilliod D.S., Arkle R.S. & McIlroy S.K. (2016). Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. *Remote Sensing of Environment*, 185: 233-242.

Goetz S. & Dubayah R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Management*, 2: 231-244.

Gwenzi D. & Lefsky, M.A. (2014). Prospects of photon counting lidar for savanna ecosystem structural studies. In *ISPRS Technical Commission I Symposium*, 17-20th November, Denver, Colorado, USA. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL-1, pp. 141-147.

Gwenzi D., Lefsky M.A., Suchdeo V.P. & Harding D.J. (2016). Prospects of the ICESat-2 laser altimetry mission for savanna ecosystem structural studies based on airborne simulation data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 118: 68-82.

Harding D.J. & Carabajal C.C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, 32: 1-4.

Homer C., Dewitz J., Yang L.M., Jin S., Danielson P., Xian G., Coulston J., Herold N., Wickham J. & Megown K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States - Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, 81: 345-354.

Houghton R.A. (2007). Balancing the global carbon budget. *Annual Review of Earth and Planetary Sciences*, 35(1): 313-347.

Houghton R.A., Butman D., Bunn A.G., Krankina O.N., Schlesinger P. & Stone T.A. (2007). Mapping Russian forest biomass with data from satellites and forest inventories. *Environmental Research Letters*, 2: 7pp.

Hu T., Su Y., Xue B., Liu J., Zhao X., Fang J. & Guo Q. (2016). Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. *Remote Sensing*, 8(7): 565.

Hudak A.T., Lefsky M.A., Cohen W.B. & Berterretche M. (2002). Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, 82: 397-416.

Jenkins J.C., Chojnacky D.C., Heath L.S. & Birdsey R.A. (2003). National-scale biomass estimators for United States tree species. *Forest Science*, 49: 12-35.

Le Toan T., Quegan S., Davidson M.W.J., Balzter H., Paillou P., Papathanassiou K., Plummer S., Rocca F., Saatchi S., Shugart H. & Ulander L. (2011).The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, 115: 2850-2860.

Lefsky M.A., Harding D.J., Keller M., Cohen W.B., Carabajal C.C., Del Bom Espirito-Santo F., Hunter M.O. & de Oliveira R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, 32: 1-4.

Lefsky M.A., Keller M., Pang Y., de Camargo P.B. & Hunter M.O.(2007). Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing*, 1: 1-18.

Leigh H.W., Magruder L.A., Carabajal C.C., Saba J.L. & McGarry J.F. (2015). Development of onboard digital elevation and relief databases for ICESat-2. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4): 2011-2020.

Liaw A. & Wiener M. (2002). Classification and Regression by randomForest. *R News*, Vol. 2/3, pp 18-22.

Markus, T., Neumann T., Martino A., Abdalati W., Brunt K., Csatho B., Farrell S., Fricker H., Gardner A., Harding D., Jasinski M., Kwok R., Magruder L., Lubin D., Luthcke S., Morison J., Nelson R., Neuenschwander A., Palm S., Popescu S., Shum C.K., Schutz B.E., Smith B., Yang Y. & Zwally J. (2017). The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sensing of Environment*, 190: 260-273.

Marselis S., Armston J. & Dubayah R. (2016). Summary of the second GEDI science team meeting. *The Earth Observer*, November - December 2016, 28 (6), pp 31-36.

Martino A. (2010). ATLAS Performance Spreadsheet. Available online: http://icesat.gsfc.nasa.gov/icesat2/data/sigma/sigma_data.php (accessed: 14.02.17).

Montesano P.M., Rosette J., Sun G., North P., Nelson R.F., Dubayah R.O., Ranson K.J. & Kharuk V. (2015). The uncertainty of biomass estimates from modeled ICESat-2 returns across a boreal forest gradient. *Remote Sensing of Environment*, 158: 95-109.

MRLC. (2018). Multi-resolution land characteristics consortium (MRLC): national land cover database. Available online: http://www.mrlc.gov/ (accessed: 14.02.19).

Narine, L.L., Popescu S., Neuenschwander A., Zhou T., Srinivasan S. & Harbeck K. (2019). Estimating Aboveground Biomass and Forest Canopy Cover with Simulated ICESat-2 Data. *Remote Sensing of Environment*, 224: 1-11.

National Aeronautics and Space Administration. (2019). GEDI - Global Ecosystem Dynamics Investigation Lidar. Available online: http://science.nasa.gov/missions/gedi/ (accessed: 11.10.18).

National Aeronautics and Space Administration. (2017). ICESat & ICESat-2. Available online: https://science.nasa.gov/ (accessed: 10.05.18).

Nelson R., Margolis H., Montesano P., Sun G., Cook B., Corp L., Andersen H.-E., deJong B., Pellat F.P., Fickel T., Kauffman J. & 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.

Neuenschwander A.L. & Magruder L.A. (2016). The potential impact of vertical sampling uncertainty on ICESat-2/ATLAS terrain and canopy height retrievals for multiple ecosystems. *Remote Sensing*, 8(12): 1-16.

Neuenschwander A., Popescu S., Nelson R., Harding D., Pitts K., Pederson D. & Sheridan R. (2017). Ice, Cloud, and Land Elevation Satellite (ICESat-2) Algorithm Theoretical Basis Document (ATBD) for Land - Vegetation Along-track products (ATL08), 78 p.

Popescu S.C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy*, 31: 646-655.

Popescu S.C. & Wynne R.H. (2004). Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering and Remote Sensing*, 70: 589-604.

Popescu S.C., Wynne R.H. & Nelson R.F. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29: 564-577.

Popescu S.C., Zhou T., Nelson R., Neuenschwander A., Sheridan R., Narine L. & Walsh K.M. (2018). Photon counting LiDAR: An adaptive ground and canopy height retrieval algorithm for ICESat-2 data. *Remote Sensing of Environment*, 208: 154-170.

Schutz B.E., Zwally H.J., Shuman C.A., Hancock D. & DiMarzio J.P. (2005). Overview of the ICESat mission. *Geophysical Research Letters*, 32(4).

Simard M., Pinto N., Fisher J.B. & Baccini A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research-Biogeosciences*, 116: 1- 12.

Texas A&M Forest Service. (2019). East Texas Forestlands, 2012. Available online: http://texasforestservice.tamu.edu/fia/publications/ (accessed: 11.10.18).

USDA Forest Service. (2018). Sam Houston National Forest. Available online: https://www.fs.usda.gov/detail/texas/about-forest/districts/?cid=fswdev3_008443 (accessed: 06.09.18).

Zald H.S.J., Wulder M.A., White J.C., Hilker T., Hermosilla T., Hobart G.W. & Coops N.C. (2016). Integrating Landsat pixel composites and change metrics with lidar plots to predictively map forest structure and aboveground biomass in Saskatchewan, Canada. *Remote Sensing of Environment*, 176: 188-201.

Zwally H.J., Schutz B., Abdalati W., Abshire J., Bentley C., Brenner A., Bufton J., Dezio J., Hancock D., Harding D., Herring T., Minster B., Quinn K., Palm S., Spinhirne J. & Thomas R. (2002). ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34(3-4): 405-445.

4. SYNERGY OF ICESAT-2 AND LANDSAT OPTICAL IMAGERY DATA FOR MAPPING FOREST ABOVEGROUND BIOMASS WITH DEEP LEARNING

4.1. Introduction

As forests continue to be altered and lost as a result of land use changes, among other causes, it has become increasingly vital to monitor their structure and extent to better understand the effects including those on the global carbon cycle and climate (Hall et al. 2011). Up-to-date and accurate maps of vegetation structure and AGB support sustainable management of forest resources (Zald et al. 2016), can be used to estimate other terrestrial carbon components (e.g. belowground biomass) (Goetz and Dubayah 2011), reduce uncertainties with carbon exchanges and the carbon budget (Goetz and Dubayah 2011), and facilitate an improved understanding of the carbon cycle (Hall et al. 2011). Light detection and ranging (lidar) remote sensing technology and specifically airborne and spaceborne lidar, have demonstrated the capability of estimating and mapping AGB (e.g. Hu et al. 2016; Chi et al. 2015; Nelson et al. 2017). Lidar systems measure the travel time for an emitted pulse of laser energy to reach the surface and then reflect back to the sensor, which facilitates a distance measurement and subsequently, unique XYZ location of or near to the surface (Glenn et al. 2016; Popescu 2007). There are currently two earth-orbiting lidars which were launched in 2018; the Advanced Topographic Laser Altimeter System (ATLAS) onboard the Ice, Cloud and land Elevation Satellite (ICESat)-2 (NASA 2017) and the Global Ecosystem Dynamics Investigation (GEDI) lidar attached to the International Space Station (ISS) (Stysley et

al. 2016). The data collected by these instruments may be used to estimate or derive forest vegetation parameters, including canopy heights and AGB, and could therefore play a crucial role in assessing and monitoring forest resources up to global scales (Nelson et al. 2017).

