EMBRY-RIDDLE
Aeronautical University
SCHOLARLY COMMONS

PhD Dissertations and Master's Theses

Spring 2022

Locality of Topographic Ground Truth Data for Salt Marsh Lidar DEM Elevation Bias Mitigation Using Machine Learning

Joelle S. Bobinsky
Embry-Riddle Aeronautical University, bobinsj1@my.erau.edu

Follow this and additional works at: https://commons.erau.edu/edt

Scholarly Commons Citation
Bobinsky, Joelle S., "Locality of Topographic Ground Truth Data for Salt Marsh Lidar DEM Elevation Bias Mitigation Using Machine Learning" (2022). *PhD Dissertations and Master's Theses*. 661.
https://commons.erau.edu/edt/661

This Thesis - Open Access is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in PhD Dissertations and Master's Theses by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

This thesis is dedicated to my mom, Brenda. Thank you for sacrificing so much so my sisters and I could get to where we are. You're an amazing mom.

1 ACKNOWLEDGEMENTS

This thesis is a result of research funded by the National Oceanic and Atmospheric Administration's National Centers for Coastal Ocean Science under award NA16NOS4780208. This work has also been funded in part by NOAA's NCCOS Effects of Sea Level Rise (ESLR) Competitive Research Program. They fully supported my stipend and covered the cost of tuition during the summer of 2021 for which I am extremely grateful. I would also like to acknowledge ASPRS for giving me the opportunity to present this research at the 2021 Fall Workshop.

I appreciate the support and assistance that I have received throughout the writing of this thesis. Especially for my thesis advisor, Dr. Medeiros, Assistant Professor of Civil Engineering, who has guided me in my research endeavors and given invaluable career advice. Thank you for being the perfect balance between challenging and encouraging. It has truly been a pleasure to work for you.

I would also like to thank my committee members, Dr. Jeff Brown and Dr. Marwa El-Sayed for their sincere feedback on my thesis. Your dedication to your students does not go unnoticed. I would also like to acknowledge the entire faculty in the Civil Engineering Department for their support throughout my undergraduate and graduate studies. Thank you to Rosa Criado for all of your help throughout this process and for your dedication to the department.

Thank you to Scott Steman, your advice has been extremely helpful and I honestly wouldn't be in this position if it weren't for your support. Thank you for being so generous with your experience and for your commitment to guiding the next generation of engineers.

Also, thank you to all my friends for giving me boosts of serotonin when I needed it most. Especially to Kelly, Gianna, Chance and all the fellows at the Center for Faith and Spirituality**,** as well as Dan, John, Eilidh and the rest of the climbing squad, to my ultimate Frisbee teammates especially Joe, Sophie, Lynn, Chuck, and Stephen. Thank you for being a friend.

Thank you to my parents, for inspiring me to reach my full potential. You are the best support system and I am so grateful to have you as my parents. Thank you for sticking by my side through it all, you were right all along. I love you.

To my sisters, thank you so much for being the positive voice in my ear when I had thoughts of failure. I love you so much and I'm so thankful I get to go through this life with you two by my side. I couldn't have done it without you.

2 ABSTRACT

Light detection and ranging (lidar) digital elevation models (DEMs) are crucial for modeling coastal salt marsh systems, simulating the coastal dynamics of sea level rise (SLR), and predicting storm surge inundation depth and duration. Improvements in lidar acquisition technology and data processing over the last decade have led to increased accuracy. However, the lidar-derived DEMs for coastal salt marshes that are densely vegetated are generally unreliable without adjustment based on local ground truth elevations. In this study, Random Forest (RF) DEM adjustment models are trained for two similar Northern Gulf of Mexico salt marshes. The need for local topographic ground truth data to train the models is also investigated. Two Real-Time Kinematic (RTK) GNSS field surveys were conducted by others to acquire ground truth elevations near St. Marks, Florida (n=377) and Pascagoula, Mississippi (n=610). These elevations, along with lidar elevations and Sentinel-2A multispectral satellite imagery (MSI) reflectance values were used to train the RF salt marsh DEM adjustment models and apply them under two scenarios: local and non-local. A local adjustment relies on data collected within the adjustment domain to train the model whereas a non-local adjustment uses data collected outside the adjustment domain. The RF-local models achieved the lowest mean absolute error (MAE) values for St. Marks and Pascagoula. The predictions using non-local RF models were unsatisfactory. The evidence suggests that local ground truth data are necessary for mitigating bias in salt marsh lidar DEMs, although future work should investigate if increasing the data set size could narrow the accuracy gap. This mitigation adjustment technique can be replicated in other coastal regions with similar vegetation profiles. As the world becomes increasingly vulnerable to the effects of climate change and SLR, it is important to accurately characterize the current state of the system to model marsh restoration and migration, natural and nature based protective infrastructure, and land use planning policies, for example.

3 TABLE OF CONTENTS

4 LIST OF FIGURES

5 LIST OF TABLES

6 NOMENCLATURE

AGBM	Above-ground biomass
AIC	Akaike's Information Criteria
BIC	Bayesian Information Criteria
ANN	Artificial Neural Network
BOA	Bottom of Atmosphere
C-CAP	Coastal Change Analysis Program
DEM	Digital elevation model
ESA	European Space Agency
ESLR	Effects of sea-level rise
FDOT	Florida Department of Transportation
FPRN	Florida Permanent Reference Network
GCGN	Gulf Coast Geospatial Center
GSF	Geoid separation file
HydroMEM	Hydrodynamic Marsh Equilibrium Model
Lidar	Light detection and ranging
LR	Linear Regression
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MEM	Marsh Equilibrium Model
MHW	Mean high water
MSI	Multi-Spectral Instrument
NAD83	North American Datum of 1983

NAVD88	North American Vertical Datum of 1988
NCCOS	National Centers for Coastal Ocean Science
NGOM	Northern Gulf of Mexico
NIR	Near-infrared
NOAA	National Oceanic and Atmospheric Administration
nRMSE	Normalized Root Mean Square Error
NWI	National Wetland Inventory
RF	Random Forest
RGB	Red, green, blue
RMSE	Root Mean Square Error
RTK-GNSS	Real Time Kinematic - Global Navigation Satellite System
RTCM	Real Time Correction Message
S2A	Sentinel 2A
SLR	Sea level rise
TOA	Top of Atmosphere
UTM	Universal Transverse Mercator
VRS	Virtual Reference Station

7 Introduction

Lidar digital elevation models (DEMs) are known to be inaccurate in coastal salt marshes mainly because the laser pulses cannot always reflect off the true marsh platform surface (Figure 7.1). The presence of standing water and dense vegetation (typically tall grasses with peak seasonal heights greater than 1 m) are the primary causes for the widespread biases in topographic elevation data products such as point clouds and bare earth DEMs (Alizad et al., 2016; Alizad, Medeiros, Foster-Martinez, & Hagen, 2020; Buffington, Dugger, Thorne, & Takekawa, 2016; Hladik & Alber, 2012)

7.1 Importance of Research

Salt marsh vegetation has adapted to conditions that are inundated with water and exposed to the air according to the tides. Sea level rise causes coastal salt marshes to recede because the marsh vegetation needs time without water to grow. If the vegetation is constantly inundated, the plants will eventually drown and die. Therefore, it is critical for sea level rise models to accurately characterize the marsh system as either inundated or exposed to the air. This relies on accurate elevation representations of the marsh platform. However, lidar DEMs are inaccurate in coastal salt marshes by a magnitude of about 0.1 - 0.6m (Table 8.1) depending on the location of the DEM (Hladik & Alber, 2012; Medeiros, Hagen, Weishampel, & Angelo, 2015).

The ramifications of the persistent elevation bias are evident when lidar DEMs are used as input data for marsh evolution models in microtidal environments (Alizad et al., 2020). Models such as the Marsh Equilibrium Model (MEM) (J. Morris, Sundareshwar, Nietch, Kjerfve, & Cahoon, 2002) and HydroMEM rely on the marsh platform topography to serve as a starting point for future projections of marsh migration and evolution (Alizad et al., 2016; J. Morris, 2007). When the initial state of the marsh platform is erroneously biased to a higher elevation

than the upper part of the tidal inundation frame (mean high water or MHW), the modeled marsh is not accurately inundated in the simulations and subsequently does not receive a realistic sediment loading. This was the case in the northern Gulf of Mexico when lidar data from 2007-2008 were used to model a marsh with a tidal range of approximately 34 cm (Alizad et al., 2016). Unless this bias is addressed, marsh biomass density and zonation projections will be inaccurate from the start and all emergent effects from future conditions such as sea level rise (SLR) will be unreliable.

Accurate representation of coastal terrain is also an important factor in hurricane storm surge modeling, regardless of the local tide range (Bilskie, Coggin, Hagen, & Medeiros, 2015; Bunya et al., 2010; Dietrich et al., 2011). Coastal salt marshes are often the first or second physical buffer zone standing between an incoming storm surge and upland property. Therefore, it exerts substantial influence over the propagation (depth, extent, and timing) of storm surge inundation. Since this is a major cause of the destruction and economic disruption associated with tropical cyclones, accurate coastal flood predictions for both immediate event-scale decisions such as evacuation orders as well as longer term risk assessments and resilient infrastructure planning rely on coastal DEMs as a primary input data source (Baradaranshoraka, Pinelli, Gurley, Peng, & Zhao, 2017).

This thesis presents the development of a random forest (RF) model structure for mitigating the elevation bias in lidar DEMs and applies it to two ecologically similar estuarine salt marsh systems in the northern Gulf of Mexico. Sentinel-2A multispectral imagery, the lidar DEM elevations, and field measured topographic elevations were used to train, test, and validate the models. The trained RF models were also tested at the non-local site to investigate the necessity of obtaining local spot elevations in marshes for the purpose of developing a bias mitigation

model. The overall objective of the lidar DEM bias mitigation effort was to provide a topographically accurate marsh platform model for two locations to be used in simulations of salt marsh evolution and migration in response to sea level rise.

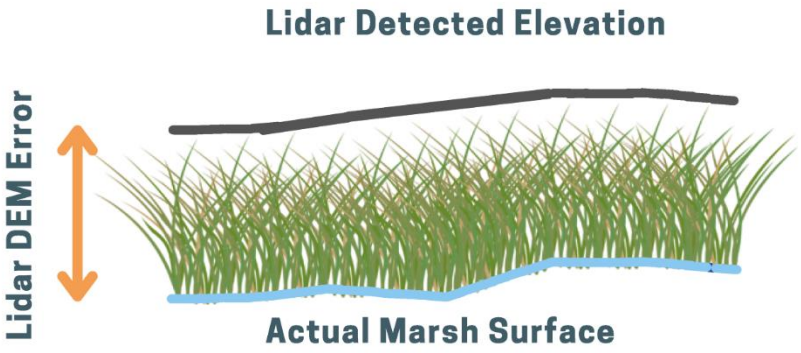

Figure 7.1: The DEM bias due to dense vegetation. The laser bounces off of the vegetation and is absorbed by standing water, which causes the error. (Martin et al., 2022)

8 Review of the Relevant Literature

This chapter will cover how others have adjusted for the bias in lidar derived DEMs and compare how researchers use machine learning techniques for environmental applications. There are numerous ways to predict the bias including using different predictors as well as using a different prediction method.

8.1 Lidar Digital Elevation Model Adjustment

To mitigate the persistent high elevation bias in salt marsh lidar DEMs, corrections must be applied to the marsh surface topography. However, it is impractical to adjust multi-county or regional scale DEMs based on field data alone. To address this, techniques have been developed that rely on vegetation characteristics such as height (Hladik & Alber, 2012), remotely sensed biomass density (Medeiros, Hagen, Weishampel, et al., 2015), local tidal frame elevations (Alizad et al., 2018), and lidar waveform data (Rogers, Parrish, Ward, & Burdick, 2018). In addition to the high elevation bias, lidar DEMs also tend to flatten out the underlying microtopography including small tidal creeks, making the spatial distribution of the bias magnitudes non-linear (Medeiros, Hagen, & Weishampel, 2015). Approaches such as random forest (Alizad et al., 2016; Cooper, Zhang, Davis, & Troxler, 2019) multiple regression (Medeiros, Hagen, Weishampel, et al., 2015), and gradient boosted nonparametric regression (Rogers et al., 2018) have been shown to be effective in the past and the research is trending towards simpler models that require fewer field measured vegetation characteristics for model training.

