UNIVERSITY OF OKLAHOMA GRADUATE COLLEGE

REMOTE SENSING AND MACHINE LEARNING AS CONTEMPORARY INDIGENOUS KNOWELDGE AND CULTURAL PRAXIS: MAPPING SAGITTARIA LANCIFOLIA IN SOUTH FLORIDA AND THE CARIBBEAN

A THESIS SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By AIYESHA M. GHANI

Norman, Oklahoma

2025

REMOTE SENSING AND MACHINE LEARNING AS CONTEMPORARY INDIGENOUS KNOWELDGE AND CULTURAL PRAXIS: MAPPING SAGITTARIA LANCIFOLIA IN SOUTH FLORIDA AND THE CARIBBEAN

A THESIS APPROVED FOR THE

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SUSTAINABILITY

BY THE COMMITTEE CONSISTING OF

Dr. Nishan Bhattarai, Chair

Dr. Dawn Drake

Dr. Yanhua Xie

Dr. Xiangming Xiao

Dr. Adam Terando

Table of Contents

List of Tables and Figures

Acknowledgements

Indigeneity can be described as existing within deep, long-term relationality of care that has been cultivated by the bio geophysical nuances of a particular place over centuries. Many cosmologies around the world assert place as a function of relation. This project would not have been possible without the lands, waters, and winds of Borikén (Puerto Rico) and Kahayatle (South Florida Everglades) that deeply resonate within me as my two most core relations of ‘home’.

Often, bodies of knowledge that exist beyond Western globalized structures are dismissed as inferior – a longstanding consequence of what can been described as “parachute science”. Without the wisdom, care, guidance, and encouragement of my academic Advisor and Committee Chair, Dr. Nishan Bhattarai, this project would also not have been possible.

I would also like to acknowledge ally and mentor Dr. Adam Terando, for sharing with me opportunity to participate in such an important project –mapping the reproductive habitat of the iconic coqui frog- that will hopefully result in better protections against habitat loss for the coqui in Borikén.

A very big thank you goes out to Dr. Mark Wiltermuth, Dr. Jaime Collazo for their support, as well as to Dr. Neftali Rios-Lopez, Rafael Chaparro, and Thomas Clayton, for the good fun field times in waist-deep water.

A very special acknowledgement goes to my family and friends that supported me on this path; familia Ghani RiveraCandelaria, Rev. Houston Cypress and Love the Everglades Movement, for helping me collect field data in South Florida, and their overall support from my existence and endeavors.

I would also like to acknowledge and thank committee members, Dr. Dawn Drake, Dr. Xiangming Xiao, and Dr. Yanhua Xie for their patience, understanding, and feedback. Gratitude goes out as well as the University of Oklahoma - Department of Geography and Environmental Sustainability, and the College of Atmospheric and Geographic Sciences, and the Graduate College for facilitating my Master of Science degree.

Indigenous (First) Peoples have woven science into culture through textiles, art, language, and ritual/customs for tens of thousands of years. The purpose of this work is to continue the practice of weaving science and culture, through cultural resource mapping, guided by deep, long-term knowledge from direct relation -- and position Boricua-Arawak (Jibaro) culture amongst other indigenous cultures globally, while elevating indigenous research and placemaking within the contexts of modernity and scientific technology.

This work is dedicated to the song of the coqui, and the places they fill.

May we always hear them sing.

Funding for this project was made possible by the Southeast Climate Adaptation Center (SE-CASC), the Cooperative Ecosystems Studies Unit (CESU), the USGS Gap Analysis Program, The Oklahoma Biological Survey, and OU Office of Vice Presidents and Research Partnerships

Abstract

Mapping dominant vegetative species across coastal wetlands provides useful insights into traditional ecological and cultural assets, as well as the hydrological and biogeochemical responses to weather variability and shifts in local and regional climate regimes. These areas are often inaccessible given the vast remoteness, complexity, and extent of such landscapes. Remote sensing is considered a vital tool for accurately mapping vegetation communities in coastal wetlands, supporting indigenous traditional ecological knowledge and cultural preservation. The primary objective of this research was to use a random forest (RF) classification method on combined multi-source remote sensing and ancillary data to map an important wetland species, *Sagittaria lancifolia,* or 'bull-tongue arrowhead' in the coastal wetlands and Everglades of South Florida and the Northern Karst of Puerto Rico. Data sources include remotely sensed datasets from harmonized Landsat - Sentinel-2(HLS-S30/L30), Sentinel-1 synthetic aperture radar, NASA's Soil Moisture Active -Passive Level 4 Soil Moisture product, and ancillary datasets such as digital elevation models, soil water content, and soil porosity. Observational data were obtained from the Global Biodiversity Information Facility and field surveys in South Florida and Puerto Rico. The RF classification method was then applied to extracted pixel values of occurrence or non-occurrence of *S. lancifolia* and yielded an overall accuracy of ~87% accuracy (~65% Kappa statistic) in South Florida and ~95% accuracy (85% Kappa statistic) in Puerto Rico, when compared against test data not used in model training. The RF method produced slightly better accuracy with Landsat 8 data compared to the Sentinel-2 data. The RF-based variable importance analysis showed that elevation and soil

water content were the most important factors distinguishing the occurrence of *Sagittaria lancifolia* from other nearby vegetation species in these regions. The trained model was used to create maps of *Sagittaria lancifolia* across three study sites in South Florida and two study sites in Puerto Rico. Most classification errors resulted from commission errors, likely due to the variety of ecosystems that *S. lancifolia* occur in, and the surrounding vegetation that it may be found embedded within. Additionally, sampling data was not an exhaustive list of *Sagittaria lancifolia* occurrences in the study regions. Nonetheless, this research demonstrated the potential of using satellite data and machine learning techniques in mapping a target vegetation species of high cultural and ecological value, *S. lancifolia,* within wetlands. Future work will include a larger spatial extent and be enhanced by ground truthing (*in-situ)* fieldwork to collect observation data in Puerto Rico. Studies can use maps derived from this method to monitor the distribution of *S. lancifolia*, or other species of interest, which serve a critical role in blue carbon accounting for future economic development and environmental monitoring, as well as serving a role in cultural heritage preservation by better understanding the reproductive habitat distribution of the endangered coqui *Eleutherodactylus juanariveroii*.

Chapter 1: Introduction

1.1 The Significance of *Sagittaria* in South Florida and the Caribbean

Wetlands serve a critical role in hydrological and biogeochemical cycles globally. Coastal wetlands in the southeastern US and the US Caribbean are vital ecotones, as they support diverse ecosystems and offer crucial ecosystem services, such as flood protection, erosion control, carbon and methane sequestration, biodiversity, and habitat. While their importance and the services they provide are well documented in the scientific literature (Costanza, Robert, et al., 1997; Barbier, Edward B., et al., 2011; Duffy, J. E., 2006; Mitsch, William J., and James G. Gosselink, 2015; Pendleton, Linwood, et al., 2012; Day, J. W., et al., 2000; Temmerman, Stijn, et al., 2013; Kumar, Pushpam, 2010) their true potential and value remain likely underrepresented. Zhao et. al. (2022) observed connections between soil-air-water temperatures and found a significant buffering effect by wetland vegetation canopies. Other studies have shown connections between surface temperature modulation, wetland canopy, the atmospheric/planetary boundary layer, and reduced Bowen ratio (measure of the ratio between sensible and latent heat) which make wetlands critical in modulating climate and weather patterns (Taylor, 2018). Additionally, soils within seagrass meadows, mangrove swamps, and tidal salt marshes are rich in ‘blue carbon’.