ICESat-2 will operate between 88° north and south latitudes during its three-year mission and will provide a nine times increase in spatial coverage than its predecessor (Markus et al. 2017) covering more of the earth's surface than GEDI which operates in mid latitudes, between 52° north and 52° south latitudes (Marselis et al. 2016; Sun 2012). While ATLAS onboard ICESat-2 was primarily designed to determine changes in ice sheet elevation and mass, it will provide information about vegetation that may be used to estimate AGB. ATLAS is a photon counting system, operating in the visible wavelengths, at 532 nm (Glenn et al. 2016). It generates 3 pairs of tracks, with each pair approximately 3.3 km apart and each track within a pair separated by 90 m (Markus et al. 2017). Lidar footprints are produced every 70 cm in the along-track direction and measure approximately 14 m in diameter (Neuenschwander and Magruder 2016). Given the unprecedented coverage and spatial detail from ICESat-2, translating ICESat-2 measurements to AGB estimates would allow for large-scale AGB and forest carbon assessments.

Variables derived from lidar data, particularly height metrics (Shao et al. 2017) and horizontal canopy structure metrics such as canopy cover are related to biomass reference data to estimate AGB (Lu et al. 2016). As the first and only satellite lidar operating from 2003 to 2009, the Geoscience Laser Altimeter System (GLAS) sensor

aboard ICESat acquired data that was utilized to estimate AGB (Popescu et al. 2011) and facilitated the mapping of forest resources at global scales (Hu et al. 2016; Lefsky 2010; Simard et al. 2011). While satellite lidar, including ICESat-2 and GEDI, will not provide spatially comprehensive measurements, the availability of other remotely sensed data, such as passive optical sensor data could be integrated to achieve a full coverage AGB product. For instance, in a study by Hu et al. (2016), a global wall-to-wall AGB product at 1 km resolution was produced using a combination of GLAS data and Moderate Resolution Imaging Spectroradiometer (MODIS) derived Normalized Difference Vegetation Index (NDVI) and land cover, and climatic and topographic variables. The random forests (RF) algorithm was used to extrapolate GLAS parameters and develop regression models with the spatially continuous variables. Chi et al. (2015) also used RF regression models to generate a nationwide wall-to-wall AGB map in China by extrapolating GLAS footprint-estimated AGB and MODIS data.

RF (Breiman 2001) is a machine learning technique that has been widely used for producing spatially explicit AGB estimates with multisource data (e.g. Baccini et al. 2004; Chi et al. 2015; Hu et al. 2016). The application of nonparametric machine learning regression algorithms, such as RF, Support Vector Regression (SVR) and k-nearest neighbor (k-NN) have become more predominant and demonstrate the ability to outperform popular parametric approaches used with remotely sensed data, like multiple linear regression (MLR) (Garcia-Gutierrez 2015; Tian et al. 2014; Wang et al. 2018). More recently, Deep Learning (DL) has been highlighted as a feasible approach for handling complex data (Hatcher and Yu 2018) with many examples in the remote

sensing literature focusing on classification and object detection tasks (Lv et al. 2019; Shafaey et al. 2019; Xiang et al. 2017). Few studies utilize DL models for forest parameter estimation and mapping, although promising results with lidar-derived variables have been reported (Shao et al. 2017). For instance, Garcia-Gutierrez (2016) indicated that autoencoders increased the accuracy of MLR predictions by 15-30%. Shao et al. (2017) found that a DL model, specifically stacked sparse autoencoder (SSAE) model, outperformed multiple stepwise linear regression, k-NN, support vector machine, back propagation neural network and random forest models, for estimating airborne lidar-derived AGB with variables from Landsat 8 Operational Land Imager (OLI) and synthetic-aperture radar (SAR) backscattering coefficients.

Deep Learning is a subset of machine learning that stems from cognitive and information theories which aim to mimic the learning process of neurons in the human brain (Hatcher and Yu 2018, p. 24412). DL is the application of multi-neuron, multi-layer neural networks to learn data representations (Hatcher and Yu 2018, p. 24413). In a neural network layer consisting of multiple nodes, each neuron is initiated with a different weight and neurons simultaneously learn the input data (Hatcher and Yu 2018). Weights are updated based on a loss function and in the case of multiple layers, each neuron learns from all output from preceding layers. The use of deep learning architectures to any type of data, including numerical, visual and audio has propelled DL to a dominant position for developing predictive systems (Hatcher and Yu 2018). Regression is one of the two primary supervised learning tasks carried out with DL, with classification being the second (Hatcher and Yu 2018) and more popular task undertaken

(Trier et al. 2018). Deep neural networks (DNNs) in particular, are capable of extracting combinations of the input that are not easily described by humans (Trier et al. 2018, p. 338). Given the capabilities of DL and increasing amounts data from remote sensing systems, including current and upcoming space lidar missions, DL models could be investigated for modeling and mapping AGB and other forest structural parameters.

One of ICESat-2's data products is ATL08 or Land-Vegetation along-track products which will consist of terrain heights, canopy heights and canopy cover estimates for non-polar regions covered by the satellite (Neuenschwander et al. 2017; Neuenschwander and Pitts 2019). Estimates will be provided at a step-size of 100 m in the along-track direction and the products will be used as input to generate gridded heights and canopy cover products or ATL18 after the three-year mission (Markus et al. 2016; Neuenschwander et al. 2017). Several studies demonstrate the use of satellite lidar (GLAS) data for mapping forests (e.g. Baghdadi et al. 2014; Chi et al. 2015; Harding & Carabajal 2005; Hu et al. 2016; Lefsky et al. 2007, Simard et al. 2011, Nelson et al. 2017) and fewer studies have compared techniques for modeling AGB with space lidar (e.g. Liu et al. 2017). However, literature focused on ICESat-2 data for vegetation studies (Glenn et al. 2016; Gwenzi et al. 2016; Montesano et al. 2015; Narine et al. 2019; Neuenschwander and Magruder 2016; Popescu et al. 2018) is limited, including studies aimed at exploring approaches for AGB mapping with the data or its products (Narine et al. in review).

Over the next two years, ICESat-2 will sample the earth's surface to provide a grid of measurements (Markus et al. 2016, p. 269) that will enable estimation of a forest

attributes. Rather than waiting for data, this study investigated an approach for mapping AGB using simulated ICESat-2 data over two years of preplanned track locations and Landsat data, in preparation for utilizing the actual data for vegetation studies as soon as it becomes available. Given the growing interest in DL (Hatcher and Yu 2018) and potential of DNNs (Trier et al. 2018), a methodology for mapping aboveground biomass (AGB) using regression-based feedforward neural network models was explored. Even though the topology of a neural network substantially affects the results (Stathakis 2009), the optimum numbers of layers and nodes are not automatically selected. Thus, a primary objective of this study was to investigate parameter settings that would yield optimum predictive performance, specifically the number of hidden neurons and hidden layers as well learning rates. Acknowledging expected differences in noise levels associated with daytime and nighttime operation of ATLAS and associated impacts on canopy height estimation (Popescu et al. 2018), an examination of AGB mapping under different data scenarios was carried out. Thus, parameter tuning was undertaken separately for each scenario and the best models were applied to generate the AGB maps at 30 m spatial resolution. AGB maps were produced for the following data scenarios: (i) daytime scenario, (ii) nighttime scenario, and (iii) without the impact of noise (no noise scenario).

4.2. Materials and Methods

4.2.1. Study Area

The study was carried out in in Sam Houston National Forest located in south-east Texas, USA (Latitude 30° 42′ N, Longitude 95° 23′ W). Elevations range from 62 m to 105 m, with an average of 85 m (Popescu 2007). Approximately 58% of the region or 80% of its forested area (NLCD classes: deciduous forest, evergreen forest, mixed forest, and woody wetland) is classified as evergreen forests (Homer et al. 2015; MRLC 2017) and the site is predominated by pine forests, which include Loblolly pine (*Pinus taeda*) plantations and old growth Loblolly pine stands (Popescu 2007). Planned ICESat-2 track locations for the first two years of the mission over the study area (Neuenschwander et al. 2017) are depicted on Figure 11.

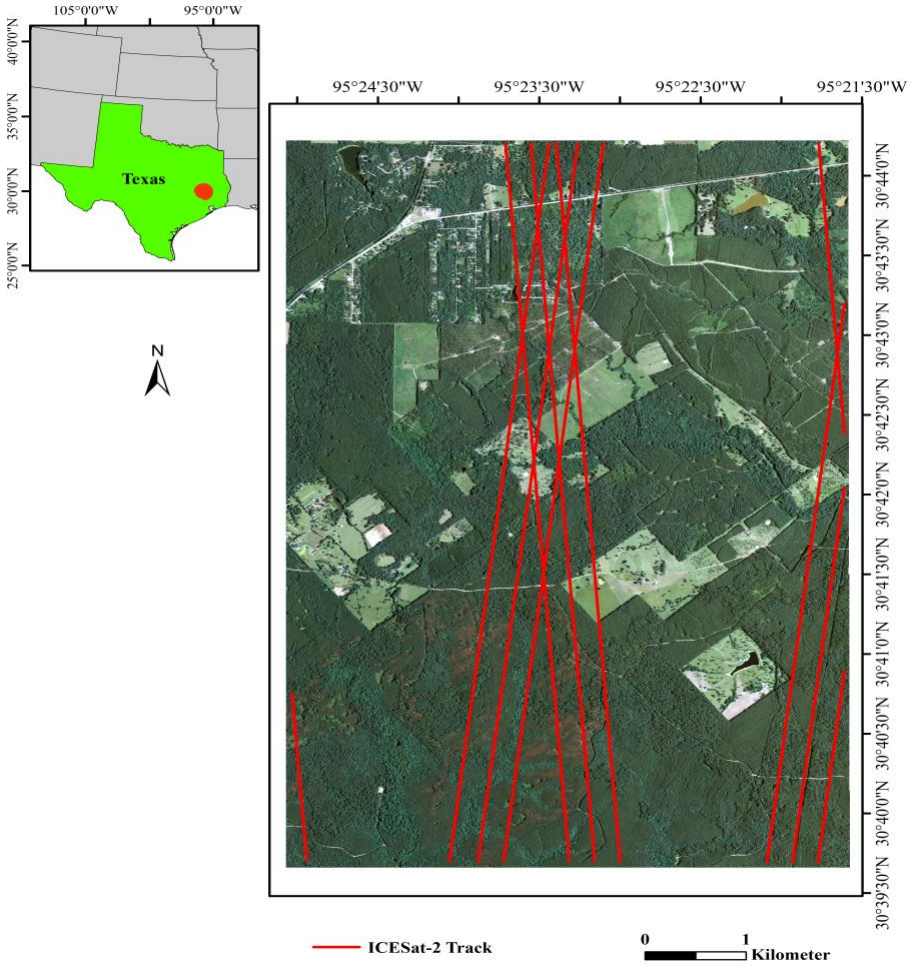


Figure 11. ICESat-2 tract locations overlaid on 2010 National Agriculture Imagery Program (NAIP) aerial imagery for the study area within SHNF, Texas (inset map, upper left corner)