Hladik and Alber (2012) conducted a correction of a lidar derived salt marsh digital elevation model for an area in Georgia. Their results show in comparison to the RTK ground truth data, the lidar contained a bias of 0.03 to 0.25 m depending on the cover class. They developed correction factors according to the type of vegetation. The species-specific corrections reduced the overall

DEM error from 0.10 ± 0.12 m (SD) to -0.01 ± 0.09 m (SD) and the root mean square error (RMSE) from 0.16m to 0.10m. Medeiros, Hagen, Weishampel, et al. (2015) used remotely sensed biomass density to adjust the lidar-derived DEM for the Apalachicola River Marsh area. The adjustment resulted in raw mean errors for the lidar DEM and the adjusted DEM as 0.61 ± 0.24m and 0.32 ± 0.24m, thereby reducing the high bias by approximately 49%. Further examples of lidar error reported in previous salt marsh modeling studies can be found in Table 8.1 (Alizad et al., 2020)

Table 8.1: Lidar error reported in previous salt marsh modeling studies

Study	Location	Lidar Error (m)
Hladik and Alber (2012)	Sapelo Island, GA	0.03 to 0.25
Medeiros, Hagen, Weishampel, et al. (2015)	Apalachicola, FL	0.61 ± 0.24
J. T. Morris et al. (2005)	North Inlet, SC	0.13 ± 0.065
Schmid, Hadley, and Wijekoon (2011)	Charleston, SC	0.153 ± 0.176
Thorne, Elliott-Fisk, Wylie, Perry, and Takekawa (2014)	San Pablo Bay, CA	0.10 to 0.35
Fernandez-Nunez, Burningham, and Ojeda Zujar (2017)	Odiel, Spain	0.23 ± 0.13 to 0.45 ± 0.19

8.2 Machine Learning Techniques

Cooper *et al* (2019) compared different machine learning techniques including Random Forest (RF), Support Vector Machine, k-Nearest Neighbor, and Artificial Neural Network (ANN). The RF models emerged as one of the most useful due to their computational efficiency, resistance to overfitting, ability to handle small datasets, and explainability (trained RF models natively produce feature importance metrics using out-of-bag testing). Rodriguez-Galiano,

Mendes, Garcia-Soldado, Chica-Olmo, and Ribeiro (2014) used a RF model to build different predictive models of nitrate pollution. An advantage of the RF in this study was its ability to determine variable importance. This feature was used to define the most significant predictors of nitrate pollution in groundwater using remotely sensed and in-situ data. The variable importance feature of the random forest makes it a useful machine learning tool because it allows for predictor comparison.

Although many different machine learning techniques are used to conduct adjustments, the random forest technique is especially useful in environmental applications where the training datasets are relatively small. Singh, Sihag, and Singh (2017) used a RF regression model in comparison with M5P model tree, and ANN to model the impact of water quality on infiltration rate of soil. Their results show that the RF method was the best performing model in terms of predictive accuracy. The model was trained using only 132 field measurements. Mascaro et al. (2014) used the random forest model in a different application and used remotely sensed data rather than field measurements. The main purpose of the study was to evaluate the performance of the random forest regression method for tropical forest carbon mapping. Their results conclude that spatial context should be considered when training an RF and adjustments may need to be made to avoid over fitting the data. However, the RF model with spatial context outperformed the stratification approach, which is the traditional method used in carbon stock modeling.

Additionally, Belgiu and Drăguţ (2016) discuss how the RF classifier is particularly beneficial for remote sensing purposes, including studies with multi-source data. Hu *et al.* (2020) used a RF regression model with data from remote sensing and field measurements to estimate aboveground biomass in order to produce a global mangrove forest above-ground biomass

(AGBM) map. This work supported the findings of Fassnacht *et al.* (2018) who concluded that combining lidar data with many reference sample units and a random forest model produce biomass predictions with the lowest error.

8.3 Summary

Previous work has shown that in coastal salt marshes, the bias associated with remote sensing data needs to be mitigated. The adjustment can be conducted in a number of ways. However, machine learning and specifically a random forest technique outperforms other models especially for environmental studies where the training dataset is relatively small compared to typical machine learning data sets. Ground truth data and remotely sensed data have been used to make predictions using a random forest. Additionally, ground truth data have been shown to enhance predictive results of DEM error (Buffington et al., 2016; Hladik & Alber, 2012; McClure, Liu, Hines, & Ferner, 2015). However, collecting ground truth data in salt marshes is extremely labor intensive and costly.

We will explore the predictive performance of a model trained on data from a different, but ecologically similar, location. A random forest model using remotely sensed data as predictors will be trained on in-situ data from two locations. Once the two models are trained, they will be used to predict the error for the other location.

9 Methodology

DEMs for two regions in the northern Gulf of Mexico were adjusted: Apalachee Bay in Florida and the lower Pascagoula River in Mississippi. The adjustment was achieved using a RF machine learning model that was trained on satellite imagery including red, green, blue, (RGB) and near-infrared (NIR) (Table 9.1) spectral bands and the lidar-derived DEM elevation to predict the error in the DEM. In-situ Real Time Kinematic - Global Navigation Satellite System (RTK-GNSS) spot elevation points acquired by others were used as ground truth data. The predicted error was then used to adjust the original DEM. To test the need for local topographic data for DEM adjustment, the Pascagoula RF model was applied to Apalachee Bay's data, and vice versa.

9.1 Research Setting

Both research settings were identified as areas of interest for the Effects of Sea Level Rise Program (ESLR) from NOAA's National Centers for Coastal Ocean Science (NCCOS). The first setting for this study is the Apalachee Bay / St. Marks region of Florida's northern Gulf of Mexico (NGOM) coast including parts of Gulf, Franklin, Wakulla, Jefferson, and Taylor counties (listed west to east, Figure 9.1). The second area of interest for this study is the lower Pascagoula River estuary in Mississippi and Alabama lying primarily in Jackson (MS) and Mobile (AL) counties. These locations will be referred to as APAL and PASC, respectively.

Figure 9.2 shows the polygons used to mask out the areas identified as wetlands. More specifically, the wetland types in this area are classified as estuarine and marine (Figure 9.2, magenta), and freshwater emergent wetlands (Figure 9.2, blue). Both locations are similar ecologically, primarily dominated by *Juncus roemerianus*, or black needle rush, with *Spartina alterniflora* and *Spartina cynosuroides* on the fringe of the open water areas. The NWI polygons

representing the wetlands areas of interest were used to extract sections of the DEM for adjustment.

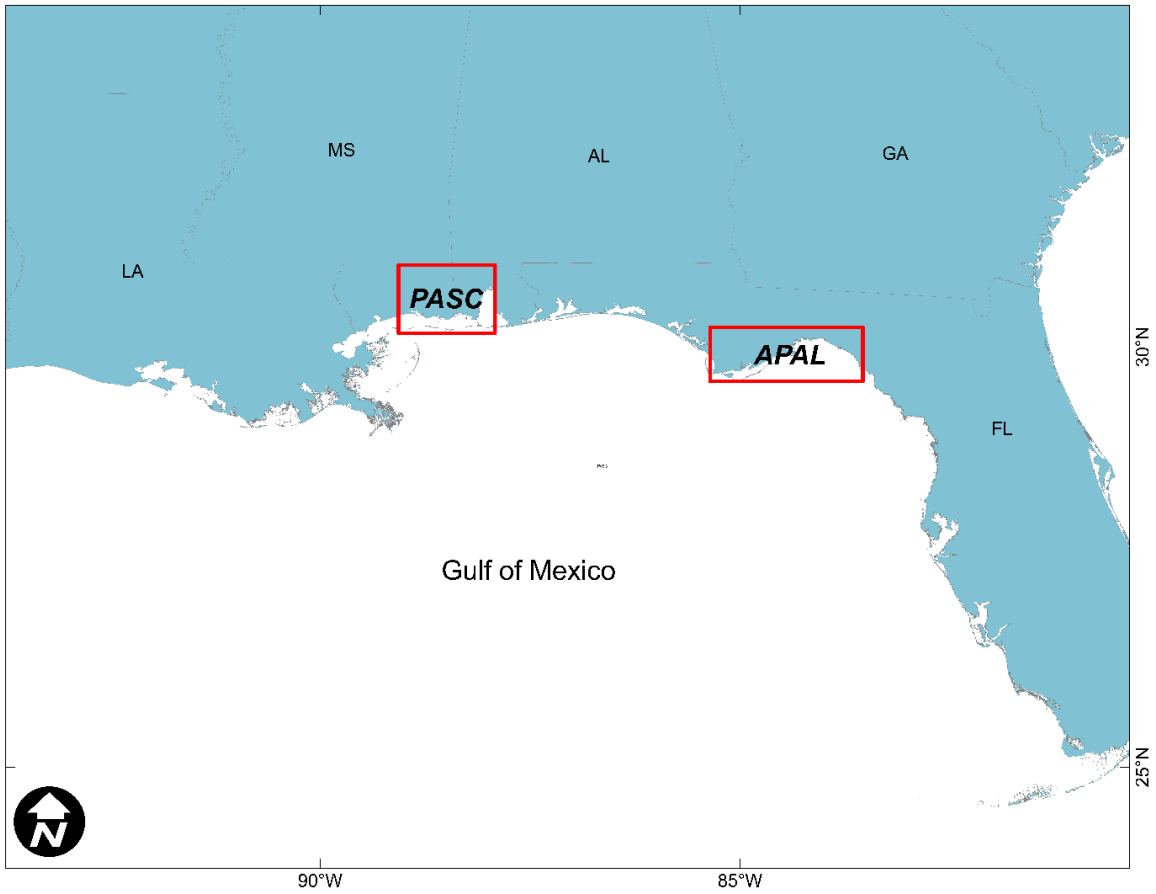

Figure 9.1: Research settings in Pascagoula, MS (PASC) and Apalachee Bay, FL (APAL).

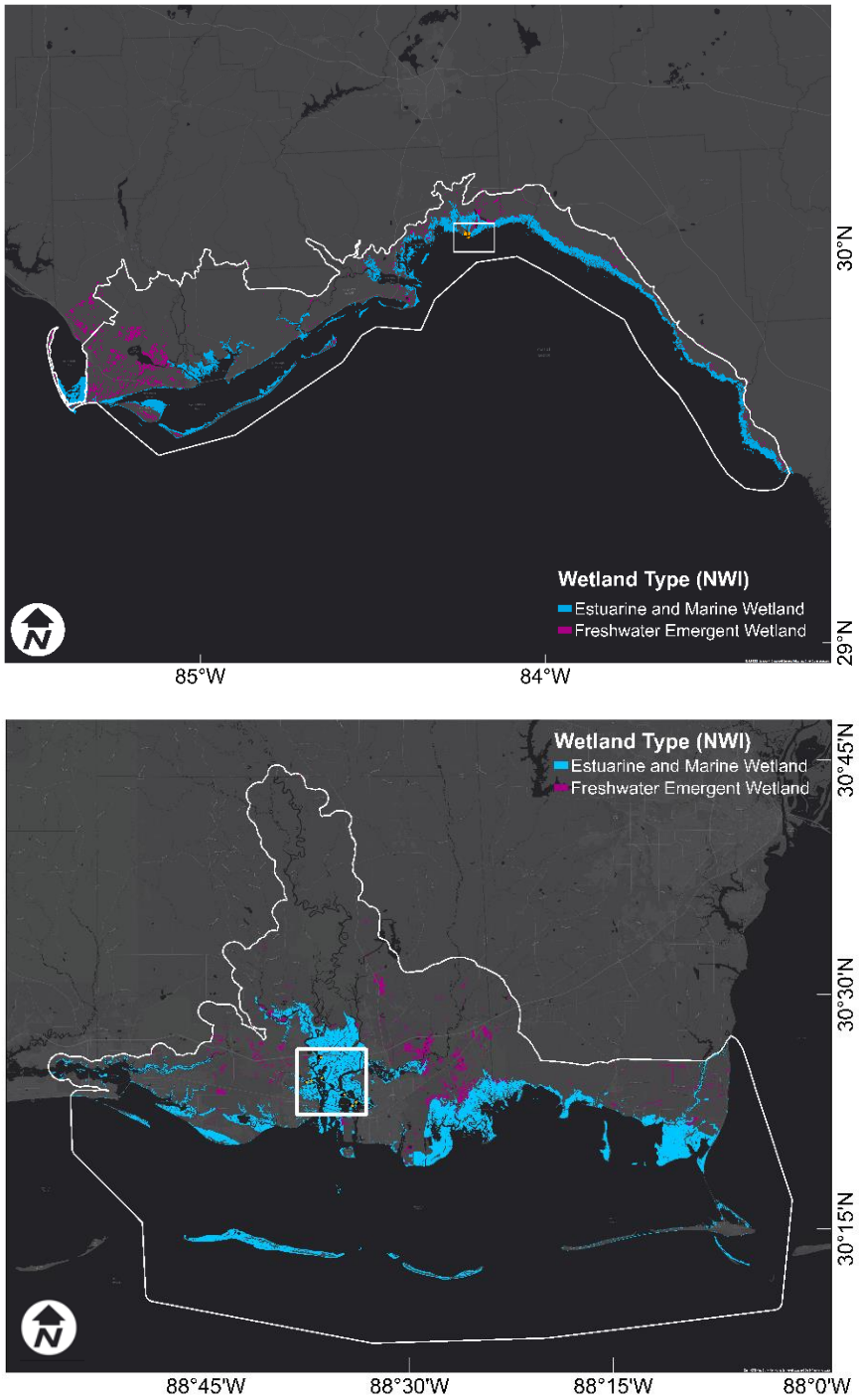

Figure 9.2: The adjustments for Apalachee Bay (top) and Pascagoula (bottom) were constrained to freshwater emergent wetlands (magenta) and estuarine and marine wetlands (blue). The white squares indicate where ground truth data was collected.