‘Blue carbon’ was a term that was introduced in 2009 to convey the significance of carbon sequestration and long-term extreme weather mitigation (Nellman and Corcoran, 2009). According to Sheehy et al. 2024, blue carbon includes all forms of carbon in marine, intertidal, and estuarine environments including sedimentary stock and carbon stored in living biomass. Soils in coastal habitats, such as coastal wetlands, sequester blue carbon and can retain large carbon stocks as these soils often remain permanently saturated with water. Coastal soil and vegetation store and remove this atmospheric carbon at a much more accelerated and efficient rate than any other ecosystem on the planet. Soil organic carbon content is often considered a primary indicator of positive soil health (Murray, et al., 2011).

Sagittaria lancifolia is a salt- and flood-tolerant, emergent herbaceous aquatic plant native to the US Southeast and Caribbean of the *Almastacea* family and is currently designated as a species of high research priority. Commonly known as 'Duck Potato' or 'Bull-tongue Arrowhead', *S. lancifolia* is widely distributed along the Eastern and Gulf Coasts of the United States, as well as within some inland areas. Within the US Caribbean, *S. lancifolia is* the only known reproductive habitat for *Eleutherodactylus juanariveroii, an* endangered coqui species, also of high research priority, which is evidenced to be deeply embedded within the local culture since potentially prior to the Archaic age.

S. lancifolia holds high cultural, nutritional, historical, and ecological significance in the southeastern US and the US Caribbean, especially for indigenous cultural practitioners inhabiting those areas. The starchy underground tubers --'*wapato*' or 'duck potato' -- harvested from *Almastacea* (the larger family of *Sagittaria spp*.) can be consumed raw, boiled, dried, baked, roasted, mashed, ground into flour, or candied with maple sugar, and are considered a highly valued food source and medicinal plant for various Native North American people. In 2011, the Yakama tribe of the Pacific Northwest saw the return of this cultural foodway, with the restoration of wetlands that were previously drained and used for agriculture during the westward expansion of the United States in the 1800's-1930's (Washines & Peltier, 2010). Within the Caribbean, *S. lancifolia* is the reproductive habitat for *Eleutherodactylus juanariveroii,* a species of '*coqui'* – endemic amphibians integral to the indigenous cosmovision and identity of Borikén (Puerto Rico) that holds high cultural, historical, and social value among Puerto Ricans worldwide.

South Florida and portions of the Caribbean are mostly composed of karstic bedrock such as oolitic limestone, thus landscapes in this region are subject to significant landscape level changes directly driven by rainfall rates and total precipitation. Sea level rise and saltwater intrusion are two of the many increasing trends that threaten coastal wetland ecosystems, and adjacent freshwater wetland communities (Casanova, 2021). Although *Sagittaria lancifolia* is a freshwater species, it is surprisingly salt-tolerant, displaying steady growth, even under high salinity conditions (Martin and Shaffer, 2005). Brewer and Grace, et.al (1990) suggested that storm-generated changes in salinity, hydrologic regime, reproductive phenology, and interspecies competition all interact to influence community structure in oligohaline tidal marshes. *S. lancifolia* often survives drastic water level fluctuations by virtue of their underground nutrient reserves, and by changing their leaf morphology dependent on water depth (Wooten, 1986). *S. lancifolia* is also flood and halo- tolerant, in the freshwater plantain/rush family Almastacea which has implications for methane sequestration, phytoremediation, and marshland restoration. Planting *S. lancifolia* during freshwater marsh restoration may mitigate saltwater intrusion events because of their substantial below ground biomass, and flood- and salt tolerance (Martin and Shaffer, 2005).

Past studies have also demonstrated that *S. lancifolia* has the potential to reduce areal methane emissions, mitigate severe precipitation and flooding, adapt wetlands to future climates, as well as remove heavy metals and harmful anthropogenic chemicals, such as detergents, from water and soil (a process known as *phytoremediation*). Although *S. lancifolia* holds widely significant cultural, ecological, and economic value, it remains one of the least studied flora species, and to date. No specific studies on *Sagittaria lancifolia* distribution have been conducted in this region. Accounting and valuation of total blue carbon assets in Puerto Rico and South Florida have yet to be studied.

1.2 Scope

This research focuses on South Florida and the main island of Puerto Rico where *Sagittaria lancifolia* is widely distributed and is considered an important species in several contexts ranging from cultural to ecological functioning. Over the past several decades *S. lancifolia* has greatly expanded its distribution since earlier studies such as Shaffer *et al.,*1992 and Visser *et al.,* 1999, and Martin and Shaffer, 2005. Yet, its true distribution as of the current day, is not well understood. This is specifically due to the complexity of wetland ecosystems and the associated challenges such as access to remote wetland areas. Mapping current distributions of *S. lancifolia* provides insight into potential habitat ranges for the culturally significant 'coqui llanero*', Eleutherodactylus juanariveroii, as* well as a cost-effective way of monitoring and assessing the efficacy of publicly funded coastal restoration projects such as the Comprehensive Everglades Restoration Plan (CERP) in South Florida (SFL).

Peninsular Florida (PF) spans roughly 24-30 degrees North latitude and contains several types of landcover, with a considerable majority of coastal and freshwater wetlands, as well as an extensive network of freshwater springs throughout the karstic terrain and oolitic bedrock. The potential for year-round plant growth at low latitudes affects wetland energy balance differently during the wet and dry periods of the year, as has been observed in some tropical terrestrial ecosystems (Malhi et al., 2002; von Randow et al., 2004). The Florida Everglades is a large (>6000 km^2) subtropical wetland (Davis et al., 1994) that exerts considerable influence over local hydrology and climate (Duever et al., 1994; Light and Dineen, 1994). Similar to South Florida (SFL), the bedrock of the Northern Karst of Puerto Rico is composed of limestone that undergoes active diagenesis (constant erosion, compaction, and fracturing, largely driven by precipitation). As such, the hydrology of the region changes over time relative to precipitation and fluctuating climate regimes. The Northern Karst region of Puerto Rico extends from the Northwest portion of the island, and roughly eastward to the El Yunque/Luquillo area. The majority of *S. lancifolia* occurrences on record are contained within the Northern Karst Region of the main island of Puerto Rico.

1.3 Remote Sensing of Coastal Wetlands

Mapping dominant wetland vegetation can provide useful insights into changes in hydroperiods (quantity of precipitation), hydrology, atmospheric and surface chemistry (methane sequestration), as well as the radiative/thermodynamic components of areal weather. Remote sensing is an invaluable tool for mapping sensitive and remote areas such as the extent of coastal wetlands. Over the last few decades, coastal wetlands have faced degradation and decline due to both direct anthropogenic activities, such as agriculture and urban sprawl, as well as indirect impacts caused by varying spatial changes to atmospheric conditions, such as changes to temperature, precipitation, and hydrology due to climate variability (Barbier, 2019). Observations of areal methane show the quantity of areal methane to be reduced with marsh and wetland restoration efforts. Increasingly more countries have stated objectives in utilizing ‘blue carbon contributions, as part of climate mitigation plans, thus the effort and global economic need to map wetlands and the associated ecosystem services at finer scales have increased.

Remotely sensed data provides an efficient method for mapping and monitoring biodiversity and habitat change across broad spatial extents and with fine resolution (Pettorelli et al. 2016; Guo et al. 2017; Navarro et al. 2017, Koma et al. 2021).
Baker et al. (2006) and Nielsen et al. (2008) noted that most of the existing research regarding wetland mapping delineates and discriminates wetlands from other types of land cover, primarily using medium resolution imagery and differentiating between major plant communities such as grassland versus forested wetlands (Kindscher et al. 1997; Johansen et al. 2007; Maxa and Bolstad 2009; Bwangoy et al. 2010). Very few studies have attempted to discriminate among wetland grass communities, particularly communities of small extent (Barbosa 2010; Gilmore et al. 2008; Harvey and Hill 2001). Due to their composition and spatial heterogeneity and their limited spectral separability. Ozesmi and Bauer (2002) reviewed most of the literature on satellite remote sensing and found that the main problem in wetland remote sensing is the spectral separability of vegetation. Identification of wetland types at a medium spatial resolution is complicated, and due to their limited

temporal and spectral capabilities, optimal lighting and water-level conditions are difficult to capture. Hence, high-resolution imagery (aerial photography or satellite imagery such as WorldView-2) is preferred for more detailed and accurate wetland mapping (Santozi, Zoltan, et al., 2015).