4.2.2. Simulated ICESat-2-estimated AGB

Discrete return lidar data acquired in November 2010 over SHNF with a point density of 4 points per m^2, were used to simulate ICESat-2's photon-counting lidar (PCL) data. Planned track locations, ICESat-2 footprint generator and simulator were made available by the mission's Science Definition Team (SDT). Footprints spaced 70

cm apart were generated along each planned ICESat-2 track and photons were simulated using the airborne lidar data. A detailed description of the PCL simulation is provided in Neuenschwander and Magruder (2016). To summarize, discrete return lidar points falling each generated ICESat-2 footprint along tracks were used to generate a pseudo-waveform (Blair and Hofton 1999), a vector with heights or elevation was then constructed for each footprint, the number of photon based on design cases were randomly determined and random sampling of height vector weighted by the pseudo-waveform was carried out (Neuenschwander and Magruder 2016). The average number of signal photons per shot was modeled at 1.9, based on the ATLAS performance model for temperate forests (Martino 2010) and the algorithm returned up to 3 photons per 14-m diameter footprint.

Anticipated noise photons, representative of solar background noise and effects of atmospheric scattering (Moussavi et al. 2014), were added to the simulated dataset. Two datasets were generated to represent the following: (1) daytime scenario with expected noise levels for daytime operation of ICESat-2, and (2) nighttime scenario with noise levels based on night operation of ICESat-2. Popescu et al. (2018) devised and applied novel noise filtering and photon classification algorithms to simulated ICESat-2 data and reported average RMSE values of 2.70 m and 3.59 m for estimating canopy heights. PCL data processing algorithms developed by Popescu et al. (2018) were applied to the data to remove noise photons and classify photons into top of canopy points and ground surface elevation points; subtracting the latter from the former yielded tree canopy height values within footprints. Another PCL simulation using above-

ground-level heights was carried out and no noise photons were added to the data. In total, data for three scenarios were analyzed; 1) daytime scenario, 2) nighttime scenario, and 3) no noise scenario.

Data representative of ICESat-2's ATL08 product was used for modeling relationships with spatially coincident airborne-lidar derived AGB. To maintain consistency with the ATL08 format, 100 m segments were extracted from the simulated datasets and height percentiles, canopy cover and canopy density metrics were calculated (Narine et al. 2019). The utilization of the simulated ICESat-2 PCL vegetation product for estimating AGB and airborne lidar-derived canopy cover are presented in Narine et al. (2019). To summarize, linear regression models relating simulated PCL metrics for a subset (n = 85) of the segments to airborne lidar-derived AGB were developed and their performance assessed with a separate test set (n = 36). AGB models yielded RMSEs of 25.35 Mg/ha, 19.23 Mg/ha, and 19.16 Mg/ha and R^2 values of 0.63, 0.79, 0.79 for the daytime scenario, nighttime scenario and no noise scenario respectively (Narine et al. 2019).

Using AGB models developed with the simulated PCL vegetation product over SHNF, AGB density (Mg/ha) was estimated for each segment over the study site and applied to 30 m pixels to match Landsat TM pixels (Narine et al. in review). AGB was assigned based on the portion of a 100 m segment across a pixel and the value was extrapolated to represent the pixel size using the estimated area of a segment across a pixel. Segments less than 7 m in a pixel were excluded from analysis and average pixel AGB was calculated in instances where there were two segments (parts) of equal lengths

traversing a pixel. The steps for assigning AGB density to pixels were repeated for each data scenario using the corresponding prediction equation developed from linear regression analysis. Then, 70% of the data (pixels) was randomly assigned to the training dataset and remaining 30% was allotted to the test set and used for model evaluation. The training and test set for each scenario consisted of 1448 and 620 30-m pixels respectively.

4.2.3. Mapped Predictors

A Landsat 5 Thematic Mapper (TM) cloud-free scene encompassing the study site, from path/row 26/39 and acquired in November, 2010 was downloaded from U.S. Geological Survey (USGS) Earth Explorer in GeoTIFF format and re-projected to UTM, WGS84, Zone 15N. The image was processed in ENVI to top-of-atmosphere (TOA) reflectance and vegetation indices were computed. Land cover and canopy cover maps from the 2011 National Land Cover Database (NLCD) (Homer et al. 2015) were also downloaded and georeferenced in the same projection as the Landsat TM image. The predictor variables used for AGB modeling were:

- Spectral Metrics from Landsat 5 TM-
    - Normalized Difference Vegetation Index (NDVI): (NIR - Red) / (NIR + Red)
    - Enhanced Vegetation Index (EVI): 2.5 * ((NIR - Red) / (NIR + 6 * Red - 7.5 * Blue + 1))

    - Soil Adjusted Vegetation Index (SAVI): ((NIR - Red) / (NIR + Red + 0.5)) * (1.5)
    - Modified Soil Adjusted Vegetation Index (MSAVI): (2 * NIR + 1 - sqrt ((2 * NIR + 1)² - 8 * (NIR - Red))) / 2
- NLCD 2011 land cover map
- NLCD 2011 US Forest Service tree canopy cover

4.2.4. Deep Neural Networks (DNNs)

DL via neural networks is comprised of stacked layers that facilitate learning successive layers of representations of the input data (Chollet 2018, p. 8). Essentially, a layer transforms the data, as specified by weights, which are referred to as the parameters of the layer (Chollet 2018). The process of learning entails finding the optimum values of the weights in each layer of a neural network, which are adjusted by an optimizer, based on the loss score. A loss function measures the difference between the observed values and predictions from the neural network and adjustments with the optimizer serves to lower the loss score (Chollet 2018). Network weights are randomly initialized resulting in a high loss score but are adjusted as each example (batch) is processed and iterated several times to produce weights that lower the loss function (Chollet 2018, p. 11). Essentially, the gradient of the loss, given the combination of the weights, is computed and parameters are subsequently moved to an extent defined by the learning rate, to reduce the loss for the batch (Chollet 2018). In terms of optimization, that there are several methods of reducing loss through gradient descent to improve

network accuracy. An activation function can also be to each layer in the model, enabling the layer to learn non-linear transformations of the data. For all neural network architectures developed in this study, the RMSProp optimizer and rectified linear unit (ReLU) as the activation function, which are regarded as good options for different tasks (Chollet 2018), were used. The activation function g is applied element-wise defined by the function $g(z) = max\{0, z\}$ (Goodfellow et al. 2016; Trier et al. 2018). Assuming h is a vector of hidden units in a layer, then $h = g(W^{\mathrm{T}}x + c)$, where W are the weights of the transformation and c are the biases (Goodfellow et al. 2016, p. 168), such that each element in the hidden layer has a different weight and bias parameter (Bengio 2009). For all models, the loss function was defined with the mean squared error (MSE), learning rate was initially set to 0.001and epochs and batch size were 100 and 32 respectively. All neural models consisted of one input layer with 6 neurons (one for each predictor) and one output node for the predicted AGB.

Determining the architecture of a neural network affects its ability to learn the data (Bengio 2009) and ultimately impacts its predictive capacity. Specifically, determining the number of hidden layers and nodes are critical (Doukim et al. 2010) since network topology can have sizeable impacts on the results (Stathakis 2009). A trial and error approach is common for determining network topology while studies have presented techniques for identifying an optimal network structure for specific applications. For example, Doukim et al. (2010) highlighted a course to fine search technique for determining the number of hidden neurons in a multi-layer perceptron (MLP) (feedforward) neural network for skin detection from images. Doukim et al.

(2010) used a binary search, with hidden neurons set to 1, 2, 4, 8, 16, 32, 64, and 128 and then a sequential search around a specific range indicated by the binary search to fine-tune the process of finding the number of hidden neurons that resulted in the smallest MSE. In this study, a network structure consisting of 112 hidden neurons was used to construct a neural network for skin detection. In another study, Guang-Bin (2003) demonstrated that a two layer feedforward neural network is sufficient for learning with minimal error and suggests that the number of nodes in the first layer is given by:

$$\sqrt{(m+2)N} + 2\sqrt{N/(m+2)}$$

And the number of hidden nodes in the 2nd layer is defined by

$$m\sqrt{N/(m+2)}$$

Where N is the number of samples and m is the number of output neurons.