9.2 Lidar Digital Elevation Model – Apalachee Bay

The lidar DEM used as the basis for the bias mitigation effort for the Apalachee Bay region was downloaded from NOAA Digital Coast (Digital Coast, 2022). The lidar DEM used in this study was compiled from three separate projects (Figure 9.3). The westernmost area was covered by the 2-meter resolution Lower Choctawhatchee lidar DEM collected between 9 April and 17 May 2017 (Dewberry, 2018b). The central area was covered by the 1-meter resolution DEM for the Florida Panhandle collected between 31 March and 10 May 2018 (Dewberry, 2018a). The eastern end of the study area used an older 5-meter resolution dataset from the Florida Division of Emergency Management 2007 DEM (Dewberry, 2008). This older dataset was used due to the lack of more recent data in this area. These three DEMs were reprojected to NAD83(2011) Florida State Plane North in meters (EPSG: 6440), resampled and co-registered with 5-meter resolution, and their elevations were converted to meters NAVD88. They were mosaicked together with priority given to the more recent data (2018, 2017, and then 2007).

The mosaicked DEM was clipped to just the coastal areas (Figure 9.3) and checked for discontinuities using topographic profile transects generated in ArcMap, especially at the boundary between the 2018 and 2007 DEMs (green and blue areas, respectively). This boundary is also the upland boundary for the marsh evolution model grid that is the end-use of the adjusted DEM. There were no abrupt discontinuities that would indicate an error in the combined DEM. The 2007 nearshore region had a manually assigned elevation of zero rather than NoData and since this area was not within the wetland areas defined by the NWI, these areas were not included in the adjustment.

9.3 Lidar Digital Elevation Model – Pascagoula River

The lidar DEM used as the basis for the bias mitigation effort in the lower Pascagoula River region was developed by the United States Geological Survey (USGS) and is also freely available

for download (Digital Coast, 2022). As Figure 9.4 shows, the 3-meter resolution CoNED Topobathymetric DEM collected in 2014 (Partners, 2014) covered the entire area of interest in Pascagoula, so no other sources were needed. This DEM was reprojected to NAD83(2011) UTM Zone 16N (EPSG: 6345) and the elevations were converted to meters NAVD88. The DEM was then clipped to include the Pascagoula River coastline as well as the coastal areas and barrier islands at the mouth of the river (Figure 9.4).

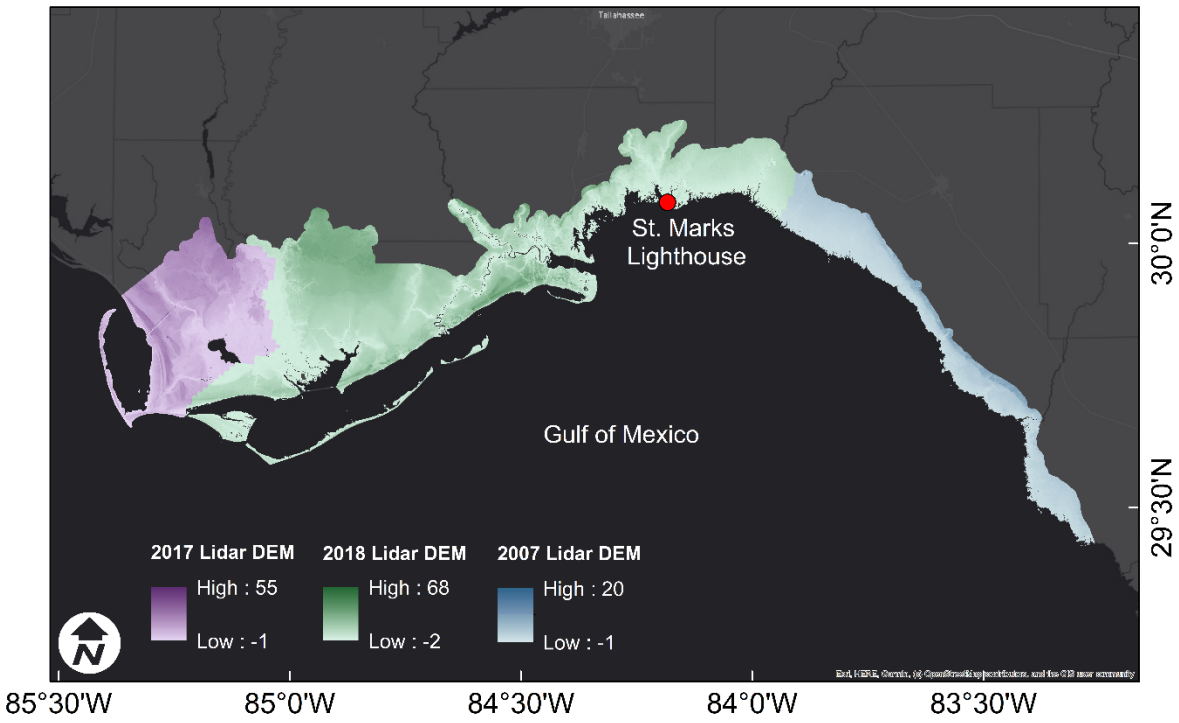

Figure 9.3: Lidar DEM products used as the bases for bias mitigation in the Apalachee domain. The western (purple), central (green), and eastern (blue) lidar were acquired in 2017, 2018, and 2007, respectively.

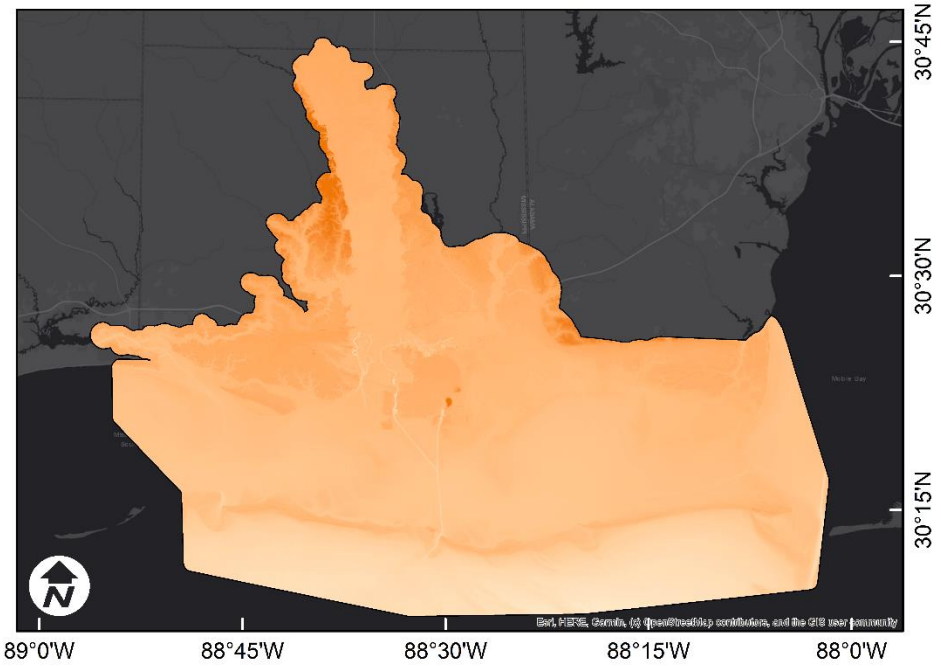

Figure 9.4: Lidar DEM of the Pascagoula River region from USGS 2014 survey

9.4 In-Situ Topographic Data Acquisition

Others collected the in-situ topographic data for the Apalachee Bay region in the salt marsh and adjacent upland areas around the St. Marks lighthouse. Figure 9.5a shows the data points in orange which were collected using RTK-GNSS survey equipment in March of 2018. The surveys were completed using a wide top shoe, 2.5” in diameter, affixed to the bottom of the survey rod. The rod was not allowed to sink into the soft surface of the marsh but rather is held suspended where the surveyor feels resistance from the sediment and rhizome. Virtual Reference Station (VRS) corrections were obtained from the Florida Permanent Reference Network (FPRN) maintained by the Florida Department of Transportation (FDOT) in Real Time Correction Message (RTCM) version 3.1 format. A total of 377 spot elevations were collected across various transects located with an emphasis on capturing the topographic gradient from the water surface to the high marsh and uplands.

The in-situ topographic data for the Pascagoula region were collected in a similar manner using RTK-GNSS in March of 2019. Single base corrections were obtained from the Gulf Coast Geospatial Center (GCGC) Real Time Network operated by the University of Southern Mississippi in RTCM 3.1 MAX format. Figure 9.5b shows the 610 spot elevations that were collected in Pascagoula along transects designed to capture topographic gradient like the protocol in St. Marks. Topographic data at both sites were converted to orthometric heights in meters NAVD88 in real time by using a geoid separation file (GSF) of the Continental US based on GEOID12B.

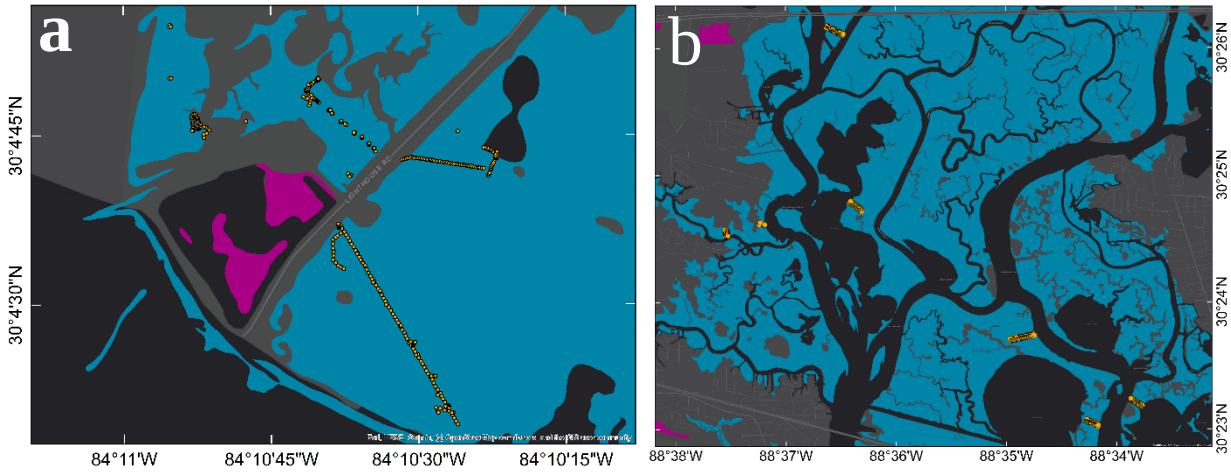

Figure 9.5: In-situ topographic data points are shown in orange for Apalachee Bay (left) and Pascagoula (right). In-situ data was collected only in wetlands identified as Freshwater Emergent Wetlands (magenta) and Estuarine and Marine Wetlands (blue). These show the areas identified by the white rectangles in Figure 9.2.

9.5 Satellite Imagery Acquisition and Processing

Sentinel-2A MSI Level 1C data were downloaded from the European Space Agency (ESA) Copernicus website. The following are the tile numbers (Txxxxx) and acquisition dates for the satellite imagery.

Apalachee Bay:

- T16RFT, 8 March 2018
- T16RFU, 31 March 2018
- T16RGT, 8 March 2018
- T16RGU, 8 March 2018
- T17RKN, 28 March 2018
- T17RKP, 8 March 2018

Pascagoula:

- T16RCU, 19 March 2019
- T16RCV, 19 March 2019

The images were chosen because they were the most cloud-free and were captured at roughly the same time as the ground truth topographic data. Each scene was downloaded at Level 1C top of atmosphere (TOA) reflectance and processed to Level 2A bottom of atmosphere (BOA) reflectance products using the Sen2Cor software provided by ESA. For each scene, bands 2, 3, 4 and 8 (Table 9.1) were extracted (Thales Alenia Space, 2021), mosaicked, and reprojected to NAD83(2011) Florida State Plane North for Apalachee Bay, and NAD83(2011) UTM zone 16N for Pascagoula, using the ESA SNAP software. These bands were selected because they are available at 10 m resolution, the other bands were available at either 30 or 60m resolution. Additionally, these bands are commonly used in remote sensing in studies of soils or vegetation.