Conversely, Baker et al. (2006) suggested that multispectral image classification of wetlands can yield maps with similar accuracy and is more repeatable relative to human photo interpretation. Studies have shown that classification algorithms, such as maximum likelihood (ML), support vector machine (SVM), decision trees (DTs), and artificial neural networks (ANNs) (Baker et al. 2006; Belluco et al. 2006; Sesnie et al. 2008; Szantoi et al. 2013; Wang et al. 2004; Wright and Gallant 2007) are promising alternatives to photo-interpretation for mapping wetlands (Santozi, Zoltan, et al., 2015). Ozesmi and Bauer (2002) also noted that in most cases, improved classifications were obtained when classification techniques such as ML, DT, or ANN were used with image texture and multispectral data incorporated in the analysis (Ghedira et al. 2000; Wang et al. 2004; Fuller 2005; Wright and Gallant 2007). This research uses a random forest (RF) approach on a combined dataset from multiple sources including harmonized NASA imagery (HLS-L30 and HLS-S30), NASA Soil Moisture Product 4.0, Sentinel 1- Synthetic Aperture Radar (SAR), as well as soil properties such as soil porosity and saturated water content, to map *S. lancifolia* in South Florida and Puerto Rico.

1.4 Hypothesis and Research Objectives

The primary goal of this project is to develop an accessible remote sensing methodology to map *S. lancifolia* across the Southeastern US and Caribbean. The central hypotheses of this research are:

i) Different satellite sensors, namely harmonized Landsat 8/9 and Sentinel-2 (HLS-L30/S30), exhibit distinct capabilities for identifying *S. lancifolia* remotely, thereby resulting in different classification accuracy in each study region, and

ii) Machine learning algorithms are a useful tool for mapping *S. lancifolia* in these regions.

The key objectives for this study are to:

i) Investigate the spectral differences between *Sagittaria lancifolia* occurrence points and non-occurrence points at fine and medium spatial scales (10m and 30m)

ii) Test the ability of a random forest algorithm to map occurrences of *Sagittaria lancifolia* using multi-sensor remotely sensed, ground verified, and ancillary data.

iii) Map target wetland vegetation species, *Sagittaria lancifolia,* across specific sites in Florida and Puerto Rico.

1.5 Organization of this Document

Chapter 1 briefly introduces the region of the US Southeast and Caribbean, describes the target species *Sagittaria lancifolia,* and identifies the need to remotely map *S. lancifolia*. Key analysis and results answering these three research questions are written in journal paper format in Chapter 2. The concluding summary and future recommendations are presented in Chapter 3.

Chapter 2. Detecting *Sagittaria lancifolia* Occurrence in South Florida and the Caribbean Using a Machine Learning Approach

ABSTRACT

Understanding vegetation dynamics is important across coastal wetlands like South Florida and the Caribbean that possess highly dynamic ecotones with complex interactions between terrestrial and marine ecosystems. *Sagittaria lancifolia* is an emergent herbaceous wetland rush considered native to the southeast US which has been designated a species of high-research priority due to the discovery of its use as a reproductive habitat by the endangered coqui llanero *(Eleutherodactylus juanariveroii).* Yet, to date, there has been limited work mapping this species of vegetation across South Florida and the US Caribbean because of limited occurrence data availability and inaccessibility of remote wetland regions. This paper utilizes the global database of the occurrence of the species, along with field observations, and globally available satellite and ancillary data to train and test a random forest (RF) classification algorithm to detect the occurrence of *S. lancifolia* at a high spatial resolution (~10-30m). Key datasets and predictors include the multispectral satellite imagery composites from the Harmonized Landsat Sentinel (HLS) datasets, Sentinel -1 synthetic aperture radar data, global soil moisture datasets, and United States Geologic Survey digital elevation model. The trained RF model was tested across sites that were not used in training, which includes all sampled points in two sites on the main island of Puerto Rico and three study sites in Florida. Overall, the occurrence model based on spectral properties and ancillary data was robust, with training and testing accuracy typically ranging between 85% - 95%. HLS-L30 (Landsat 8/9) data showed a slightly better ability to train a random forest classifier to

recognize the potential occurrence of *S. lancifolia* in Florida and was selected over HLS-S30 (Sentinel-2) to test and implement this classification model in Puerto Rico. The model trained in South Florida and applied in Puerto Rico returned an accuracy of roughly 83%, with Kappa values below 60%. Despite those results, the overall findings suggest good potential for using an RF-based algorithm on combined datasets to map *Sagittaria lancifolia*, and other specific wetland vegetation, across the southeastern US and Puerto Rico.

2.1 Introduction

The significance of coastal wetland ecosystems and the ecosystem services they provide are well-recognized in the scientific literature (Costanza, Robert, et al., 1997; Barbier, Edward B., et al., 2011; Duffy, J. E., 2006; Mitsch, William J., and James G. Gosselink, 2015; Pendleton, Linwood, et al., 2012; Day, J. W., et al., 2000; Temmerman, Stijn, et al., 2013; Kumar, Pushpam, 2010). However, the true value of these ecosystem services is likely underestimated due to the complexity of studying these wetland ecosystems. Wetland vegetation is specifically adapted to thrive in hydric soil and plays a vital role in distinguishing wetlands from other land cover or open water (Kokaly et al. 2003; Floyd and Anthony, 2008). One such case is the distribution of *Sagittaria lancifolia,* an herbaceous freshwater wetland species found prolifically worldwide but considered native within the southeastern US and the Caribbean. This specific wetland vegetation species is known to host egg clutches of *Eleutherodactylus juanariveroii*, an endangered amphibian species of high conservation priority discovered in 2007 (Rios- Lopez 2007) in Puerto Rico. Commonly known as 'Coquí llanero' or Plains Coqui, *Eleutherodactylus juanariveroii* is one of the 17 species of 'coquí', an endangered amphibian endemic to Puerto Rico. The conservation of these endangered species and their critical reproductive habitat, *S. lancifolia*, is an important consideration in not just preserving the cultural and ecological heritage of Puerto Rico, but also in designing adaptations for future climate regimes, mitigation strategies, and economic development via blue carbon stocks in these regions (Collazo, et al 2019). Mapping *Sagittaria lancifolia* can provide useful insights into the potential habitat and distribution of *Eleutherodactylus juanariveroii* and to the study of changes in the region's climate, yet few efforts have been made to map distributions *Sagittaria lancifolia* utilizing blended remote sensing methodologies.

Remotely sensed data provides an efficient method for mapping and monitoring biodiversity and habitat change across broad spatial extents and with fine resolution (Pettorelli et al., 2016; Guo et al., 2017; Navarro et al., 2017, Koma et al. 2021). The Florida Everglades and the Northern Karst of Puerto Rico cover some of the highly impacted, managed, and vast subtropical wetlands that greatly influence the hydrology, weather, and climate of each respective region. Baker et al. (2006) and Nielsen et al. (2008) noted that most of the existing research regarding wetland mapping delineates and discriminates wetlands from other types of land cover, primarily using medium resolution imagery and differentiating between major plant communities such as grassland versus forested wetlands (Kindscher et al. 1997; Johansen et al. 2007; Maxa and Bolstad 2009; Bwangoy et al. 2010). Very few studies have attempted to discriminate among wetland grass communities, particularly communities of small extent (Barbosa 2010; Gilmore et al. 2008; Harvey and Hill 2001), due to their composition and spatial heterogeneity and their limited spectral separability. Ozesmi and Bauer (2002) suggested that the main challenge in accurately identifying wetland species using remotely sensed data is the spectral separability of vegetation (Santozi, Zoltan, et al., 2015). Identification of wetland vegetation types using readily available, moderate spatial resolution data is often challenging due to the low temporal resolution of these sensors, and difficulty in timing for consistently capturing optimal lighting and water-level conditions that might contaminate the raw data. Hence, high-resolution imagery (aerial photography or satellite imagery) such as WorldView-2 is generally preferred for more detailed and accurate wetland mapping (Szantoi et al., 2013). However, most high-resolution datasets are not freely available.