For comparison purposes, the equations for determining the number of hidden nodes in the first and second hidden layers proposed by Guang-Bin (2003) were applied to the training dataset, which equates to 98 neurons in the first layer and 20 hidden nodes in the 2nd layer. Since determining optimum network structure was a primary focus in order to understand the applicability of deep architectures with simulated ICESat-2 data, the number of hidden neurons and layers were varied. Exploratory analysis involving gradual increases in the number of hidden neurons in the first hidden layer (e.g. binary search in Doukim et al. 2010) did not result in substantial changes in model performance. As a result, the number of neurons were varied in intervals of 20 up to 200

neurons and then in intervals of 100 with a maximum of 1000 hidden neurons in the first hidden layer. Results from the application of each trained model to the separate test set were compared RMSE and R^2 values. Guang-Bin (2003) proposed a neural network architecture consisting of a large first hidden layer and a second hidden layer that is considerably narrower. To adopt this feature to the network architecture for deeper models, successively narrow layers were applied. A second hidden layer was added to neural network that gave the highest R^2 and lowest RMSE value in each scenario and the number of neurons was varied in increments of 20 up to the number of hidden neurons in the previous layer. Layers were added until there were no further improvements in model performance.

With the final selected network structure for each scenario, the learning rate used by the RMSprop optimizer was changed from 0.001 to 0.1, 0.01, and 0.0001 (Xie et al. 2017) and model performance with the test set were compared for each hyperparameter setting. Regarded as the most important hyperparameter of DL algorithms (Goodfellow et al. 2016, p. 424), the learning rate affects the model's capacity, as indicated by model error. Too large to too small learning rates could increase training error as well as training times in the case where rates are too small and suboptimal. Models yielding the lowest RMSE and highest R^2 for each scenario were used to generate AGB maps. Modeling was done with the Keras (Chollet 2018) with TensorFlow backend in R and data visualization, with ArcGIS 10.4 software.

4.3. Results

An assessment of neural networks predicting simulated PCL-estimated AGB for the daytime, nighttime and no noise data scenarios, indicated models with 300 hidden neurons, 600 hidden neurons, and 500 hidden neurons in the first hidden layer performed best (Table 6). These models explained 40%, 47% and 48% of the variance in simulated PCL-estimated AGB with RMSEs of 19.90 Mg/ha, 19.72 Mg/ha and 20.29 Mg/ha for the daytime, nighttime and no noise scenarios respectively. Incremental increases in the number of neurons in the first hidden layer yielded R^2 values ranging from 0.38 to 0.40, 0.45 to 0.47, and 0.46 to 0.48 with the daytime, nighttime and no noise scenarios. In comparison, using the formula proposed by Guang-Bin (2003), a DNN with 98 neurons in the first layer and 20 hidden nodes in the 2nd layer yielded a R^2 and RMSE of 0.40 and 19.92 Mg/ha, 0.45 and 19.95 Mg/ha and 0.47 and 20.27 Mg/ha for the daytime, nighttime and no noise scenarios respectively.

Table 6. Model performance with different number of neurons in the first hidden layer

Number of neurons in 1st hidden layer	Daytime Scenario		Nighttime Scenario		No Noise Scenario	
	R^2	RMSE (Mg/ha)	R^2	RMSE (Mg/ha)	R^2	RMSE (Mg/ha)
20	0.40	19.95	0.45	19.97	0.46	20.60
40	0.40	20.01	0.45	19.98	0.46	20.55
60	0.40	19.98	0.46	19.92	0.47	20.49
80	0.40	19.94	0.46	19.91	0.47	20.49
100	0.40	19.94	0.45	19.99	0.47	20.44
120	0.40	19.95	0.45	19.94	0.47	20.46
140	0.40	20.00	0.46	19.89	0.47	20.44
160	0.40	20.02	0.45	19.97	0.47	20.38
180	0.39	20.08	0.46	19.84	0.47	20.43

Table 6 Continued

Number of neurons in 1st hidden layer	Daytime Scenario		Nighttime Scenario		No Noise Scenario	
	R^2	RMSE (Mg/ha)	R^2	RMSE (Mg/ha)	R^2	RMSE (Mg/ha)
200	0.39	20.06	0.45	19.98	0.47	20.46
300	0.40	19.90	0.46	19.88	0.47	20.36
400	0.39	20.09	0.46	19.85	0.47	20.32
500	0.39	20.05	0.45	20.08	0.48	20.29
600	0.40	19.91	0.47	19.72	0.47	20.34
700	0.39	20.13	0.46	19.87	0.47	20.37
800	0.40	19.93	0.46	19.75	0.47	20.35
900	0.40	19.98	0.46	19.84	0.47	20.36
1000	0.38	20.27	0.45	20.00	0.48	20.30

To better understand the applicability of DNNs using simulated PCL-estimated AGB and predictors consisting of Landsat spectral metrics, landcover and canopy cover, model performance from varying the number of hidden neurons in each additional hidden layer were assessed and compared. With the daytime scenario, a second hidden layer was added and neurons were increased in increments of 20 to a maximum of 300 hidden neurons. The DNN consisting of 160 neurons in the second hidden layer performed best, in terms of R^2 and RMSE and addition of a third hidden layer did not improve model performance (Figure 12). With the nighttime scenario, a second hidden layer with hidden neurons varied from 20 to 500 was investigated, leading to DNN structure consisting of 2 hidden layers with 600 in the first hidden layer and 400 in the second. The lowest reported RMSE corresponding to the DNN with a third hidden layer is shown in Figure 12. However, DNN with two hidden layers achieved the best results. The best performing model with the no noise scenario consisted of five densely connected layers; one input layer, 3 hidden layers with 500, 300 and 60 neurons in the

first, second and third hidden layer respectively, and 1 output layer. DNNs with four hidden layers did not result in better models (Figure 12).

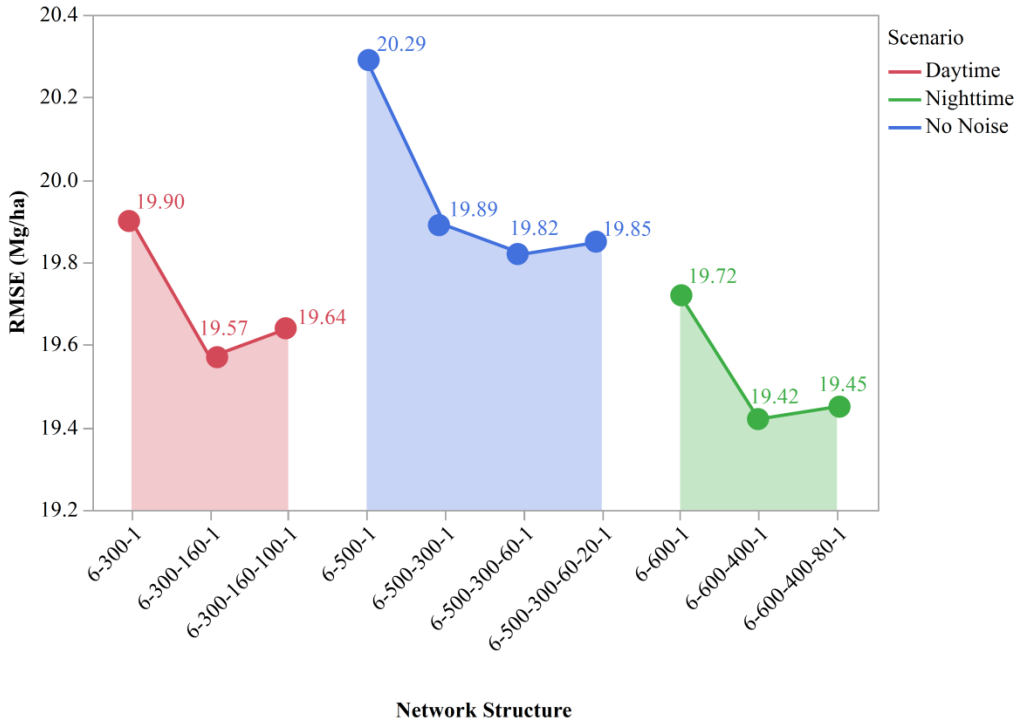


Figure 12. Neural network structures for predicting AGB with the daytime, nighttime and no noise scenarios

Findings from tuning the learning rate for each DNN (Table 7) resulted in final models with the parameter set to 0.0001 for the daytime and nighttime scenarios and 0.001 for the no noise scenarios. Substantial decreases in R^2 were noted when the learning rate was increased to 0.01 and 0.1 indicating overall suitability of the 0.001 and 0.0001 level with the data. For example, with the DNN model for the daytime scenario, the learning rate set to 0.0001 yielded the highest R^2 and lowest RMSE of 0.42 and 19.55 Mg/ha, while a learning rate of 0.1 resulted in R^2 of 0.25 and RMSE of 22.32

Mg/ha. With the nighttime scenario, a change in the learning rate from 0.1 to 0.0001 increased the R^2 value from 0.34 to 0.49 leading to 0.0001 as the final learning rate used in the model for producing the AGB map. The model for the no noise scenario yielded R^2 and RMSE of 0.50 and 19.82 Mg/ha when the learning rate was set to 0.001 or 0.0001, compared to 0.40 and 21.66 Mg/ha when the rate was 0.1.