The red and NIR bands especially are used to indicate vegetation health in the canonical Normalized Difference Vegetation Index (Tahsin, Medeiros, & Singh, 2021) and its derivatives.

Table 9.1: Spatial Resolution Bands (10m) for Sentinel 2A satellite imagery used for training the random forest.

Band Number	Central Wavelength (nm)	Bandwidth (nm)
B2 (Blue)	492.4	66
B3 (Green)	559.8	36
B4 (Red)	664.6	31
B8 (Near-Infrared)	832.8	106

9.6 Geospatial Data Fusion

All data mentioned previously including the lidar-derived DEM, in-situ topographic data, and satellite imagery, were integrated using ArcGIS to produce the training and validation data sets for the bias mitigation machine learning models.

First, the topographic spot elevations were converted to a point feature class containing xyz coordinates measured in the field. The lidar DEM elevations along with the reflectance values from the four multispectral satellite imagery spectral bands were interpolated onto the in-situ spot elevations. The difference between the field measured marsh platform elevation and lidar DEM elevation was calculated and added as a field; this represents the DEM error and serves as the label in the machine learning model explained below. The following is a list of the fields in the point data corpus (Note that * indicates predictor fields and ** indicates the target value or label for training the machine learning model):

- Northing (meters), field measured y coordinate
- Easting (meters), field measured x coordinate
- Elevation (meters), field measured z coordinate

- DEM_Elevation (meters)*, elevation interpolated from lidar DEM
- B2_Blue*, representing the 492.4 nm band reflectance
- B3_Green*, representing the 559.8 nm band reflectance
- B4_Red*, representing the 664.6 nm band reflectance
- B8_NIR*, representing the 832.8 nm band reflectance
- ERROR**, elevation difference between the lidar DEM and field measured elevation, calculated from other fields as DEM_Elevation – Elevation

The BOA S2A MSI reflectance data are provided in digital number (integer) format. The S2A MSI specifications state that these integer reflectance values are computed by multiplying the floating-point reflectance values by a quantification value (Thales Alenia Space, 2021). In the metadata for all scenes used in this study, the quantification value is 10,000. Therefore, this was the value used to convert BOA reflectances to floating point values, which can be used as input to a machine learning model without normalization.

The projected, clipped and NWI-masked lidar DEM, which represents the collective set of points that need to be adjusted, was converted to a point feature class in ArcGIS. Like the field data, the reflectance values from the four satellite imagery spectral bands were interpolated onto the points to form the application data corpus. All the predictor fields listed above, along with Northing and Easting, are present in the application data corpus.

9.7 Elevation Adjustment Model

Other machine learning models including Support Vector Machine, k-Nearest Neighbor, or Artificial Neural Network were considered, but Random Forest (RF) has proven to make the best predictions in marsh environments (Cooper et al., 2019). Therefore, the elevation adjustment used to mitigate the bias in the lidar DEM was determined using an RF. An RF is an ensemble technique, consisting of many decision trees where each decision tree is trained on a random

subsample of the data to make either a classification or a regression prediction (Belgiu & Drăguţ, 2016). Each tree makes its own prediction, which are aggregated to produce a single prediction from the forest (Figure 9.6). For a classification problem, each tree produces one vote for the classification and the aggregate prediction is the majority vote. However, the DEM error prediction requires a regression model, so each tree produces one numeric prediction and these are averaged to produce one aggregate prediction for the forest.

The following is a strongly simplified example of the inner workings of an RF model. In addition to demonstrating the training process, we will also simulate sending an unlabeled point down the tree to produce a prediction. Each decision tree is "grown" using a random subsample of the predictors as well as a random subsample of the data.

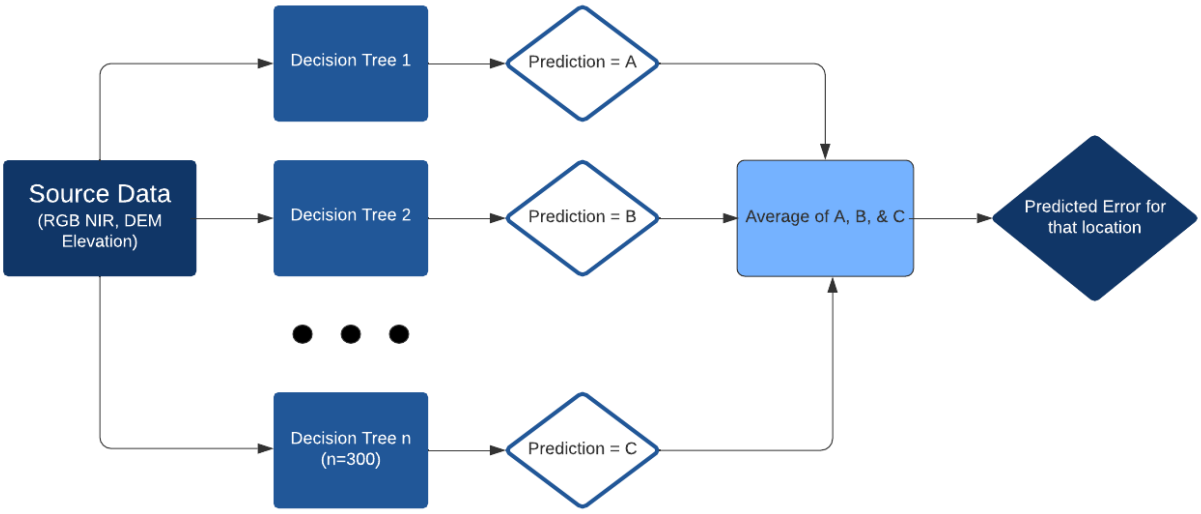

Figure 9.6: In a random forest, the source data (RGB, NIR, and DEM elevation) filters through the decision trees to produce a single prediction from each tree. The predictions from each tree are averaged to produce the predicted error for that location. The process is repeated until there is an adjustment for every location.

Table 9.2: Random subsample of predictors and sample data from the training data corpus taken with replacement.

		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10
Predictors	B2	0.150	0.131	0.197	0.183	0.162	0.170	0.186	0.138	0.112	0.187
	B8	0.211	0.233	0.217	0.389	0.27	0.266	0.288	0.287	0.277	0.36
	DEM Elevation	0.091	0.103	0.105	0.155	0.07	0.087	0.2416	0.131	0.109	0.038
Label	True DEM Error*	0.021	0.083	0.0865	0.0041	0.023	0.060	0.1147	0.084	0.089	0.0058

** True DEM Error = DEM Elevation – Ground Truth Elevation*

*Table 9.3: S*ubsample B2 reflectances, sorted from smallest to largest.

	Sample 9	Sample 2	Sample 8	Sample 1	Sample 5	Sample 6	Sample 4	Sample 7	Sample 10	Sample 3
B2 reflectance	0.112	0.131	0.138	0.150	0.162	0.170	0.183	0.186	0.187	0.197

Table 9.4: Potential splits for B2 reflectances

	Sample 9	Sample 2	Sample 8	Sample 1	Sample 5	Sample 6	Sample 4	Sample 7	Sample 10	Sample 3
TOTAL	0.112	0.131	0.138	0.150	0.162	0.170	0.183	0.186	0.187	0.197
0.0282	0.0091		0.0191							
0.0259	0.0107			0.0152						
0.0252	0.0136				0.0116					
0.0256	0.0169					0.0086				
0.0250	**0.0195**						**0.0055**			
0.0278	0.0226							0.0052		
0.0294	0.0243								0.0051	
B2 Reflectance Split Value Threshold = 0.1765										

Table 9.5 Subsample B8 reflectances, sorted smallest to largest. Split from the last predictor remains

	Branch 1						Branch 2			
	Sample 1	Sample 2	Sample 6	Sample 5	Sample 9	Sample 8	Sample 3	Sample 7	Sample 10	Sample 4
B8 reflectance	0.211	0.233	0.266	0.27	0.277	0.287	0.217	0.228	0.36	0.389

Table 9.6: Split for Branch 2 of B8 reflectance.

Branch 2				
Sample No.	3	7	10	4
B8 Reflect.	0.2174	0.228	0.36	0.389
.0203	0.0055		0.0148	
Split Value Threshold = 0.294				

Table 9.7: Split Value Threshold for DEM Elevation

	Branch 2			
	Branch 2a		Branch 2b	
Sample Number	3	7	10	4
DEM Elev.	0.1045	0.2416	0.0383	0.0828
	Split Value Threshold = 0.173		**Split Value Threshold = 0.061**	

Table 9.8: Last split produces terminal or "leaf" nodes using the label, true DEM error.

Sample Number	3	7	10	4
DEM Error	0.0865	0.1147	0.0058	0.0041

In this example, B2, B8, and the DEM Elevation predictors were selected as the predictors and 10 random samples were taken, with replacement, from the Apalachee training data corpus (Table 9.2). The B2 reflectance was randomly chosen as the first predictor feature in the tree. Then the subsample is sorted from smallest to largest by B2 reflectance (Table 9.3). Table 9.4 shows potential split locations for the next branches. The variations are taken at each potential split and the split is chosen where the variation is minimized according to a variety of metrics including the Gini coefficient (Zheng, 2020), and in this example, the standard deviation of the values on each side of the split (Table 9.4) . However, an important parameter when generating an RF is the minimum number of values in a branch; for the purposes of this example this is set

to two except at the terminal or "leaf" node in a branch. The split with the lowest total standard deviation (left + right) is chosen.

The key result of this step is the split value threshold, defined as the average of the values on either side of the split. In this example, the split occurs between sample 6 and 4 because this is where the variation on either side of the split was minimized (Table 9.4). The average of those reflectances is 0.1765, so a point with B2 reflectance of less than or equal to 0.1765 will proceed down the left branch and a B2 reflectance of greater than 0.1765 proceeds down the right. Next, a new predictor feature is randomly chosen, B8 reflectance in this case, shown in Table 9.5. The same process is used to determine the split, noting that the sub-sampled features contained in each branch from the split above persist at this node. The reflectances in each branch are again sorted from smallest to largest (Table 9.5). To determine where the next split in each branch will be, the variations are quantified again (Table 9.6). This process is repeated for the next predictor, DEM Elevation (Table 9.7). The next split is the last and produces terminal or "leaf" nodes. The known target value or label is used for each subsample which completes this decision tree (Table 9.8).

The RF model was implemented in Python using the scikit-learn module (Pedregosa, Varoquaux, Gramfort, & Michel, 2011). The RF model hyperparameters were left at their default values with the exception of the number of trees being set to 300 (n_estimators parameter). Preliminary tests indicated that more than 300 trees offered no increase in prediction accuracy. The random state parameter was also set to an arbitrary number (59) for reproducibility purposes.

To summarize, 300 decision trees are trained as described above to construct the RF. Each tree makes a prediction of the DEM error for a given point and the predictions from all 300 trees

are averaged to produce one aggregate prediction for that set of feature data. To predict a point whose DEM Error is unknown, the point would be sent down the decision tree. For example, Table 9.9 and Figure 9.7 show the logic that would be followed while sending an unlabeled point down the tree. The point with a B2 reflectance of 0.2594, B8 reflectance of 0.2001, and a DEM Elevation of 0.1747 would have a predicted DEM Error of 0.1147 from this tree.

Table 9.9: Example of a random forest prediction of an unlabeled point.