Recent efforts to produce freely available harmonized datasets that combine multiple sensors, taking advantage of spatial, spectral, and temporal capabilities of multiple sensors, show promise for mapping coastal wetlands. Fan et al. 2023 used harmonized Landsat and Sentinel-2 imagery to map complex phenology of wetlands by manually harmonizing the data directly from Google Earth Engine. However, the National Aeronautics and Space Administration (NASA) has released 30-meter harmonized Landsat and Sentinel-2 data (HLS-S30 and HLS-L30) which provides higher temporal, spatial, and spectral resolution than previously available. Image texture is defined by the spatial arrangement and variation frequency of value levels of a contiguous set of pixels in a local neighborhood of an image. Textural data can improve the spectral separability of vegetation (Aytekin et al., 2013, Jin et al., 2014), and shorter wavelengths, such as C-band (~ 5.6 cm) and X-band (~3.1 cm), are preferred to discern non-forested wetland classes such as bog, fen, and marsh, from open water (Brisco et al., 2009).

Though there are several remotely sensed datasets available to map landscape level surface features including coastal wetlands like those in Florida and the US Caribbean, there is not a universally accepted classification algorithm for identifying specific vegetation occurrences or vegetation assemblages. Ozesmi and Bauer (2002) noted that in most cases, improved classifications were obtained when classification techniques such as machine learning, deep learning, and artificial neural networks using multispectral data are employed (Ghedira et al. 2000; Wang et al. 2004; Fuller 2005; Wright and Gallant 2007).

Sentinel-1 SAR (Synthetic Aperture Radar) is global, free of charge, and has a 5-6 days of revisit period. SAR can continuously and stably obtain earth observation data and information of vegetation structure and downward canopy, also be highly sensitive to water and soil moisture (Hui et.al, 2022). Mohammadimanesh et al. (2018) showed that the classification accuracy of RF could be improved by combining SAR intensity, interferometry, and polarimetry data. Backscatter coefficients of wetland woodlands and grasslands are significantly related to timing changes in water level. Typically, time series SAR data would be optimal for wetland classification and monitoring.

High resolution wetland studies at fine spatial scales in rice paddies and other wetland scenarios compare the efficacy of pixel based and object-based classification of imagery (Bolin et al. 2017). In the Bolin et al. (2017) study, the use of SAR and optical data resulted in 89.64% overall accuracy when using object-based classification and indicated the importance of textural data (SAR) to classification.

This study aims to test the utility of a random forest algorithm using pixel-based classification of multi-source remotely sensed and other ancillary data to differentiate *S. lancifolia* within its vegetative communities across Florida and Puerto Rico. The specific objectives are to identify which NASA HLS (Harmonized Landsat-Sentinel) dataset, namely HLS-S30 (Sentinel-2) and HLS-L30 (Landsat 8/9), is more effective in the task of mapping *S. lancifolia* occurring vegetation communities, given the diversity of canopy types where there are known occurrences, and which variables are most important to consider in developing a pixel-based classification methodology for high-resolution mapping of this species in subtropical and tropical regimes.

2.2 Methods and Materials

2.2.1 Study Area

A combination of ground-verified, known occurrence points and the limited record in the Global Biodiversity Information Facility (GBIF) database defined the study area and focus sites of this study (Figure 1). The study area includes South Florida and parts of Puerto Rico with a specific focus on five sites of known *S. lancifolia* occurrence, three in South Florida and two in Puerto Rico of approximately 5 kilometers. While the occurrence and non-occurrence points across all study areas were used to test the model accuracies, these sites were used to demonstrate the use of trained algorithms to create maps of *S. lancifolia* distribution. These sites were sampled in October 2023 with each sample area covering approximately 0.5 arc-seconds (15 meters)

and consisted of 4 to 8 georeferenced occurrence point samples. The sites in Florida include Pah-Hay-Okee Trail, Anhinga Wilderness Trail, and Hagen Ranch Road, all located in different watersheds within South Florida. The two experimental sites within the Northern Karst landscape of Puerto Rico include Sabana Seca and Caño Tiburonés (Figure 1). Distinctive in their dominant vegetation of Pterocarpus swamp, and spring fed, precipitation driven elephant grass prairies (respectively) Caño Tiburonés and Sabana Seca are located along the Northern Karst of the main island of Borikén.

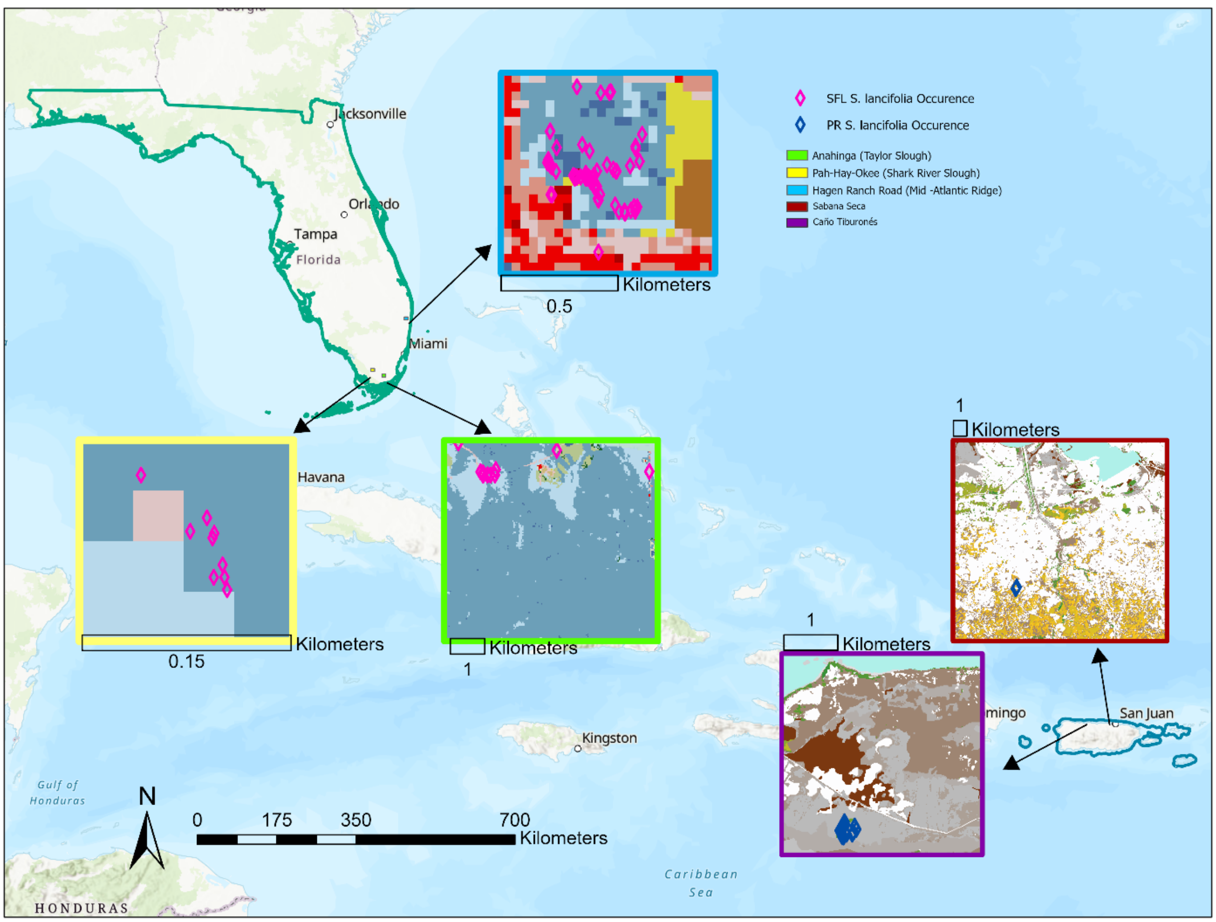

Figure 1. Map of *S. lancifolia* occurrences in Peninsular Florida (PF) and Puerto Rico (PR) study sites with sampled occurrences marked in pink and blue diamonds.