Scatterplots with simulated PCL-estimated AGB vs DNN predicted AGB (Figure 13) show the line of best fit adjacent to the 1:1 line. However, DNN models tended to underestimate AGB, where this trend was most pronounced with the daytime scenario and least prominent with the no noise scenario. Average AGB predictions from maps (Figures 14-16) produced with the final trained DNN models for the daytime, nighttime and no noise scenarios were 43.16 Mg/ha, 45.85 Mg/ha and 46.52 Mg/ha respectively. Overall, maps correspond with vegetation patterns in the study area with the southern portion of the site predominated by forests, primarily mature pines, and lower AGB values in northern portions, which includes young pine stands. Higher AGB values are evident in the map generated with the no noise scenario while lower AGB ranges are more prevalent with the map for the daytime scenario. Maximum predicted AGB density for 30 m cells was 78.88 Mg/ha for the daytime scenario, 85.73 Mg/ha with the nighttime scenario and 101.35 Mg/ha with the no noise scenario. With the training data average AGB were 51.21 Mg/ha, 52.24 Mg/ha and 52.09 Mg/ha for the daytime, nighttime and no noise scenarios and maximum values were 139.61 Mg/ha, 119.75 Mg/ha and 150.64 Mg/ha. A comparison of DNN models with RF models (Narine et al. in press) produced with the same data (Table 8) highlighted comparable predictive

abilities. Except for the no noise scenario, R^2 values were identical for the two sets of models and RMSEs were in the same range.

Table 7. DNN model performance for the daytime and nighttime and no noise scenarios, for different learning rates

	Daytime Scenario Model Structure: 6-300-160-1		**Nighttime Scenario Model Structure: 6-600-400-1**		**No Noise Scenario Model Structure: 6-500-300-60-1**	
Learning Rate	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
0.1	0.25	22.32	0.34	22.01	0.40	21.66
0.01	0.39	20.17	0.45	20.02	0.41	21.48
0.001	0.42	19.57	0.48	19.42	0.50	19.82
0.0001	0.42	19.55	0.49	19.35	0.50	19.82

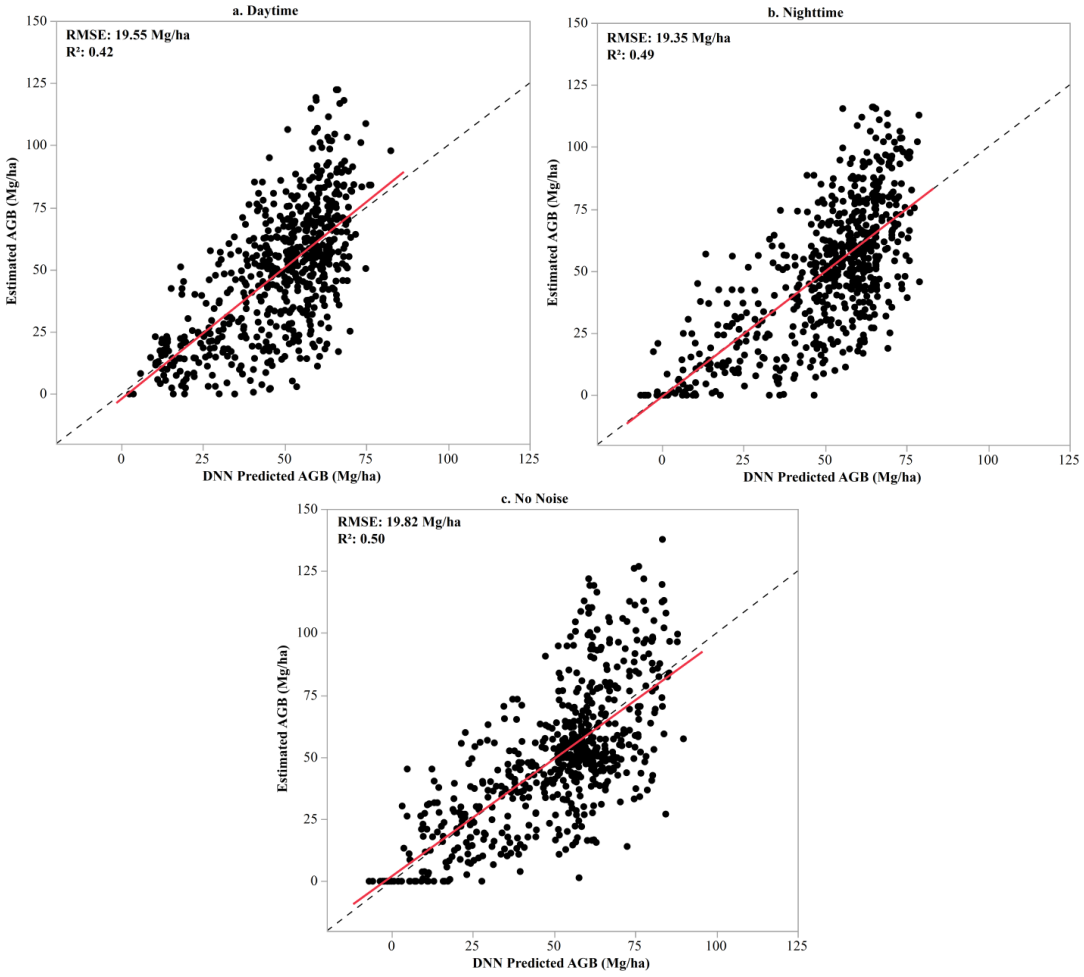


Figure 13. Simulated PCL AGB estimated from linear regression vs DNN predicted AGB with test data for the daytime scenario; (b) Simulated PCL AGB estimated from linear regression vs DNN predicted AGB with test data for the nighttime scenario (c) Simulated PCL AGB estimated from linear regression vs DNN predicted AGB with test data for the no noise scenario. The dashed line in each graph is the 1:1 line

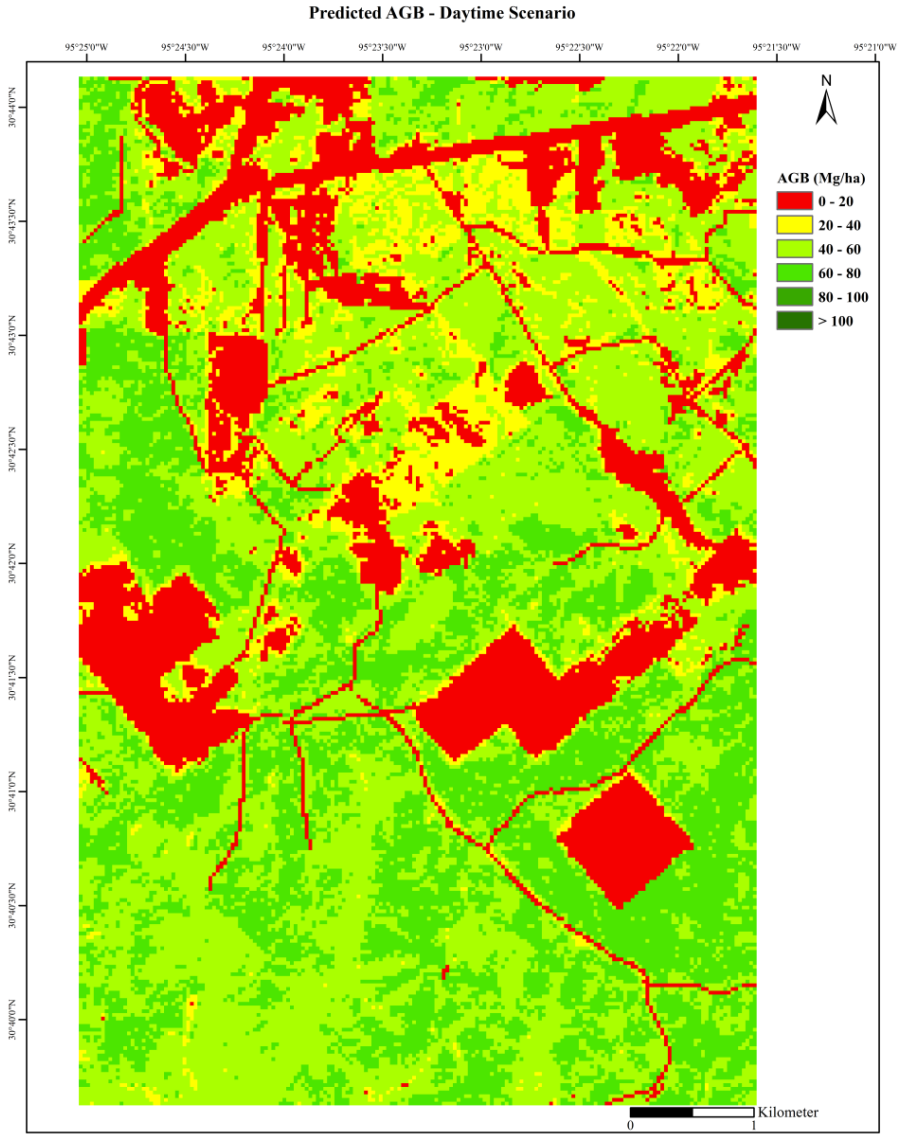


Figure 14. DNN predictions of AGB density at a 30 m spatial resolution for the daytime scenario