Reflectance	Relationship to Split Threshold Value	Result
B2 = 0.2594	0.2594 > 0.1765	Go through Branch 2
B8 = 0.2001	0.2001 < 0.294	Go through Branch 3
DEM Elevation = 0.1747	0.1747 > 0.17305	*Predicted* DEM Error = **0.1147**

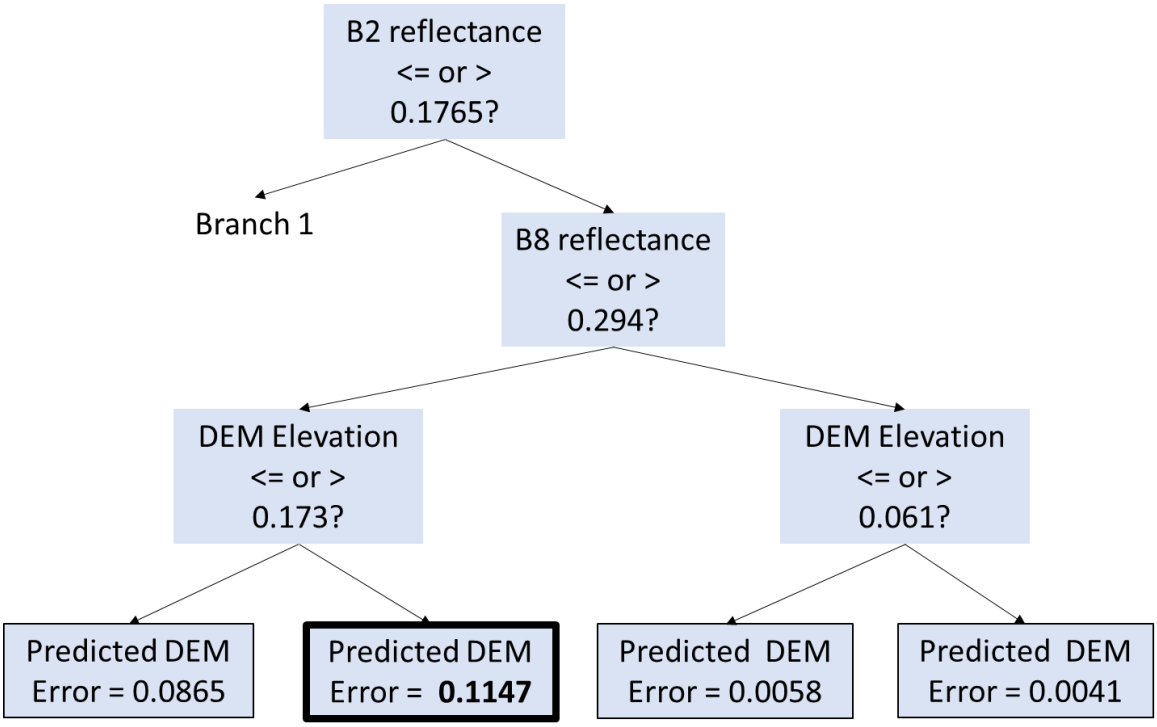

Figure 9.7: Decision tree trained from the given values from Table 9.3-Table 9.8. An unlabeled point sent down the decision tree with B2 ref. = 0.2594, B8 ref. = 0.2001, and DEM Elev. = 0.1747 would produce a predicted DEM error of 0.1147.

9.7.1 Model Validation

Due to the relatively small (n = 377 for St. Marks, n = 610 for Pascagoula) size of the field data corpus, RF model validation was executed using a bootstrapped or leave-one-out cross-validation protocol. One record in the training dataset was held out, and the model was trained on the remaining data. Then the trained model was used to predict the held-out value. This was repeated for all records in the training data to prevent over-fitting the model and maximizing the training dataset. The statistical metrics used included a 1 to 1 plot with coefficient of determination (R^2), root mean squared error (RMSE), normalized root mean square error (nRMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

For comparison, a linear regression (LR) model was also constructed and validated in the same manner on the same data to justify the use of the more complex RF model. Also, RF and LR models trained using only the DEM_Elevation field were developed in order to investigate the contribution of the satellite imagery to prediction accuracy. Statistical tests such as Akaike's Information Criteria (AIC) or Bayesian Information Criteria (BIC) were considered to differentiate between RF and LR, but it was determined that these tests are better suited to differentiating parameters in the same model not for comparing two different models (Burnham & Anderson, 2004).

10 Results

As described in the Methodology, a random forest model was used to predict the error in the lidar-derived DEM. The results of building the model and conducting the adjustment are discussed in the following sections.

10.1 Lidar Bias Mitigation – Apalachee Bay, FL

The results of the bootstrap cross-validation procedure for St. Marks are shown in Figure 10.1. Recall that this reflects the results of using 376 data points to train the model and testing it on the single held-out value for a total of 377 training / test cycles. The MAE is commonly used in model evaluation studies and in this case it represents the average amount of error in the DEM. The data were more scattered around the 1 to 1 line than expected with the RF model achieving an MAE = 0.054 m and the linear regression (LR) model achieving an MAE = 0.08 m. During the cross-validation procedure, the RF was the better performing model. The unadjusted DEM had an MAE of 0.177m. Therefore, the RF model adjustment improved the MAE of the DEM from 0.177 m to 0.054 m, improving the accuracy of the DEM by about 69%.

The satellite imagery used in this analysis, while publicly available, is not trivial to process therefore, its contribution to the prediction accuracy must be investigated. To accomplish this, we also ran the same training and test procedure using only the DEM_elevation field. Using only this field, the cross-validation procedure yielded an MAE value of 0.083 m for RF and 0.102 m for LR for the Apalachee region. Under these scenarios, the RF model improved the accuracy of the DEM by about 53% and the LR model by about 42%. Compared to an MAE of 0.054 m for RF with satellite imagery, the inclusion of the satellite imagery improved RF model performance. It is likely that the DEM_elevation field represents both the likelihood of inundation and a measure of bare ground visibility, as both are known to vary within the tidal elevation frame. Subsequently, the satellite imagery from wavelengths known to indicate the

variability in vegetation health (red and NIR) further predicts the interference of vegetation (Tahsin et al., 2021).

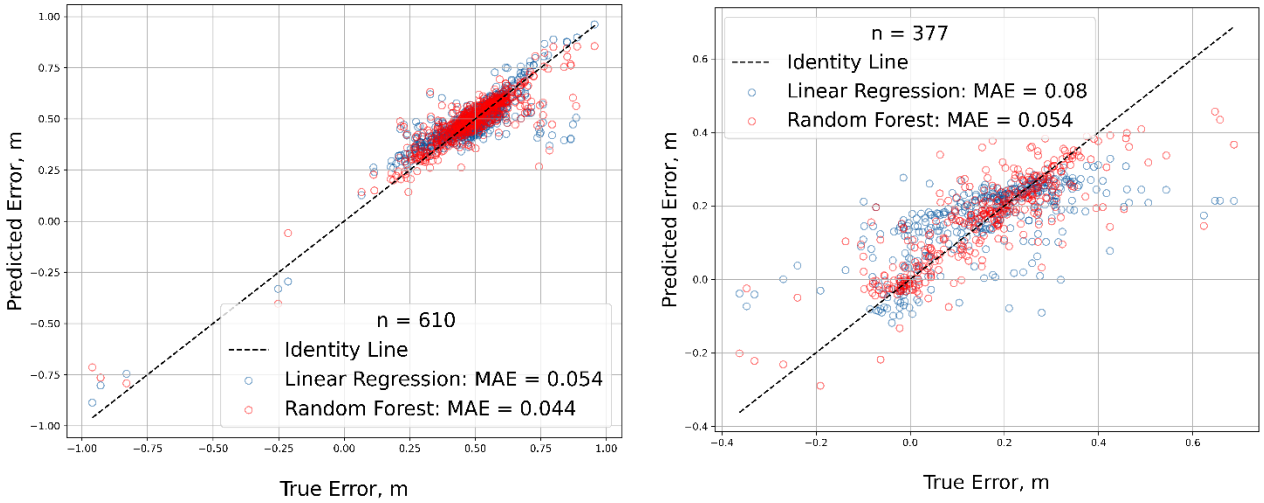

Figure 10.1: Leave-one-out validation plots of Pascagoula (left) and Apalachee Bay (right). The RF model improved the accuracy by about 69% for Apalachee Bay and about 90% for Pascagoula.

The trained RF model was applied to the lidar DEM for Apalachee Bay and the final adjusted DEM is shown in Figure 10.2. At this scale, the differences between the unadjusted and adjusted DEMs are difficult to see. The adjustment lowered the elevation by an average of 0.29 m. On a smaller scale differences are detectable. For example, Figure 10.3 shows the region where the field data were collected. The arrow is pointing to a manmade impoundment which appears to be relatively flat in the source DEM. However, the adjusted DEM recovers the underlying microtopography that exists in that area. The source DEM shows a higher elevation, with an MAE = 0.177 m. The adjusted DEM tends to lower the elevation which appears as the darker blue color in Figure 10.3. Figure 10.4 shows the actual adjustments that were made to the DEM. These subtraction values were calculated as the unadjusted lidar elevation minus the adjusted elevation.

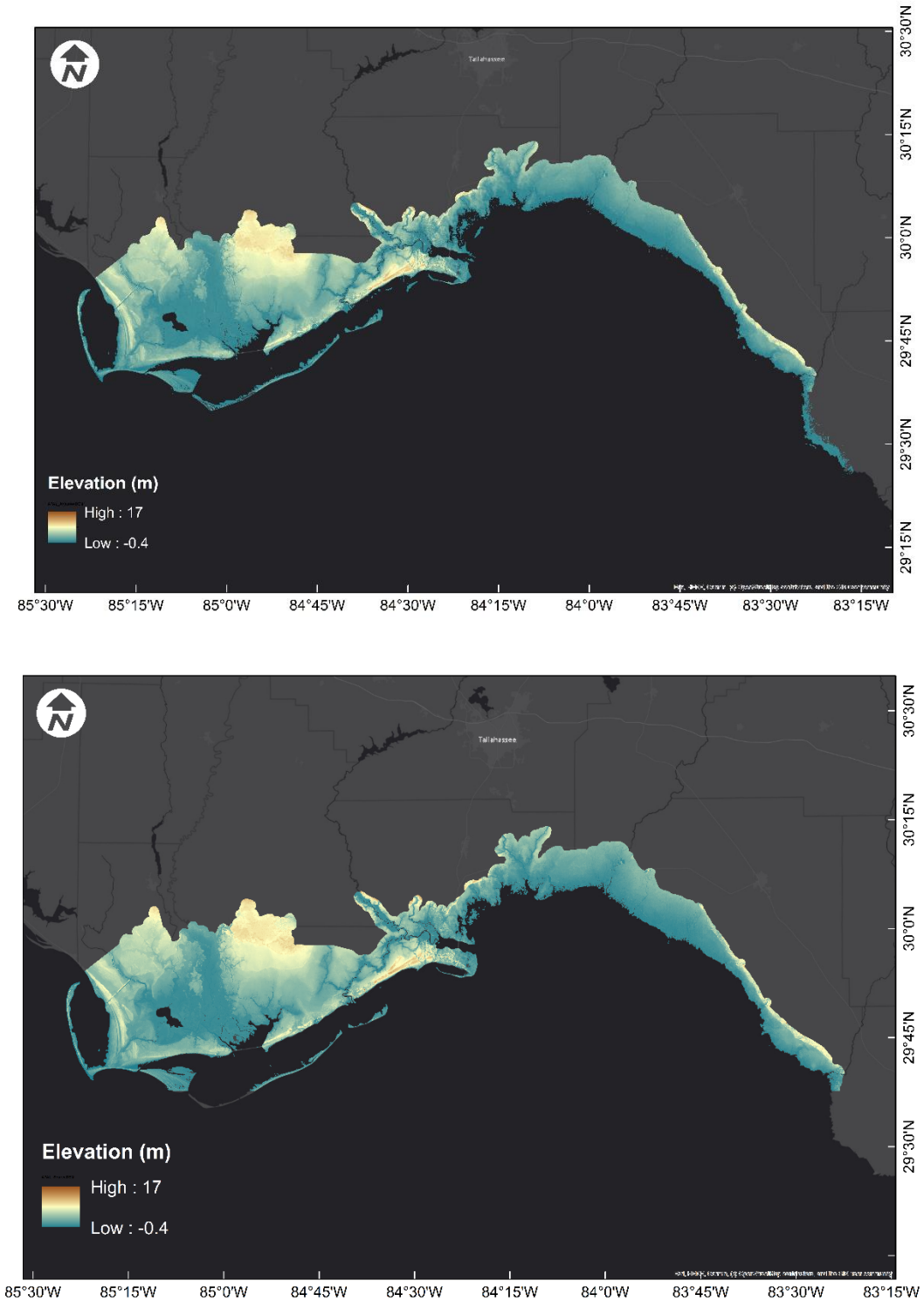

Figure 10.2: Comparison of the source (top) and adjusted (bottom) DEMs for Apalachee Bay