2.2.2 Data

2.2.2.1 *Sagittaria lancifolia* Occurrence Data

Occurrence data (presence, year, and location) for *S. lancifolia* across the state of Florida and the main island of Puerto Rico was obtained from the GBIF database (www.gbif.org). Established in 2001, GBIF is an open access collaborative data infrastructure and network funded by international governments. Information from this database is drawn from diverse sources including museum specimens, DNA barcodes, and georeferenced smartphone photograph observations. A polygon with latitudes ranging from 26.69213° to - 24.78502° and longitudes ranging from -82.11768° to -79.98663° in Florida, covering these occurrence points was used as a bounding box in Google Earth Engine (GEE). Similarly, in Puerto Rico, large polygons covering the known occurrences of *S. lancifolia* (at long-term field study sites, Caño Tiburonés, and Sabana Seca) in the Northern Karst regions were created. Occurrence point coordinates of *S. lancifolia* were collected at both sites in late October 2023 and aggregated to the existing occurrence point data. A total of 1,125 occurrence points in Florida, identified during the years 2022-2023 in the GBIF dataset, were used for training and testing the image classification algorithm (described later in section 2.2.3 Methods). For Puerto Rico, the sample size was relatively small with 80 occurrence points (20 from GBIF and 60 from field observations).

2.2.2.2 Key Remote Sensing and Ancillary Data

The remote sensing data for developing the classification algorithm and mapping *S. lancifolia* in the selected study sites primarily was derived from the NASA Harmonized Landsat Sentinel-2 surface reflectance data (Masek, et al., 2021). This includes multi-spectral data from Landsat 8 and 9– Operational Land Imager (OLI) harmonized (HLS-L30), and Sentinel 2 –Multi-spectral Imager (MSI) harmonized (HLS-S30). Other remotely sensed data include the C-band Sentinel-1 synthetic aperture radar (SAR) and products derived from remotely sensed data, such as the United States Geological Survey (USGS) National Land Cover Dataset and digital elevation model (DEMs) and NASA SMAP Level 4 Soil Moisture data product.

Table 1. Description of Data used in this study

Data Source	**Usage/Data**	**Variables Used**	**Reference**
NASA Harmonized Sentinel-Landsat (HLS-L30 /-S30- 30m)	Spectral, thermal bands	B2-B7 (surface reflectance) B8,8A- (Red Edge-NIR) B10 and 11(land surface temperature)	NASA/Google Earth Engine, Masek et al. (2021)
Sentinel -1 SAR (C-band)	Backscatter values, surface texture	AscDec Mean of 'vv', 'vh' dual and cross polarization	Li et al.(2013), Li and Chen (2005), Lang et al. (2008), Brisco et al. (2009)
NASA Soil Moisture Active and Passive Level - 4	Soil surface and root zone moisture	sm_Surface sm_Rootzone	Richle, et al. (2022)
HyHydroSoilv2.0	Soil porosity	'wcsat' saturated water content	Simons, et al. 2020
USGS DEM – 3DEP (10m)	Elevation	Elevation	usgs.gov
National Land Cover Dataset (2021)	Florida land cover designation	water, urban/impervious	Dewitz (2023)
Puerto Rico USFS 2020 landcover map	Puerto Rico land cover designations	water, urban/impervious	mrlc.gov
Global Biodiversity Information Facility	*Sagittaria lancifolia* historical occurrence, PFL and PR	occurrence	gbif.org

Sentinel 2 Multi-Spectral Instrument measures the Earth's reflected radiance in 13 spectral bands at 10m, 20m, and 60m spatial resolutions. Twelve of the Sentinel 2 MSI (10m) bands and eight (8) of Landsat 8 Operational Land Imager bands, which have a 15m panchromatic and 30m– multispectral spatial resolution, were used for the preliminary imagery analysis. The Harmonized Landsat Sentinel-2 (HLS) project provides consistent surface reflectance (SR) data by combining data from the Operational Land Imager (OLI), which includes the joint NASA/USGS Landsat 8 and Landsat 9 satellites, and Sentinel-2 Multi-Spectral Instrument (MSI) from Europe's Space Agency, which includes Sentinel-2A and Sentinel-2B satellites. The combined sensors enable global observations of the land every 2 to 3 days at 30m spatial resolution. The HLS project uses a set of algorithms to obtain seamless products from OLI and MSI that include atmospheric correction, cloud and cloud-shadow masking, spatial co-registration and common gridding, illumination and view angle normalization, and spectral bandpass adjustment.

Cloud masks using quality assessment bands were applied to the collections of 2022-2023 Landsat 8 and Sentinel-2 images to create cloud-free annual mean composites from each harmonized sensor. The NLCD 2021 dataset, which is primarily based on Landsat, has a spatial resolution of 30m. NLCD land cover maps were used to mask urban areas from the study areas. The NLCD 2021 dataset is not available for Puerto Rico, however the USFS and the Multi-Resolution Landcover Characteristic Consortium, which is approximately 5 years old, were used to filter prediction results. Landcover categories that were omitted to filter binary results in Puerto Rico were: hay and row crops (55), rocky cliffs (57), gravel beaches (58), artificial barrens (65), high-density urban (63), salt production (62), low-density urban (64), and saltwater (67). Landcover categories used to filter South Florida predictions were impervious surface and water, since none of the study sites in South Florida contained the same land cover types as sites in Puerto Rico.

Sentinel-1 SAR data was retrieved for the same time period to capture the responses of surfaces to changing moisture conditions. Sentinel-1 SAR variables include the average ascendant and descendant backscatter coefficients of cross-polarized and single polarized variables over the study period. The NASA SMAP (Soil Moisture Active and Passive) radiometer data is collected roughly every 3 hours and assimilated with the NASA Catchment Land Surface Model to generate the Level -4 data product used in this methodology. HLS- mean composite imagery was derived from a 2-year period (2022 and 2023) and values were extracted from each sample point of occurrence or non-occurrence from Google Earth Engine into '.csv' format and cleaned in *R*. The HiHydroSoil v2.0 Soil Hydraulic Properties dataset provides hydraulic properties of soils for hydrological modeling derived from the ISRIC's high resolution dataset, SoilGrids. The saturated water content/capacity ('wcsat') layer from the HiHydrosoil v2.0 global dataset was downloaded as a raster for the same time period (2022-2023), and pixel values were extracted in *R*. Incomplete data points within the South Florida dataset were omitted, while missing data for Puerto Rico were imputed using the 'mice' package in *R*, since the total number of points for SFL (n= >2000) was roughly 200% more than the number of points used for modeling in Puerto Rico (n = 138).

2.2.3. Methods

2.2.3.1 HLS S30 / L30 Spectral Reflectance Curves

Firstly, a collection of images of Peninsular Florida was compiled from the Sentinel-2 harmonized (S30) and Landsat 8/9 harmonized (L30) dataset (Claverie et.al 2018, Masek et.al 2021). Their quality assessment bands were used to mask out all the clouds and shadows to derive annual cloud-free surface reflectance composite (2022-2023) multi-band optical and thermal raster imagery. The surface reflectance (Both Sentinel-2 and Landsat 8) and land surface temperature (LST; Landsat 8 only) values for the occurrence point were extracted from these raster layers. Using a random point generator within ArcGIS Pro, a set of 2500 non-occurrence point features were generated to represent non-*Sagittaria* occurring points and were manually verified. Then, spectral reflectance data for each occurrence and non-occurrence point were compiled and extracted from Google Earth Engine. Spectral reflectance curves for occurrence and non-occurrence points, showing their spectral responses across different spectral regions of the electromagnetic spectrum covered by each sensor were derived.