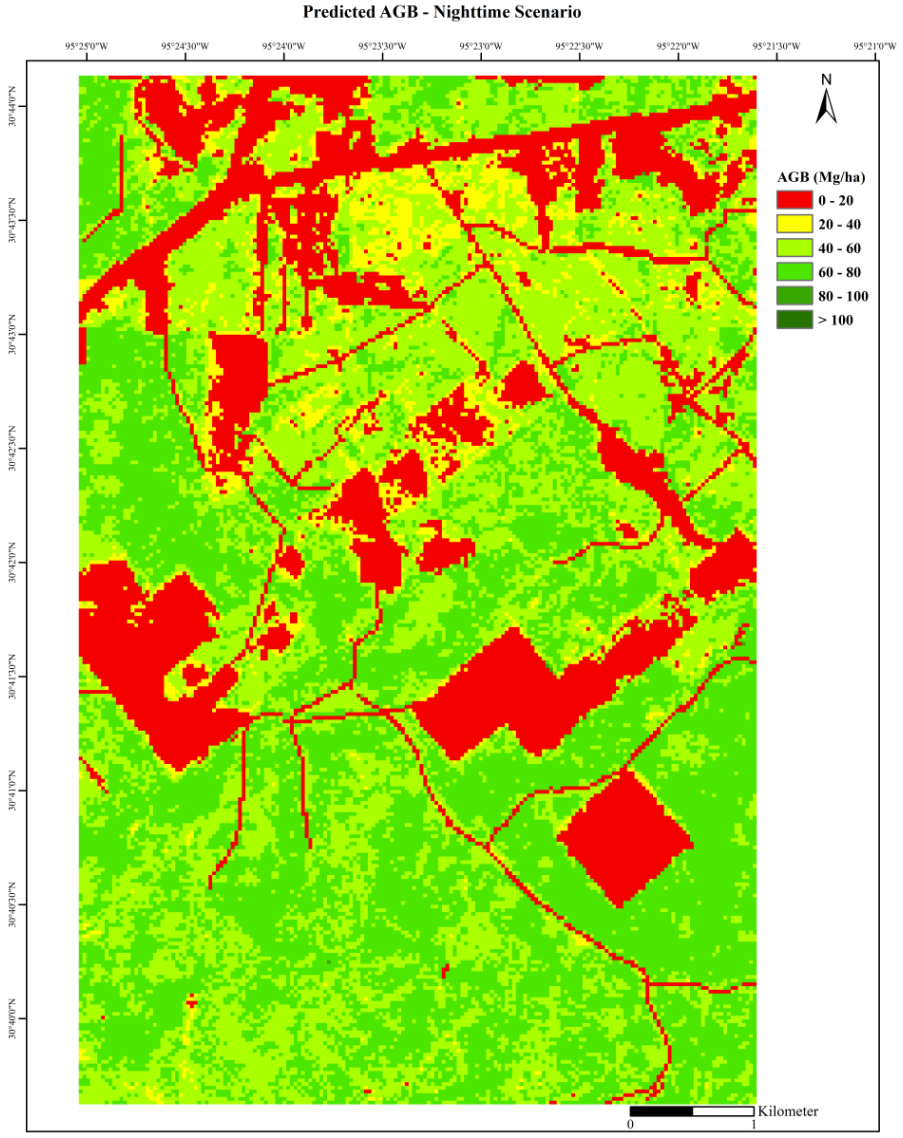


Figure 15. DNN predictions of AGB density at a 30 m spatial resolution for the nighttime scenario

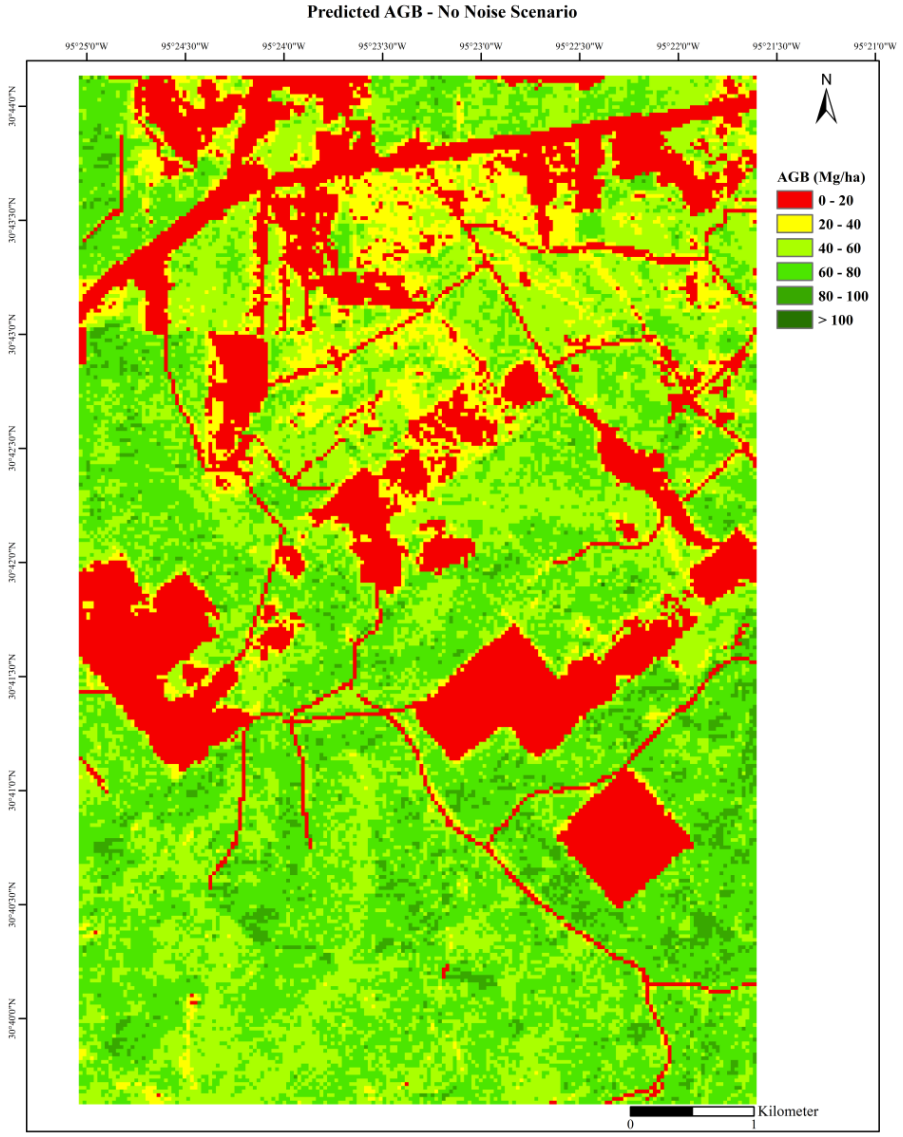


Figure 16. DNN predictions of AGB density at a 30 m spatial resolution for the no noise scenario

Table 8. Comparison of RF predicted AGB with AGB predicted from DNN models under three scenarios; no noise, daytime and nighttime scenarios

	RF Model		DNN Model	
Scenario	**R^2**	**RMSE**	**R^2**	**RMSE**
Daytime	0.42	19.69 Mg/ha	0.42	19.55 Mg/ha
Nighttime	0.49	19.30 Mg/ha	0.49	19.35 Mg/ha
No Noise	0.51	19.72 Mg/ha	0.50	19.82 Mg/ha

4.4. Discussion

With ICESat-2 already in launch, data will soon be available for meeting the mission's science objectives, one of which has a direct benefit to the ecosystem community (Markus et al. 2017). The mission's ATL08 product will provide canopy heights and canopy cover which will be used to produce global maps of these attributes at the end of the mission (Neuenschwander et al. 2017; Markus et al. 2017). The investigation of approaches for utilizing ICESat-2 data using simulated data could potentially offer a better understanding of how the data can be used for vegetation studies, provide insights about the expected accuracies of techniques examined, and potentially facilitate quick adoption using the actual data. The relationship between simulated PCL metrics from 100 m segments and reference airborne lidar-estimated AGB along ICESat-2 profiles over the study site in SHNF was modeled in a previous study (Narine et al. 2019) and the resulting prediction equations were applied to estimate AGB for daytime, nighttime and no noise scenarios. Using freely available data, including satellite data which offers global coverage, AGB estimates were up-scaled from 3% of 30 m pixels which comprise the area, to produce spatially explicit AGB coverage.

In this study, simulated PCL estimated AGB along 2 years of tracks over the study site were used in synergy with spectral metrics derived from Landsat imagery and NLCD products to develop regression-based models for application to the rest of the study site. Deep learning was implemented using feedforward neural networks consisting of multiple nonlinear hidden layers. A key task was to determine an appropriate number of hidden layers and hidden neurons or architecture of models predicting AGB with the data. The final model for the daytime scenario consisted of four densely connected layers; one input layer, two hidden layers and one output layer. With the nighttime scenario, the neural network consisted of two non-linear hidden layers and the neural network structure for the no noise scenario consisted of three hidden layers. All neural networks used fully connecting layers and at least 300 neurons in the first hidden layer of each model, resulting in architectures with hundreds of thousands of parameters or weights. Increasing the width of the neural networks led to some improvements in terms of RMSEs with negligible differences in training time and with added depth, model performance improved with consistently better results from the nighttime scenario than daytime scenario. One possible reason for R^2 values being less than expected could be dissimilarities in the range of biomass represented in the test data and training data, as mentioned in a biomass mapping study with neural networks by Foody et al. (2001). Thus, additional training data representative of the range of AGB, including data in the test set, may improve model accuracies. It is also important to note that additional layers of representation to the final selected network structures did not improve model performance and findings demonstrate that the DNN models for all data

scenarios yielded results that were comparable to RF models. With more training data, the use of a deeper architecture could be beneficial and likely improve results (Liu 2017).