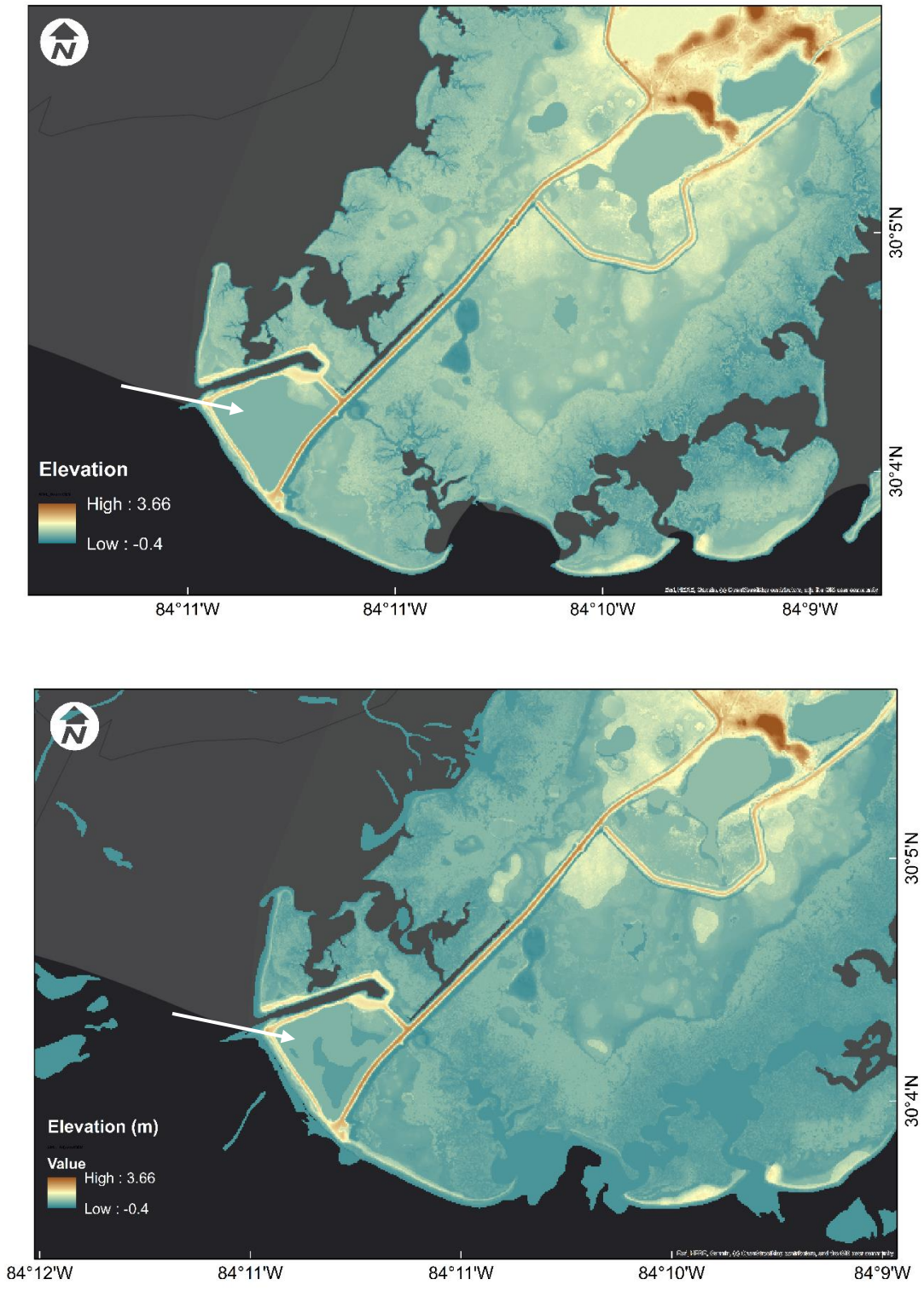

Figure 10.3: Comparison of the source (top) and adjusted (bottom) DEMs for Apalachee Bay zoomed in to the area where RTK spot elevations were taken. Notice the adjusted DEM shows underlying topography in the manmade impoundment which is called out by the white arrow.

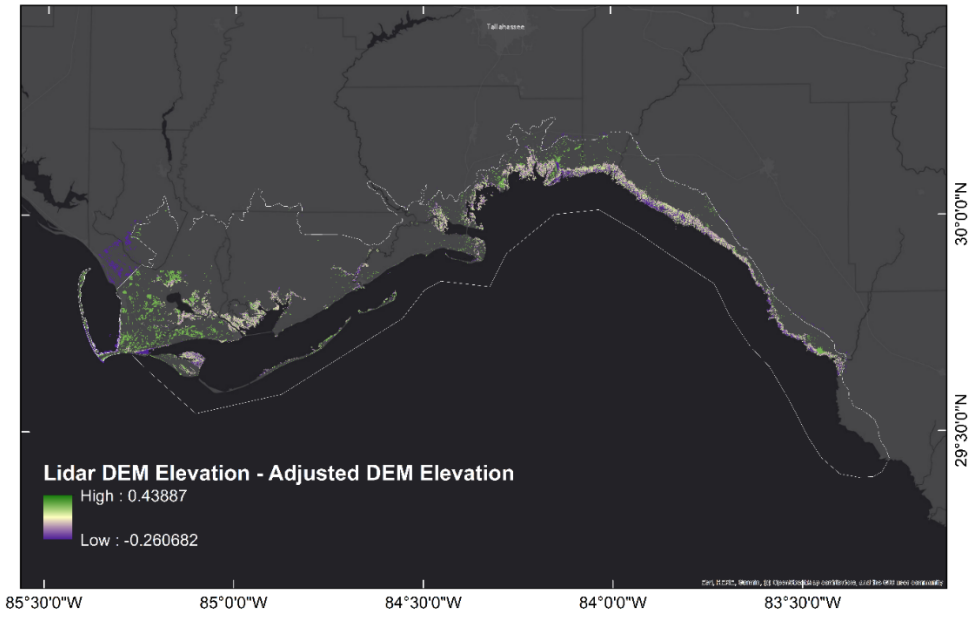

Figure 10.4: Map showing the adjustments made to the DEM calculated as lidar elevation minus adjusted elevation for Apalachee Bay. The purple area outside the boundary on the west side of the image contains a value of zero because it was outside the area of adjustment. The adjustment lowered the elevation by an average of 0.29 m.

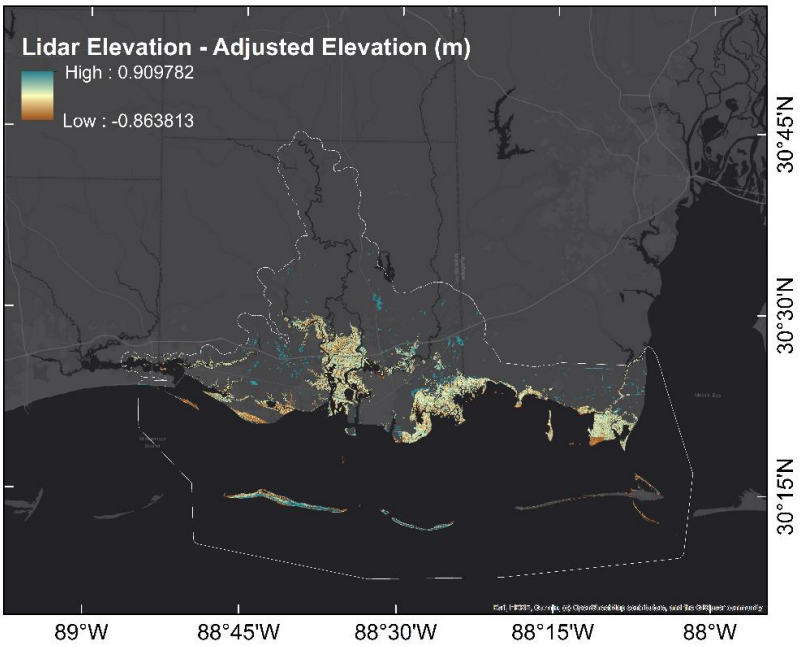

Figure 10.5: Map showing the adjustments made to the DEM calculated as lidar elevation minus adjusted elevation for Pascagoula. The adjustment lowered the elevation by an average of 0.56 m.

10.2 Lidar Bias Mitigation – Pascagoula Region

The results of the cross-validation procedure for Pascagoula are shown in Figure 10.1. Recall that this reflects the results of using 609 data points to train the model and testing it on the single held-out value for a total of 610 training / test cycles. This is 1.6 times the amount of data points that was used for the Apalachee Bay region.

The MAE for the source DEM was 0.493m. The RF model achieved an MAE = 0.044 m and the LR model achieved an MAE = 0.054 m. During the cross-validation procedure, the RF was the better performing model. Additionally, the RF model improved the accuracy of the DEM by about 90% and achieved a sub 5 cm MAE when used to predict the bias in the lidar DEM.

The trained RF model was applied to the lidar DEM of the Pascagoula domain. The adjusted DEM for the entire area is shown in Figure 10.6. As with Apalachee, the differences are difficult to see at this scale. The adjustment lowered the lidar DEM by an average of 0.56m. When zoomed in to the area east of Portersville Bay, the differences between the source and adjusted DEM are more apparent (Figure 10.7). The adjusted DEM in this area appears to be a darker shade of purple which corresponds with a lower elevation compared to the source DEM.

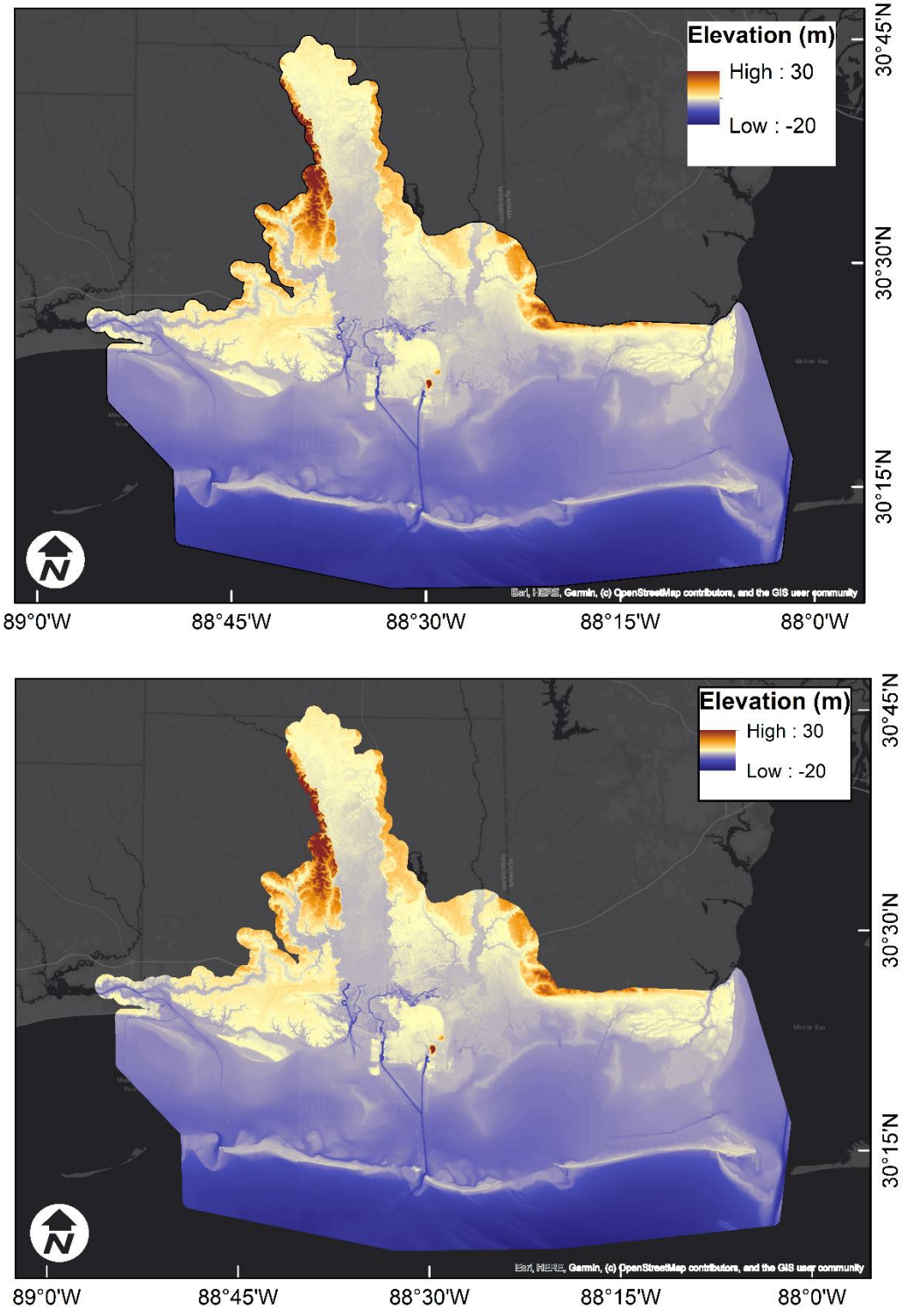

Figure 10.6: Source DEM (top) compared to the adjusted DEM (bottom) for the entire Pascagoula domain. Differences are difficult to see at this scale. Figure 10.7 shows the differences more clearly.