2.2.3.2 Classifying *Sagittaria lancifolia* using Random Forest Algorithm at a point-pixel scale using 2022-2023 composite imagery

The 'randomForest ()' package in *R* (Breiman 2001; Liaw and Wiener 2002; Biau and Scornet 2016; Cutler et al. 2007) was used to classify the occurrence and non-occurrence of S. *lancifolia*. This Random Forest (RF) model was conceptualized referring to Leo Breiman (2001) and Adele Cutler (2007), following steps outlined in Wang (2024):

(i) Use N for the number of training samples and M for the number of features.

(ii) Choose the number of input features (m) for determining decisions at each tree node, where m is considerably less than M.

(iii) Employ bootstrapping by randomly sampling N times with replacement, creating a training set, and evaluating errors on the remaining unsampled examples.

(iv) Randomly select m features for each node and compute optimal splitting based on these features.

(v) Allow each tree to grow fully without pruning (Wang, 2024)

The random forest classifier is an ensemble classifier that can handle categorical, unbalanced, and missing data. This classifier also provides the relative importance of unique features during the classification process. Random forest classifiers consist of N trees, where N is the number of trees to be grown, which can be of any value defined by the user. To classify a new dataset, each case of the dataset is passed down to each of the N trees. The forest chooses a class having the most out of N votes, for that case (Pal, 2005). Furthermore, the random forest classifier provides a way to detect outliers by using proximity analysis between pairs which can be used for unsupervised learning; however most studies have not investigated the behavior of the random forest classifier by assessing the influence of training data quality/noise and variations in training data set size on classifier performance (Rodriguez-Galiano, et al 2012)

Some of the advantages of using the random forest classifier proposed by Breiman (2001) include efficient execution for large databases, high number of input variables without variable detection, estimate of significance of variables on classification, unbiased estimation of generalization error, identification of outliers, and it is computationally lighter than other tree ensemble classification methods (Rodriguez-Galiano, et al. 2012).Random forest classifiers utilize the Gini Index (Breiman et al, 1984) which measures the impurity of an attribute for any given class, as a selection measure. For any given training set T, selecting one case (pixel) at random and saying that it belongs to some class Ci, the Gini index can be written as:

$$\sum\sum_{j\neq i}(f(C_i, T)/|T|)(f(C_j, T)/|T|) \qquad (1)$$

where $f(C_i, T)/|T|$ is the probability that the selected case belongs to class C_i.

The coordinates of a total number of 1,125 occurrence points of *S. lancifolia* and an additional randomly generated 2,500 non-occurrence points were used in designing the RF classification model. These sites were randomly divided into 70/30 (calibration/validation) sites, where 70% of the data was used to train the random forest (RF) classification algorithm and the remaining 30% was used to test the accuracy trained model. A 10-tree random forest classification scheme was implemented on the composite imagery with additional SAR and saturated water content values (soil moisture data) (Figure 2).

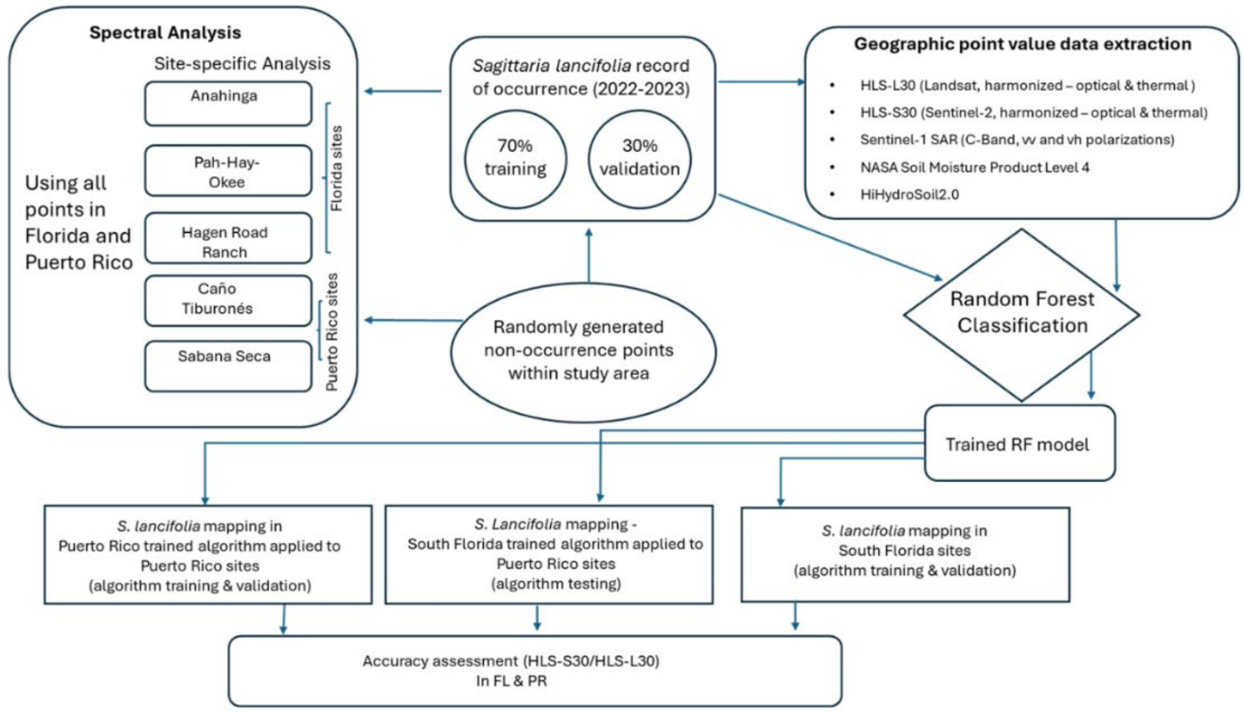

Figure 2. Workflow diagram for machine learning model focusing on detection of *S. lancifolia* using spectral analysis and a random forest classification.

2.2.3.3 Mapping *Sagittaria lancifolia* across Specific Study Sites Using Trained Random Forest Algorithm

Using the combined dataset created for this specific study, a Random Forest algorithm was applied to classify point pixels from HLS-S30 and HLS-L30 (see Figure 2). Since preliminary accuracy testing showed the HLS-L30 dataset to have a higher accuracy in Peninsular Florida, a model for mapping *S. lancifolia* was derived from each HLS-L30 optical, infrared and thermal bands combined with Sentinel 1- SAR, HiHydroSoil2.0, and NASA SMAP Level 4 data layers. The trained Random Forest model was applied to the layered data matrix to classify each point as an occurrence or non-occurrence. The resulting predictive layer of occurrence was further filtered by land cover classes by eliminating anything classified as ‘open water’ or ‘impervious surface’.

2. 3. Results and Discussion

2.3.1 Spectral Responses of *S Lancifolia* in Florida

Overall, *Sagittaria lancifolia* and nearby non- *S. lancifolia* vegetation species across the three South Florida study sites exhibited similar spectral patterns (reflectance approximately between (0.2 -0.4) with peak reflectance in the near-infrared region and lower reflectance in the visible and shortwave infrared (SWIR) regions (Figure 3). However, non-*S. lancifolia* vegetation species demonstrated consistently higher reflectance than *S. lancifolia* from both Landsat 8/9 and Sentinel-2 harmonized instrument retrievals. Peak reflectance for *S. lancifolia* occurrence and non-occurrence samples were detected by Landsat-8 around 900 nm wavelength, while Sentinel-2 detected peak reflectance closer to 950 nm wavelength. Specifically, reflectance from non-*Sagittaria* species in the shortwave and near infrared regions of the spectrum was noticeably higher. Landsat 8 showed a higher reflectance of non-*Sagittaria* points in the near infrared region, while Sentinel-2 showed a higher reflectance in the shortwave infrared region of the spectrum.