A comparison of different learning rates for selected model structures revealed apparent differences in reported R^2 values and RMSEs. In one study, comparison of learning rates using a regression-based, deep auto-encoder model with two hidden layers for gene expression prediction yielded MSEs ranging from 0.289 to 0.292, with the best results attributed to the model with highest learning rate; 0.1 (Xie et al. 2017). Learning rates investigated were 0.1, 0.01, 0.001, 0.0001 and 0.00001 (Xie et al. 2017). Conversely, findings from this study emphasized the importance of the lowest learning rates investigated (0.001 and 0.0001) for predicting AGB in each data scenario and with highest RMSE and R^2 values corresponding to DNNs when the learning rates were set to 0.1.

The use of RF has been successfully demonstrated in the literature for mapping AGB with data from ICESat-2's predecessor (e.g. Chi et al. 2016), but with the ability to handle large, complex datasets, DL algorithms could be explored for generating a spatially explicit AGB product from an integration of multi-source data, including GEDI, ICESat-2, satellite optical imagery and radar data with other ancillary data. For instance, a combination of predicted AGB density from GEDI and ICESat-2 data could be combined with mapped variables derived from Landsat data, for generating a wall-to-wall AGB map. With current and upcoming space-based vegetation missions like NASA-ISRO Synthetic Aperture Radar (NISAR) (Rosen et al. 2016) and the European

Space Agency P-band radar BIOMASS mission (Carreiras et al. 2017), the volume of data that can potentially provide rich insights about the world's forests will grow considerably and DL models could play a larger role in predicting forest attributes. Given new data from satellite sensors and multiple, open-source DL frameworks (e.g. TensorFlow), DL architectures for predicting AGB can be explored more extensively. Future work could entail an examination and comparison of multiple DL techniques, like the SSAE model (Shao et al. 2017), autoregressive neural network (e.g. Nunez et al. 2018), and recurrent neural network (e.g. Lee and Lee 2018), for producing a wall-to-wall AGB product with ICESat-2 and other satellite remote sensing data. Finally, limited examples of DL models for estimating forest biophysical parameters in the literature suggest the need for further research to better understand or identify potential benefits from DL for forest mapping applications.

4.5. Conclusions

In this study, an approach for mapping AGB was developed by integrating simulated PCL-estimated AGB with Landsat imagery, NLCD canopy cover and land cover using DNNs. The effect of network structure and learning rate on model performance were evident with the latter having sizeable impacts on reported error metrics. While reported accuracies from DL models were comparable to RF models in this study, additional training data may likely improve results and deeper architectures could be investigated. Nevertheless, findings with simulated PCL-estimated data along ICESat-2 profiles, especially with the nighttime scenario ($R^2 = 0.49$), highlight the

potential for generating a wall-to-wall AGB product with ICESat-2 by adopting a synergistic approach with Landsat optical imagery, canopy cover and land cover. With growing volumes of data from current space lidars and upcoming satellite missions and learning capacity of DL, DL algorithms should be explored in future research. In doing so, multiple techniques for modeling AGB and other forest structural parameters could be implemented and compared.

4.6. References

Baccini, A., Friedl, M.A., Woodcock, C.E., & Warbington, R. (2004). Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters, 31*, 4.

Baghdadi, N., le Maire, G., Fayad, I., Bailly, J.S., Nouvellon, Y., Lemos, C., Hakamada, R., & Ieee (2014). Estimation of Forest Height and Aboveground Biomass from ICESat-2/GLAS Data in Eucaluptus Plantations in Brazil. *2014 Ieee International Geoscience and Remote Sensing Symposium (Igarss)*, 725-728.

Bengio, Y. (2009). Learning deep architectures for AI. Foundations and Trends in Machine Learning, 2(1): 1-55.

Blair, J.B., & Hofton, M.A. (1999). Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters, 26*, 2509-2512

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32.

Carreiras, J.M.B., Shaun, Q.G., Toan, T.L., Minh, D.H.T., Saatchi, S.S., Carvalhais, N., Reichstein, M., & Scipal, K. (2017). Coverage of high biomass forests by the ESA BIOMASS mission under defense restrictions. *Remote Sensing of Environment, 196*, 154-162.

Chi, H., Sun, G.Q., Huang, J.L., Guo, Z.F., Ni, W.J., & Fu, A.M. (2015). National Forest Aboveground Biomass Mapping from ICESat/GLAS Data and MODIS Imagery in China. *Remote Sensing, 7*, 5534-5564.

Chollet, F.o. (2018). *Deep learning with Python*. Shelter Island, NY: Manning Publications Co.

Doukim, C.A., Dargham, J.A., & Chekima, A. (2010). Finding the number of hidden neurons for an MLP neural network using coarse to fine search technique. In, *10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)* (pp. 606-609).

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*, 379-387.

Garcia-Gutierrez, J., Gonzalez-Ferreiro, E., Mateos-Garcia, D., & Riquelme-Santos, J.C. (2016). A Preliminary Study of the Suitability of Deep Learning to Improve LiDAR-Derived Biomass Estimation. *Hybrid Artificial Intelligent Systems, 9648*, 588-596.

Garcia-Gutierrez, J., Martinez-Alvarez, F., Troncoso, A., & Riquelme, J.C. (2015). A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing, 167*, 24-31.

Glenn, N.F., Neuenschwander, A., Vierling, L.A., Spaete, L., Li, A.H., Shinneman, D.J., Pilliod, D.S., Arkle, R.S., & McIlroy, S.K. (2016). Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. *Remote Sensing of Environment, 185*, 233-242.

Goetz, S., & Dubayah, R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Management, 2*, 231-244.

Chollet, F.o. (2018). *Deep learning with Python*. Shelter Island, NY: Manning Publications Co.

Goodfellow, I. Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org (accessed: 02.05.19).

Guang-Bin, H. (2003). Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Transactions on Neural Networks, 14*, 274-281.

Gwenzi, D., Lefsky, M.A., Suchdeo, V.P., & Harding, D.J. (2016). Prospects of the ICESat-2 laser altimetry mission for savanna ecosystem structural studies based on airborne simulation data. *Isprs Journal of Photogrammetry and Remote Sensing, 118*, 68-82.

Hall, F.G., Bergen, K., Blair, J.B., Dubayah, R., Houghton, R., Hurtt, G., Kellndorfer, J., Lefsky, M., Ranson, J., Saatchi, S., Shugart, H.H., & Wickland, D. (2011). Characterizing 3D vegetation structure from space: Mission requirements. *Remote Sensing of Environment, 115*, 2753-2775.

Harding, D.J., & Carabajal, C.C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters, 32*, 4

Hatcher, W.G., & Yu, W. (2018). A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends. *Ieee Access, 6*, 24411-24432.

Homer, C., Dewitz, J., Yang, L.M., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the Conterminous United States - Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering and Remote Sensing, 81*, 345-354.

Hu, T., Su, Y., Xue, B., Liu, J., Zhao, X., Fang, J., & Guo, Q. (2016). Mapping Global Forest Aboveground Biomass with Spaceborne LiDAR, Optical Imagery, and Forest Inventory Data. *Remote Sensing, 8*, 565.

Lee, S., & Lee, D. (2018). Improved Prediction of Harmful Algal Blooms in Four Major South Korea's Rivers Using Deep Learning Models. *International Journal of Environmental Research and Public Health, 15*, 15.

Lefsky, M.A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters, 37*, 5.

Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B., & Hunter, M.O. (2007). Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing, 1*, 18.

Liu, K., Wang, J., Zeng, W., & Song, J. (2017). Comparison and Evaluation of Three Methods for Estimating Forest above Ground Biomass Using TM and GLAS Data. *Remote Sensing, 9*, 341.

Lu, D.S., Chen, Q., Wang, G.X., Liu, L.J., Li, G.Y., & Moran, E. (2016). A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth, 9*, 63-105.

Lv, J.J., Zhang, B., & Li, X.Q. (2019). Research on Object Detection Algorithm Based on PVANet. *Advances in Computer Communication and Computational Sciences, Vol 1, 759*, 141-151.

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., Popescu, S., Shum, C.K., Schutz, B.E., Smith, B., Yang, Y., & Zwally, J. (2017). The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sensing of Environment, 190*, 260-273.

Marselis, S., Armston, J., & Dubayah, R. (2016). *Summary of the Second GEDI Science Team Meeting*, in: National Aeronautics and Space Administration, The Earth Observer, November - December 2016, Volume 28, Issue 6, pp: 31-36.

Martino, A. (2010). ATLAS Performance Spreadsheet. http://icesat.gsfc.nasa.gov/icesat2/data/sigma/sigma_data.php (accessed: 06.04.18).

Montesano, P.M., Rosette, J., Sun, G., North, P., Nelson, R.F., Dubayah, R.O., Ranson, K.J., & Kharuk, V. (2015). The uncertainty of biomass estimates from modeled ICESat-2 returns across a boreal forest gradient. *Remote Sensing of Environment, 158*, 95-109.