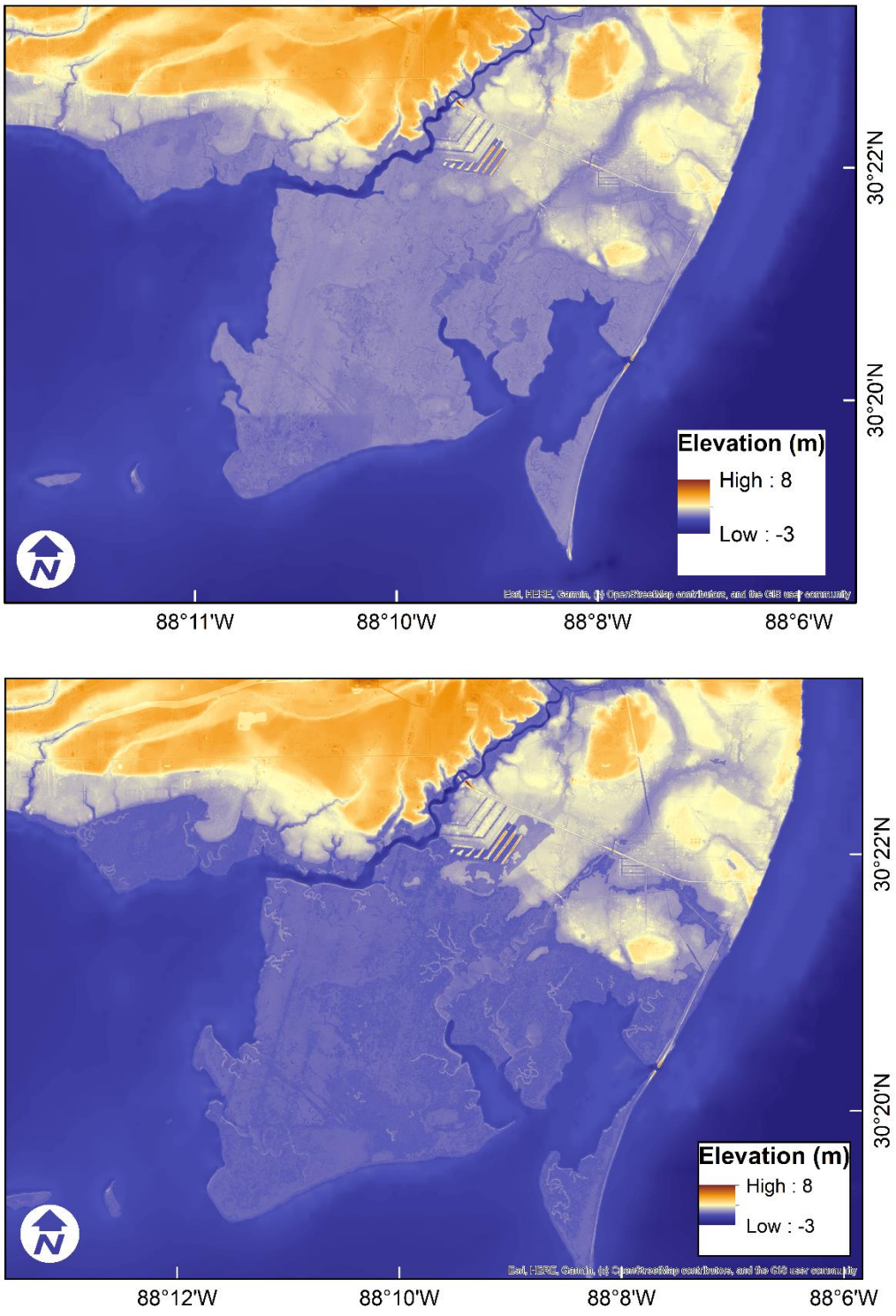

Figure 10.7: A comparison of the source (top) and adjusted (bottom) Pascagoula DEM zoomed in on the area east of Portersville Bay.

10.3 Nonlocal Validation Testing

It is expected that a lidar DEM from one location could be validated by an RF model trained on data from another location if the systems are ecologically similar. We tested this hypothesis by predicting the errors in Apalachee Bay using the Pascagoula RF model, and vice versa. Figure 10.8 shows that using a model on nonlocal data results in poor prediction accuracy because the R^2 values are negative. A negative R^2 value indicates that these predictions are worse than simply using the average of the point DEM errors as the prediction.

This graph also shows a shortcoming of machine learning models because the RF model cannot make predictions outside of the boundaries of the data with which it was trained. In Figure 10.8a, the predictions are clearly constrained between -0.2 and 0.5. The three outlier points clustered around -0.1 should have a prediction close to -1, but since the model was trained with APAL data, it did not have any points close to -1 to reference. The results show that the model relies on having data in the marsh within the adjustment domain despite being similar from an ecological perspective. The differences in the tidal hydrodynamics likely necessitate local data for each adjustment. Table 10.1 shows the hydrodynamics at each location. The hydrodynamics appear similar, but the differences relative to the tide range are significant.

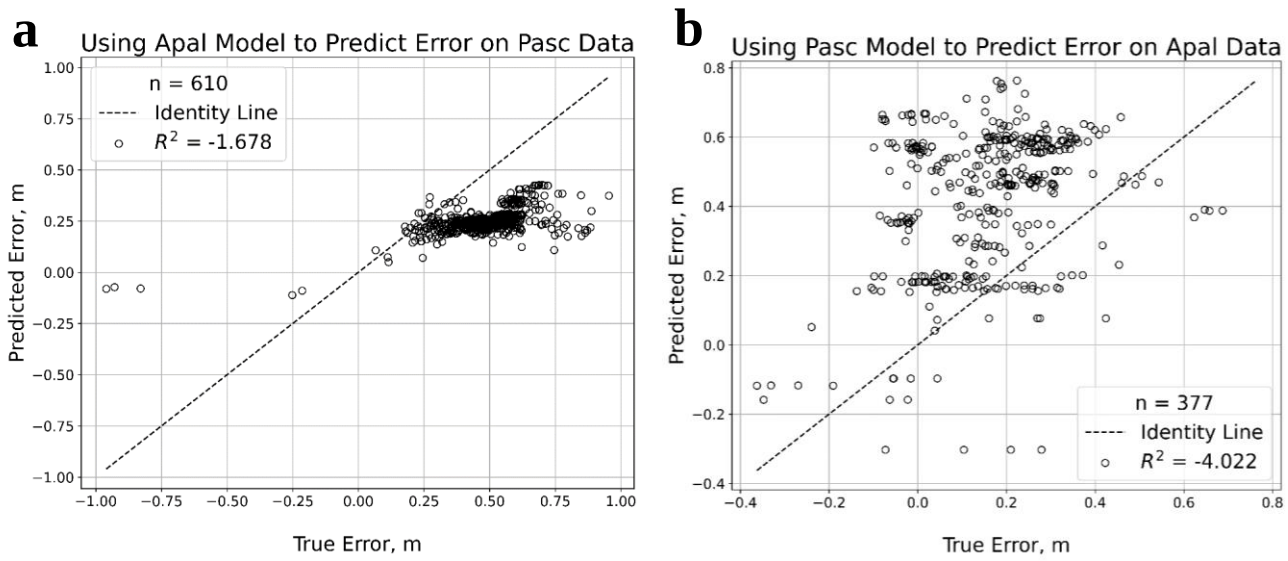

Figure 10.8: The DEM errors in the Pascagoula data were predicted using the Apalachee model (left) and the DEM errors in the Apalachee data were predicted using the Pascagoula model (right).

Table 10.1: Tidal datum elevations at each location (NAVD88)

	APAL (m)	PASC (m)
Mean High Water (MHW)	0.228	0.2339
Mean Seal Level (MSL)	0.045	0.031
Mean Low Water (MLW)	-0.110	-0.174

11 Discussions, Conclusions, and Recommendations

Using DEM in Apalachee with lidar data collected almost a decade apart creates an opportunity to analyze differences in lidar acquisition from 2007 and 2017/2018. Additionally, the most important predictor for the RF model was the DEM_elevation field. Furthermore, the nonlocal validation tests showed that local data is necessary to validate the model. Future work should include finding the optimum amount of training data needed for a model and a universal adjustment model should be explored as well.

11.1 Improvements in Lidar Acquisition Technology

To fully cover the Apalachee Bay region of interest, three data sources were mosaicked together. This was discussed in Section 9.2. Using data that was collected a decade apart (2007, 2017, 2018) created the opportunity to compare the magnitude of the error. The DEM data collected for the Apalachee Bay region in 2017/2018 is substantially better than the 2007 data in terms of the high elevation bias. The mean of the error distribution from 2007 DEM is about 0.65 compared to about 0.16 for the more recent DEM. This difference could be attributed to improvements in data collection technology as well as post-processing, specifically improvements in geometric calibration and radiometric correction. (Yan, Shaker, Habib, & Kersting, 2012) Additionally, the 2007 lidar DEM had five-meter resolution whereas the 2017 and 2018 DEMs were two meter and one-meter resolutions, respectively. Continued improvements in lidar acquisition and processing could eliminate the need for an adjustment. However, the current technology is not enough to meet the need for high accuracy digital elevation models (Alizad et al., 2020).

11.2 Most Important Predictor

Feature importance is a built-in metric for trained RF models. It allows us to compare how important each predictor is in producing predictions. Table 11.1 shows the feature

importances for each predictor of error for both research settings. The DEM_Elevation predictor had the highest relative feature importance for both locations, therefore it explains most of the variability in lidar DEM error. The elevation of the marsh platform controls inundation potential and bare ground visibility (J. Morris, 2007), while the RGB and NIR predictors likely capture vegetation vigor and subsequently its interference with the laser. Since the DEM elevation is the most important predictor of the error, we can infer that inundation potential and bare ground visibility are more influential than vegetation vigor in predicting the error for these locations.

Table 11.1: Feature importances for each model predictor and adjustment region.

Predictor	Feature Importance	
	Pascagoula	*Apalachee Bay*
DEM Elevation	0.6272	0.3725
B8 (NIR)	0.0999	0.1345
B4 (Red)	0.0896	0.1381
B3 (Green)	0.0983	0.2313
B2 (Blue)	0.0851	0.1236

11.3 Non-local Topographic Data

We reject the hypothesis that trained RF models are transferrable. Although the lower Pascagoula and Apalachee Bay are similar from an ecological perspective, the differences in the hydrodynamics require local data for each adjustment. The efficacy of a universal adjustment model trained on an expanded dataset across multiple marshes, especially those across a broad range of hydrodynamic conditions and ecologies remains to be seen.

11.4 NWI Inconsistency

A result we discovered when processing the geospatial data was an inconsistency in the NWI data for Apalachicola that appeared as unnatural discontinuities as seen in Figure 11.1. The NWI's

terminology for classifications changed in 2013, which caused the discontinuity in the data north of Apalachicola (Figure 11.1) (Federal Geographic Data Committee, 2013).

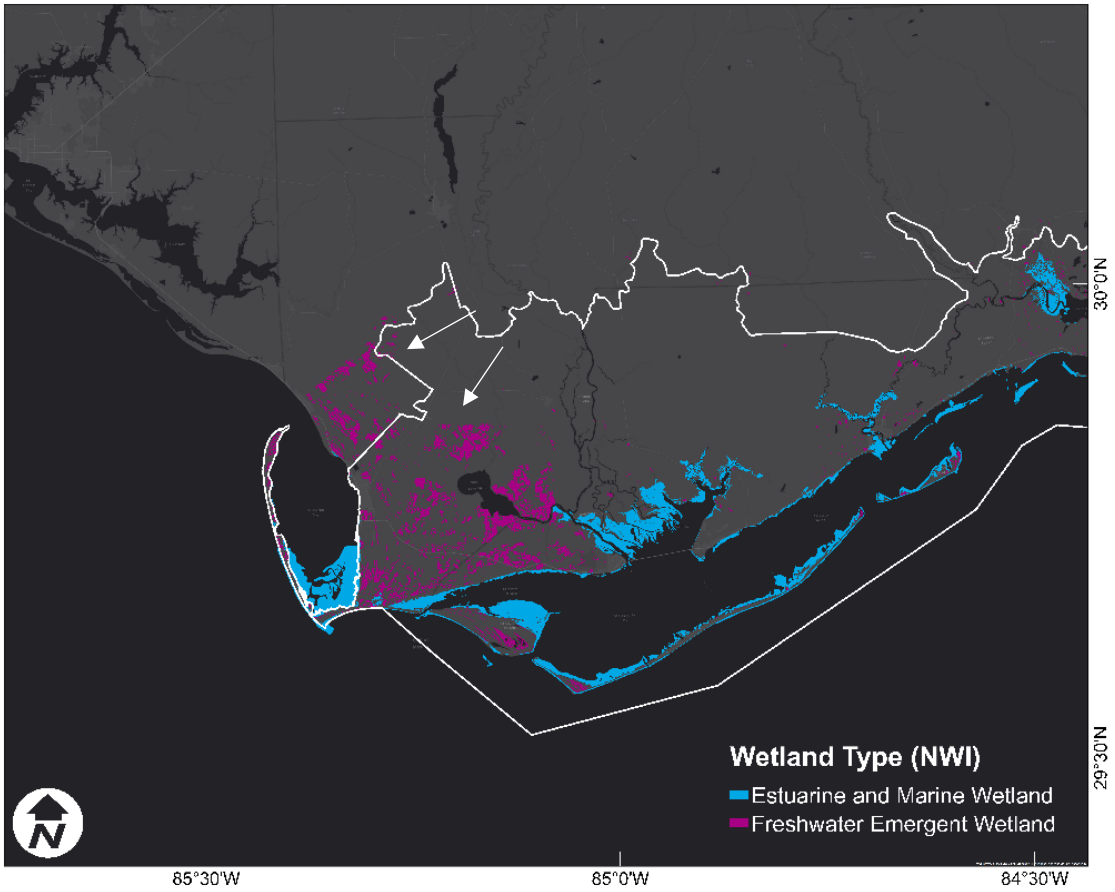

Figure 11.1: Inconsistency in the NWI mask for the Apalachee region illustrated by the unnatural breaks in the mask. This potentially caused some of the area of interest to be left out of the adjustment, which is outlined in white.