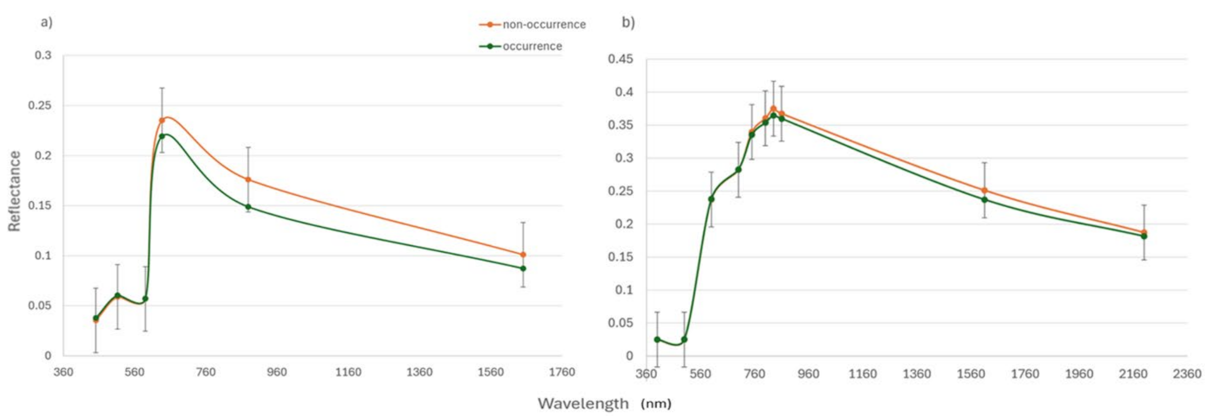

Figure 3. Spectral reflectance curves of *S. lancifolia* occurrence and non-occurrence in Peninsular Florida (PF) derived from band means from; a) HLS-L30 (Landsat 8/9), and b)

HLS-S30 (Sentinel-2). The error bars indicate the standard deviation of mean estimates from the samples.

The mean difference in reflectance from occurrence and non-occurrence points was significantly different for most bands (with t-stats ranging from -23.68 – 16.91 and *p-values* ranging from <0.0001 to 0.95 (Table 2). HLS-L30 Landsat optical Band 5, Band 6, Band 7, and thermal Bands 10 and 11 showed significant differences in values between occurrence and non-occurrence with t-stats ranging from -7.67 to 11.29 and *p-values* less than 0.0001.

Table 2. Results from Welch Two-Sample T-Test among all variables – South Florida with significant results denoted with *.

Variable	Mean Value (occurrence)	Mean Value (non-occurrence)	T-Statistic	P-Value
Elevation (Dem)*	8.043453	18.58988	-23.6794	4.16E-115
Saturated water capacity (Wcsat)*	5166.651	4894.072	16.91304	5.61E-59
B2_Landsat (HLS-L30)	0.036779	0.033997	2.911372	0.003635
B3_ Landsat (HLS-L30)	0.058927	0.056826	1.849794	0.064474
B4_ Landsat (HLS-L30)	0.055905	0.055119	0.596546	0.550868
B5_ Landsat (HLS-L30)*	0.217239	0.231868	-6.05453	1.64E-09
B6_ Landsat (HLS-L30)*	0.147109	0.17371	-10.9666	1.93E-27
B7_ Landsat (HLS-L30)*	0.086261	0.099519	-7.31936	3.24E-13
B10_L30_thermal band*	22.94725	22.33135	6.765311	1.61E-11
B11_ L30_thermal band*	21.75609	21.12836	7.852857	5.72E-15
B2_Sentinel (HLS-S30)*	0.049771	0.035398	9.896576	1.75E-22
B3_ Sentinel (HLS-S30)*	0.071761	0.057921	9.061457	3.23E-19
B4_ Sentinel (HLS-S30)*	0.069318	0.057288	7.320018	3.60E-13
B5_ Sentinel (HLS-S30)*	0.100077	0.090453	6.0404	1.82E-09
B6_ Sentinel (HLS-S30)	0.175146	0.171159	2.053071	0.040188
B7_ Sentinel (HLS-S30)	0.20364	0.201876	0.811288	0.417289
B8_ Sentinel (HLS-S30)	0.21024	0.209987	0.109618	0.912722
B8A_ Sentinel (HLS-S30)	0.227304	0.230284	-1.24919	0.211727
B11_ Sentinel (HLS-S30)*	0.149659	0.176779	-11.6214	1.44E-30
B12_ Sentinel (HLS-S30)*	0.08995	0.101607	-6.7802	1.45E-11
SMAP lvl-4 (sm_Rootzone)	0.175614	0.17325	0.310881	0.755922
SMAP lvl-4 (sm_Surface)	0.192023	0.191253	0.195522	0.845004
Sentinel1 - SAR - VH (vh_AscDec_mean)	-16.8772	-17.1465	2.538704	0.011186
Sentinel1 - SAR – VV* (vh_AscDec_mean)	-10.5281	-10.8684	3.49115	0.00049

Sentinel-2 (HLS-S30) Band 7, Band 8/8A, Band 11, and 12 also showed significant differences between occurrences and non-occurrences with t-stats ranging from 4.04 to 8.53 (*p-values* < 0.0001). SAR dual-polarization variable VV also indicated a significant difference between *S. lancifolia* occurrences and non-occurrence pixels with a t-statistic of 3.67 and a p-value of 0.0002. Saturated water content (soil porosity) and elevation appear to be the most significant factors, followed by HLS-L30 Band 6 (SWIR).

Similar reflectance patterns were observed across the three study sites in Florida, especially from the Landsat 8 sensor (Figure 4). However, there was a wide variation in reflectance patterns across the three sites in Florida. The Hagen Ranch Road site showed the highest reflectance from occurrence and non-occurrence in HLS-S30 (Sentinel-2), but the lowest reflectance for occurrence and non-occurrence from HLS-L30 (Landsat) out of all three South Florida study sites (see Figure 4). The Hagen Ranch Road site also showed the highest reflectance in *Sagittaria lancifolia* occurrence than non-occurrence among the three Southern sample sites from Sentinel-2 data. Additionally, occurrences and non-occurrences within the 0.5 arc-sec sample area at the Hagen Ranch Road site were indistinguishable using Landsat-8 and returned the same mean reflectance values. Contrarily, reflectance data from Pah-Hay-Okee Trail showed the highest reflectance for occurrence and non-occurrence from Landsat 8, but the lowest reflectance values for occurrence out of all sites from Sentinel-2.