Moussavi, M.S., Abdalati, W., Scambos, T., & Neuenschwander, A. (2014). Applicability of an automatic surface detection approach to micropulse photon-counting lidar altimetry data: implications for canopy height retrieval from future ICESat-2 data. *International Journal of Remote Sensing, 35*, 5263-5279.

MRLC (2017). Multi-resolution land characteristics consortium (MRLC): national land cover database. http://www.mrlc.gov/ (accessed: 02.09.17).

Narine, L.L., Popescu, S., Neuenschwander, A., Zhou, T., Srinivasan, S., & Harbeck, K. (2019). Estimating aboveground biomass and forest canopy cover with simulated ICESat-2 data. *Remote Sensing of Environment, 224*, 1-11.

Narine, L.L., Popescu S., Zhou T., Srinivasan S., & Harbeck K. (in press). Mapping Forest Aboveground Biomass with a Simulated ICESat-2 Vegetation Canopy Product and Landsat Data. *Annals of Forest Research*.

National Aeronautics and Space Administration, 2017. ICESat & ICESat-2. http://icesat.gsfc.nasa.gov/ (accessed: 06.04.18).

Nelson, R., Margolis, H., Montesano, P., Sun, G., Cook, B., Corp, L., Andersen, H.-E., deJong, B., Pellat, F.P., Fickel, T., Kauffman, J., & 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.

Neuenschwander, A., & Pitts, K. (2019). The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sensing of Environment, 221*, 247-259.

Neuenschwander, A., Popescu, S., Nelson, R., Harding, D., Pitts, K., Pederson, D., & Sheridan, R. (2017). ICE, CLOUD, and Land Elevation Satellite (ICESat-2) Algorithm Theoretical Basis Document (ATBD) for Land - Vegetation Along-track products (ATL08). In (p. 108).

Neuenschwander, A.L., & Magruder, L.A. (2016). The Potential Impact of Vertical Sampling Uncertainty on ICESat-2/ATLAS Terrain and Canopy Height Retrievals for Multiple Ecosystems. *Remote Sensing, 8*, 16.

Nunez, R.A.C., de la Pena, M., Irigollen, A.F., & Rodriguez, M.R. (2018). Deep learning models for the prediction of small-scale fisheries catches: finfish fishery in the region of "Bahia Magadalena-Almejas". *Ices Journal of Marine Science, 75*, 2088-2096.

Popescu, S.C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass & Bioenergy, 31*, 646-655.

Popescu, S.C., Zhao, K., Neuenschwander, A., & Lin, C. (2011). Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environment, 115*, 2786-2797.

Popescu, S.C., Zhou, T., Nelson, R., Neuenschwande, A., Sheridan, R., Narine, L., & Walsh, K.M. (2018). Photon counting LiDAR: An adaptive ground and canopy height retrieval algorithm for ICESat-2 data. *Remote Sensing of Environment, 208*, 154-170.

Rosen, P., Hensley, S., Shaffer, S., Edelstein, W., Kim, Y., Kumar, R., Misra, T., Bhan, R., Satish, R., Sagi, R., & Ieee (2016). AN UPDATE ON THE NASA-ISRO DUAL-FREQUENCY DBF SAR (NISAR) MISSION. *2016 Ieee International Geoscience and Remote Sensing Symposium (Igarss)*, 2106-2108.

Shafaey, M.A., Salem, M.A.M., Ebied, H.M., Al-Berry, M.N., & Tolba, M.F. (2019). Deep Learning for Satellite Image Classification. *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018, 845*, 383-391.

Shao, Z.F., Zhang, L.J., & Wang, L. (2017). Stacked Sparse Autoencoder Modeling Using the Synergy of Airborne LiDAR and Satellite Optical and SAR Data to Map Forest Above-Ground Biomass. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*, 5569-5582.

Simard, M., Pinto, N., Fisher, J.B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research-Biogeosciences, 116*, 12.

Stathakis, D. (2009). How many hidden layers and nodes? AU - Stathakis, D. *International Journal of Remote Sensing, 30*, 2133-2147.

Stysley, P.R., Coyle, D.B., Clarke, G.B., Frese, E., Blalock, G., Morey, P., Kay, R.B., Poulios, D., & Hersh, M. (2016). Laser Production for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar. In, *Conference on Laser Radar Technology and Applications XXI*. Baltimore, MD: Spie-Int Soc Optical Engineering.

Sun, X. (2012). Space-Based Lidar Systems. In, *Conference on Lasers and Electro-Optics 2012* (p. JW3C.5). San Jose, California: Optical Society of America.

Tian, X., Li, Z., Su, Z., Chen, E., van der Tol, C., Li, X., Guo, Y., Li, L., & Ling, F. (2014). Estimating montane forest above-ground biomass in the upper reaches of the Heihe River Basin using Landsat-TM data. *International Journal of Remote Sensing, 35*, 7339-7362.

Trier, O.D., Salberg, A.B., Kermit, M., Rudjord, O., Gobakken, T., Naesset, E., & Aarsten, D. (2018). Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data. *European Journal of Remote Sensing, 51*, 336-351.

Wang, L., Liu, J.P., Xu, S.H., Dong, J.J., & Yang, Y. (2018). Forest Above Ground Biomass Estimation from Remotely Sensed Imagery in the Mount Tai Area Using the RBF ANN Algorithm. *Intelligent Automation and Soft Computing, 24*, 391-398.

Xie, R., Wen, J., Quitadamo, A., Cheng, J.L., & Shi, X.H. (2017). A deep auto-encoder model for gene expression prediction. *Bmc Genomics, 18.*

Zald, H.S.J., Wulder, M.A., White, J.C., Hilker, T., Hermosilla, T., Hobart, G.W., & Coops, N.C. (2016). Integrating Landsat pixel composites and change metrics with lidar plots to predictively map forest structure and aboveground biomass in Saskatchewan, Canada. *Remote Sensing of Environment, 176*, 188-201.

Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine, 5*, 8-36.

5. CONCLUSIONS

The ICESat-2 mission offers an opportunity to contribute to the up-to-date characterization of vegetation structure, including spatially explicit assessments of AGB density which can be used to support sustainable forest management and reduce uncertainty in quantifying terrestrial carbon. In this study, the utilization of simulated ICESat-2 PCL data along planned track locations over SHNF for modeling and mapping AGB and estimating canopy cover was demonstrated. In addition to applying the expected photon detection rate over temperate forests, the impact of the atmosphere and solar background noise detected by the ATLAS instrument was considered in the PCL simulations with the analysis of three data scenarios representative of different noise levels; daytime, nighttime and no noise. To maintain consistency with the data format provided by ICESat-2's ATL08 land and vegetation product, the scenarios were processed to report height, canopy cover and canopy density metrics for 100 m along-track segments. Results from multiple regression models highlighted the predictive ability of the simulated PCL-derived metrics to characterize forest vegetation, with R^2 values of ranging from 0.56 to 0.83 with the daytime and nighttime scenarios. In particular, the ability of the nighttime scenario to predict AGB was comparable to simulated data without the addition of noise, yielding a R^2 0.79 and RMSE of 19.23 Mg/ha.

Using predicted AGB and mapped predictors consisting of Landsat reflectance data, landcover and canopy cover, AGB maps and AGB uncertainty maps were

produced at 30 m spatial resolution with RF. The integration of Landsat data and derived products for AGB model and map production highlighted canopy cover as the most important predictor of AGB in every scenario and indicated similarities in the predictive capabilities of RF models for no noise and nighttime scenarios, which outperformed the daytime scenario. RF regression tree models explained 42%, 49% and 51% of the variance with test data for the daytime, nighttime and no noise scenarios with RMSE values of 19.69 Mg/ha, 19.30 Mg/ha and 19.72 Mg/ha. Uncertainty in the resulting maps, estimated as the variability of predictions from 1000 independent trees used to estimate pixel AGB, emphasized areas with higher values in the daytime scenario than the corresponding nighttime scenario.

The investigation of DNNs for mapping AGB highlighted the potential of this approach for mapping AGB by integrating simulated PCL-estimated AGB with Landsat imagery, NLCD canopy cover and land cover. Findings revealed the effect of neural network structure and learning rate on model performance with the latter having sizeable impacts on reported error metrics. Overall, wider architectures (at least 300 neurons) with two to three hidden layers yielded results comparable to those achieved with RF. Models achieved R^2 values and RMSEs of 0.42 and 19.55 Mg/ha, 0.49 and 19.30 Mg/ha and 0.50 and 19.72 Mg/ha for the daytime, nighttime and no noise scenarios respectively. With additional training data, results may improve and deeper architectures could be investigated. Nevertheless, findings with simulated PCL-estimated AGB, especially with the nighttime scenario, highlight the potential for generating a wall-to-

wall AGB product with ICESat-2 by adopting a synergistic approach with Landsat data and derived products, using DNNs.

Given the coverage of ICESat-2 profiles over vegetation and free of charge and open access data used in this study, upscaling approaches could be implemented at larger spatial scales, such as regional and continental extents. In doing so, ancillary data such as climatic and topographic variables may be examined for improving AGB predictions. Similar performance of RF and DNN models also suggest that multiple techniques or approaches should be explored for generating an AGB product with ICESat-2. Additionally, with growing volumes of data from current space lidars, such as ICESat-2 and GEDI, as well as upcoming satellite missions (e.g. BIOMASS mission) and the learning capacity of DL, DL algorithms could be further explored in future research.