11.5 Conclusions

To accurately model a microtidal marsh system, the most representative DEM possible is preferred. A properly trained RF model can be used to reduce the error in the lidar-derived DEM In this study, the DEM was lowered by an average of 0.29 m for Apalachee and 0.56 m for

Pascagoula. The RF model improved the MAE of the DEM from 0.177 m to 0.054 m, improving the accuracy of the DEM by about 69% for Apalachee. For Pascagoula the results were even better, the RF model improved the MAE from 0.493m to 0.044m, improving the accuracy of the DEM by about 90%. DEM elevation is the most important parameter for predicting error in the DEM over the satellite imagery. However, the satellite imagery improves the performance of the model and should be included in the training dataset.

11.6 Recommendations

This approach can be adapted for other salt marshes, but local ground truth data are required to train and validate the model. In future work, a more consistent data source should be used to constrain the DEM adjustments to emergent wetlands in this region, such as the NOAA Coastal Change Analysis Program (C-CAP) Wetland Potential Layer. Other remotely sensed data sources could be explored with finer resolution such as Planetscope with RGB NIR at 3.7 m resolution or a UAS derived RGB NIR imagery with even finer resolution. Additionally, sediment elevation tables could be used as virtual ground truth points as opposed to collecting ground truth points by hand. A universal RF adjustment model that is trained on multiple datasets that represent a variety of hydrologic conditions could also be explored. As lidar acquisition and processing technology improves, the need for accurate elevation models will evolve as new demands arise.

12 REFERENCES

Alizad, K., Hagen, S., Medeiros, S., Bilskie, M., Morris, J., Balthis, L., & Buckel, C. (2018). Dynamic responses and implications to coastal wetlands and the surrounding regions under sea level rise. *PLoS One, 13*(10). doi:http://dx.doi.org/10.1371/journal.pone.0205176

Alizad, K., Hagen, S., Morris, J., Medeiros, S., Bilskie, M., & Weishampel, J. (2016). Coastal wetland response to sea-level rise in a fluvial estuarine system. *Earth's Future, 4*(11), 483-497. doi:10.1002/2016EF000385

Alizad, K., Medeiros, S., Foster-Martinez, M., & Hagen, S. (2020). Model Sensitivity to Topographic Uncertainty in Meso- and Microtidal Marshes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 807-814. doi:10.1109/JSTARS.2020.2973490

Baradaranshoraka, M., Pinelli, J.-P., Gurley, K., Peng, X., & Zhao, M. (2017). Hurricane Wind versus Storm Surge Damage in the Context of a Risk Prediction Model. *Journal of Structural Engineering, 143*(9), 04017103. doi:10.1061/(ASCE)ST.1943-541X.0001824

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24-31. doi:10.1016/j.isprsjprs.2016.01.011

Bilskie, M. V., Coggin, D., Hagen, S. C., & Medeiros, S. C. (2015). Terrain-driven unstructured mesh development through semi-automatic vertical feature extraction. *Advances in Water Resources, 86*, 102-118. doi:10.1016/j.advwatres.2015.09.020

Buffington, K. J., Dugger, B. D., Thorne, K. M., & Takekawa, J. Y. (2016). Statistical correction of lidar-derived digital elevation models with multispectral airborne imagery in tidal marshes. *Remote Sensing of Environment, 186*, 616-625. doi:https://doi.org/10.1016/j.rse.2016.09.020

Bunya, S., Dietrich, J. C., Westerink, J. J., Ebersole, B. A., Smith, J. M., Atkinson, J. H., . . . Roberts, H. J. (2010). A High-Resolution Coupled Riverine Flow, Tide, Wind, Wind Wave and Storm Surge Model for Southern Louisiana and Mississippi. Part I: Model Development and Validation. *Monthly Weather Review, 138*, 345-377.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference:Understanding AIC and BIC in Model Selection. *Sociological Methods & Research, 33*(2), 261-304. doi:10.1177/0049124104268644

Cooper, H. M., Zhang, C., Davis, S. E., & Troxler, T. G. (2019). Object-based correction of LiDAR DEMs using RTK-GPS data and machine learning modeling in the coastal Everglades. *Environmental Modelling & Software, 112*, 179-191. doi:10.1016/j.envsoft.2018.11.003

Dewberry. (2008). *2007 Florida Division of Emergency Management (FDEM) Lidar Project: Taylor County*.

Dewberry. (2018a). *2018 USGS Lidar: Florida Panhandle*.

Dewberry. (2018b). *FL Lower Choctawhatchee NWFWMD Lidar 2017 B17*. Retrieved from: https://coast.noaa.gov/htdata/lidar2_z/geoid18/data/8681/supplemental/NWFWMD_FL_Lower_Choctawhatchee_Topo_Lidar_Project_Report.pdf

Dietrich, J. C., Westerink, J. J., Kennedy, A. B., Smith, J. M., Jensen, R. E., Zijlema, M., . . . Cobell, Z. (2011). Hurricane Gustav (2008) waves and storm surge: Hindcast, synoptic

analysis and validation in southern Louisiana. *Monthly Weather Review, 139*, 2488-2522. doi:10.1175/2011MWR3611.1
Digital Coast. (2022). Data Access Viewer. https://coast.noaa.gov/dataviewer/#/lidar/search
Fassnacht, F. E., Latifi, H., & Hartig, F. (2018). Using synthetic data to evaluate the benefits of large field plots for forest biomass estimation with LiDAR. *Remote Sensing of Environment, 213*, 115-128. doi:10.1016/j.rse.2018.05.007
Federal Geographic Data Committee. (2013). *Classification of wetlands and deepwater habitats of the United States*. Retrieved from Washington, DC:
Fernandez-Nunez, M., Burningham, H., & Ojeda Zujar, J. J. J. o. C. C. (2017). Improving accuracy of LiDAR-derived digital terrain models for saltmarsh management. *21*(1), 209-222.
Hladik, C., & Alber, M. (2012). Accuracy assessment and correction of a LIDAR-derived salt marsh digital elevation model. *Remote Sensing of Environment, 121*, 224-235. doi:10.1016/j.rse.2012.01.018
Hu, T., Zhang, Y., Su, Y., Zheng, Y., Lin, G., & Guo, Q. (2020). Mapping the Global Mangrove Forest Aboveground Biomass Using Multisource Remote Sensing Data. *12*(10), 1690. doi:10.3390/rs12101690
Martin, S., Collini, R., Buckel, C., Passeri, D. L., Ahrabi-Nejad, S., & Medeiros, S. (2022). Adjusted Digital Elevation Models Improve Understanding of Gulf of Mexico Salt Marshes. Unpublished. In.
Mascaro, J., Asner, G. P., Knapp, D. E., Kennedy-Bowdoin, T., Martin, R. E., Anderson, C., . . . Chadwick, K. D. (2014). A tale of two "forests": random forest machine learning AIDS tropical forest carbon mapping. *PLoS One, 9*(1), e85993-e85993. doi:10.1371/journal.pone.0085993
McClure, A., Liu, X., Hines, E., & Ferner, M. C. (2015). Evaluation of Error Reduction Techniques on a LIDAR-Derived Salt Marsh Digital Elevation Model. *Journal of Coastal Research, 32*, 424 - 433.
Medeiros, S., Hagen, S., & Weishampel, J. (2015). A Random Forest Model Based on Lidar and Field Measurements for Parameterizing Surface Roughness in Coastal Modeling. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(4), 1582-1590. doi:10.1109/jstars.2015.2419817
Medeiros, S., Hagen, S., Weishampel, J., & Angelo, J. (2015). Adjusting Lidar-Derived Digital Terrain Models in Coastal Marshes Based on Estimated Aboveground Biomass Density. *7*(4), 3507-3525. doi:10.3390/rs70403507
Morris, J. (2007). Ecological engineering in intertidial saltmarshes. In (Vol. 577, pp. 161-168).
Morris, J., Sundareshwar, P. V., Nietch, C. T., Kjerfve, B., & Cahoon, D. R. (2002). Responses of Coastal Wetlands to Rising Sea Level. *Ecology, 83*(10), 2869-2877. doi:10.1890/0012-9658(2002)083[2869:ROCWTR]2.0.CO;2
Morris, J. T., Porter, D., Neet, M., Noble, P. A., Schmidt, L., Lapine, L. A., & Jensen, J. R. (2005). Integrating LIDAR elevation data, multi-spectral imagery and neural network modelling for marsh characterization. *International Journal of Remote Sensing, 26*(23), 5221-5234. doi:10.1080/01431160500219018
Pedregosa, F., Varoquaux, G., Gramfort, A., & Michel, V. (2011). Scikit-learn: Machine Learning in Python. *Journal of machine learning research : JMLR*. doi:10.5555/1953048.2078195

Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., & Ribeiro, L. (2014). Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Science of The Total Environment, 476-477*, 189-206. doi:doi:10.1016/j.scitotenv.2014.01.001

Rogers, J. N., Parrish, C. E., Ward, L. G., & Burdick, D. M. (2018). Improving salt marsh digital elevation model accuracy with full-waveform lidar and nonparametric predictive modeling. *Estuarine, Coastal and Shelf Science, 202*, 193-211. doi:10.1016/j.ecss.2017.11.034

Schmid, K. A., Hadley, B. C., & Wijekoon, N. (2011). Vertical Accuracy and Use of Topographic LIDAR Data in Coastal Marshes. *Journal of Coastal Research, 27*(6A), 116-132.

Singh, B., Sihag, P., & Singh, K. (2017). Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Modeling Earth Systems and Environment, 3*, 999-1004. doi:10.1007/s40808-017-0347-3

Tahsin, S., Medeiros, S. C., & Singh, A. (2021). Consistent Long-Term Monthly Coastal Wetland Vegetation Monitoring Using a Virtual Satellite Constellation. *Remote Sensing, 13*(3), 438.

Thales Alenia Space. (2021). *Sentinel-2 Products Specification Document* (S2-PDGS-TAS-DI-PSD). Retrieved from https://sentinel.esa.int/documents/247904/685211/Sentinel-2-Products-Specification-Document.pdf/fb1fc4dc-12ca-4674-8f78-b06efa871ab9?t=1616068001033

Thorne, K. M., Elliott-Fisk, D. L., Wylie, G. D., Perry, W. M., & Takekawa, J. Y. (2014). Importance of Biogeomorphic and Spatial Properties in Assessing a Tidal Salt Marsh Vulnerability to Sea-level Rise. *Estuaries and Coasts, 37*(4), 941-951. doi:10.1007/s12237-013-9725-x

Yan, W. Y., Shaker, A., Habib, A., & Kersting, A. P. (2012). Improving classification accuracy of airborne LiDAR intensity data by geometric calibration and radiometric correction. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 35-44. doi:10.1016/j.isprsjprs.2011.10.005

Zheng, C. (2020). A novel classification tree based on local minimum Gini index and attribute partial order structure diagram. *International Journal of Computer Applications in Technology, 64*(1), 33-45.

13 PUBLICATIONS

Medeiros, S.C, Bobinsky, J.S., Abdelwahab, K. (2022), Locality of topographic ground truth data for salt marsh lidar DEM bias mitigation, *IEEE JSTARS*, In Final Preparation

2 Data Sets (each adjusted DEM) Archived with NOAA National Center for Environmental Information (NCEI)