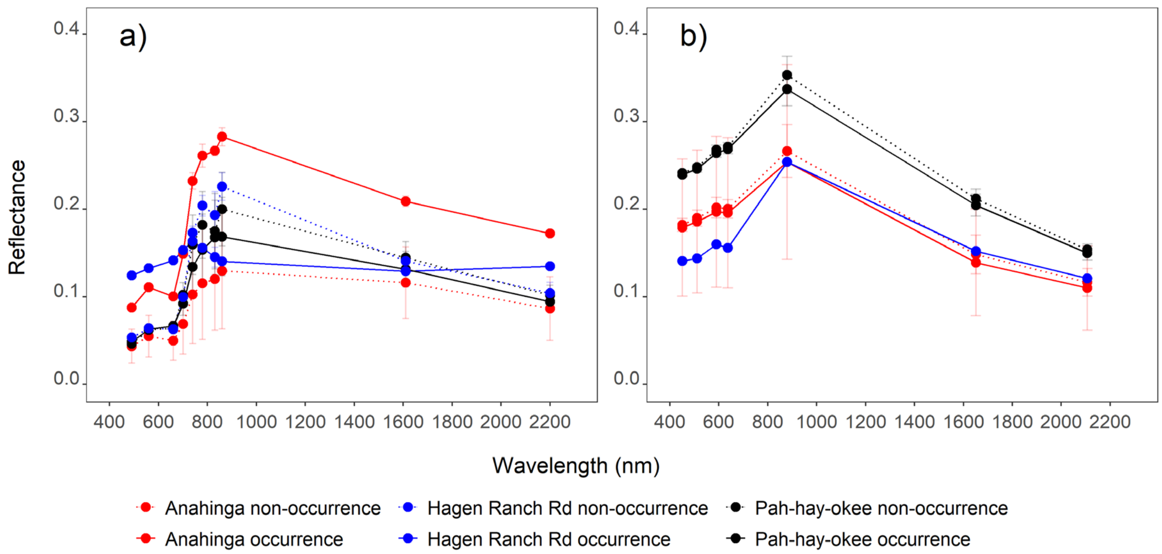

Figure 4. Spectral reflectance curve of *Sagittaria lancifolia* occurrence vs non-occurrence, a) Sentinel- 2 (HLS-S30), and b) Landsat-8/9 (HLS-L30) for South Florida study sites, Anahinga, Pah-Hay-Okee, and Hagen Ranch Road.

2.3.2 Spectral Responses of *S Lancifolia* in Puerto Rico

The spectral reflectance curves of *S. lancifolia* from the two study sites in Puerto Rico were similar to Peninsular Florida (Figure 5), as reflectance values from the non-occurrence points were consistently higher than the occurrence points. This was specifically when considering longer wavelengths on the sensor spectrum. The reflectance values from non-occurrence points varied more than those from the occurrence points, as indicated by the 95% confidence interval bars in Figure 5.

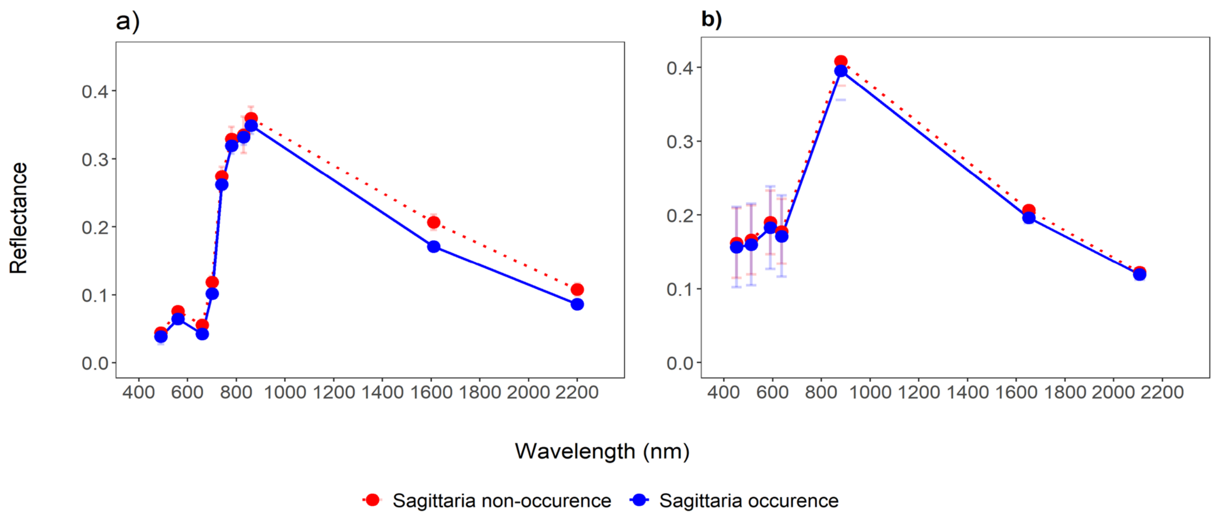

Figure 5. Mean spectral reflectance curve of *Sagittaria lancifolia* occurrence vs non-occurrence from study sites in Puerto Rico from NASA-HLS satellite sensors, a) HLS-S30 (Sentinel), and b) HLS-L30 (Landsat).

While the occurrence and non-occurrence of *S. lancifolia* in Puerto Rico follows a similar pattern to occurrence/non-occurrence in Florida, the differences in reflectance between the two sensors were less prominent. Significant spectral differences between occurrence and non-occurrence were primarily within HLS-S30 bands (see Table 3), and no significant difference was apparent within elevation or saturated water content of the soil at pixel-points. Sentinel HLS-S30 Band 6, Band 7, Band 8A, and Band 11 showed significant differences between *S. lancifolia* occurrence and non-occurrence pixels.

Table 3. Welch Two-Sample T-Test of Significance for All variables- Puerto Rico with significant variables noted *

Variable	Mean values (occurrence)	Mean values (non-occurrence)	T-Statistic	P-Value
Elevation (Dem)	15.57796639	2.167567869	1.861827209	0.067077109
Saturated water capacity	2570.164179	2572.464789	-0.026454299	0.978966372
B2_Landsat (HLS-L30)	0.021136474	0.034847418	-2.667363699	0.008697199
B3_ Landsat (HLS-L30)*	0.047672792	0.070652113	-4.004082644	0.000105542
B4_ Landsat (HLS-L30)	0.037451928	0.058208685	-3.320824355	0.001178607
B5_ Landsat (HLS-L30)*	0.284552756	0.383773944	-9.951963341	9.71E-17
B6_ Landsat (HLS-L30)*	0.137945321	0.214193662	-10.0685928	4.21E-18
B7_ Landsat (HLS-L30)*	0.063792444	0.107381455	-6.477023016	1.95E-09
B10_L30_thermal band	24.45252177	24.67178404	-1.067104686	0.288199905
B11_ L30_thermal band	23.11605721	23.1192723	-0.017696835	0.985911474
B2_Sentinel (HLS-S30)	2133.085114	2270.213913	-2.64372841	0.009494979
B3_ Sentinel (HLS-S30)	2217.313128	2365.784444	-2.991294279	0.003468041
B4_ Sentinel (HLS-S30)	2008.429087	2158.51849	-2.852453301	0.00520447
B5_ Sentinel (HLS-S30)	2602.920924	2765.110648	-3.278730566	0.001427207
B6_ Sentinel (HLS-S30)*	3640.869941	3902.0175	-5.309236757	6.71E-07
B7_ Sentinel (HLS-S30)*	4014.332494	4327.421126	-5.49563447	2.59E-07
B8_ Sentinel (HLS-S30)*	4034.002644	4290.193851	-4.489275763	1.75E-05
B8A_ Sentinel (HLS-S30)*	4172.737279	4500.436419	-5.545396645	2.18E-07
B11_ Sentinel (HLS-S30)*	2238.29856	2524.888625	-6.594320202	1.56E-09
B12_ Sentinel (HLS-S30)*	1495.545929	1694.88501	-4.806146755	4.62E-06
Sentinel1 - SAR - VH	-15.08815034	-14.86389382	-0.825889439	0.411129897
Sentinel1 - SAR - VV	-8.727790161	-9.02101922	1.009620557	0.315145952
SMAP lvl-4 (sm_Rootzone)*	0.281768632	0.318708622	-4.187934694	5.05E-05
SMAP lvl-4 (sm_Surface)*	0.250297348	0.283368149	-4.323785789	2.95E-05

Both HLS-S30 and HLS-L30 showed peak vegetation reflectance for both occurrences and non-occurrences at around 0.4, with greater deviations in the near infrared and shortwave infrared regions of the spectrum (Figure 6). It is notable that while both sites showed similar patterns, non-*S. lancifolia* points at Caño Tiburonés had noticeably higher reflectance in the near- and shortwave- infrared portions of the spectrum. Additionally, at these sites, there was a larger variation in reflectance from sampling points in HLS-L30 Landsat 8/9 datasets than in the HLS-S30 Sentinel-2 datasets.

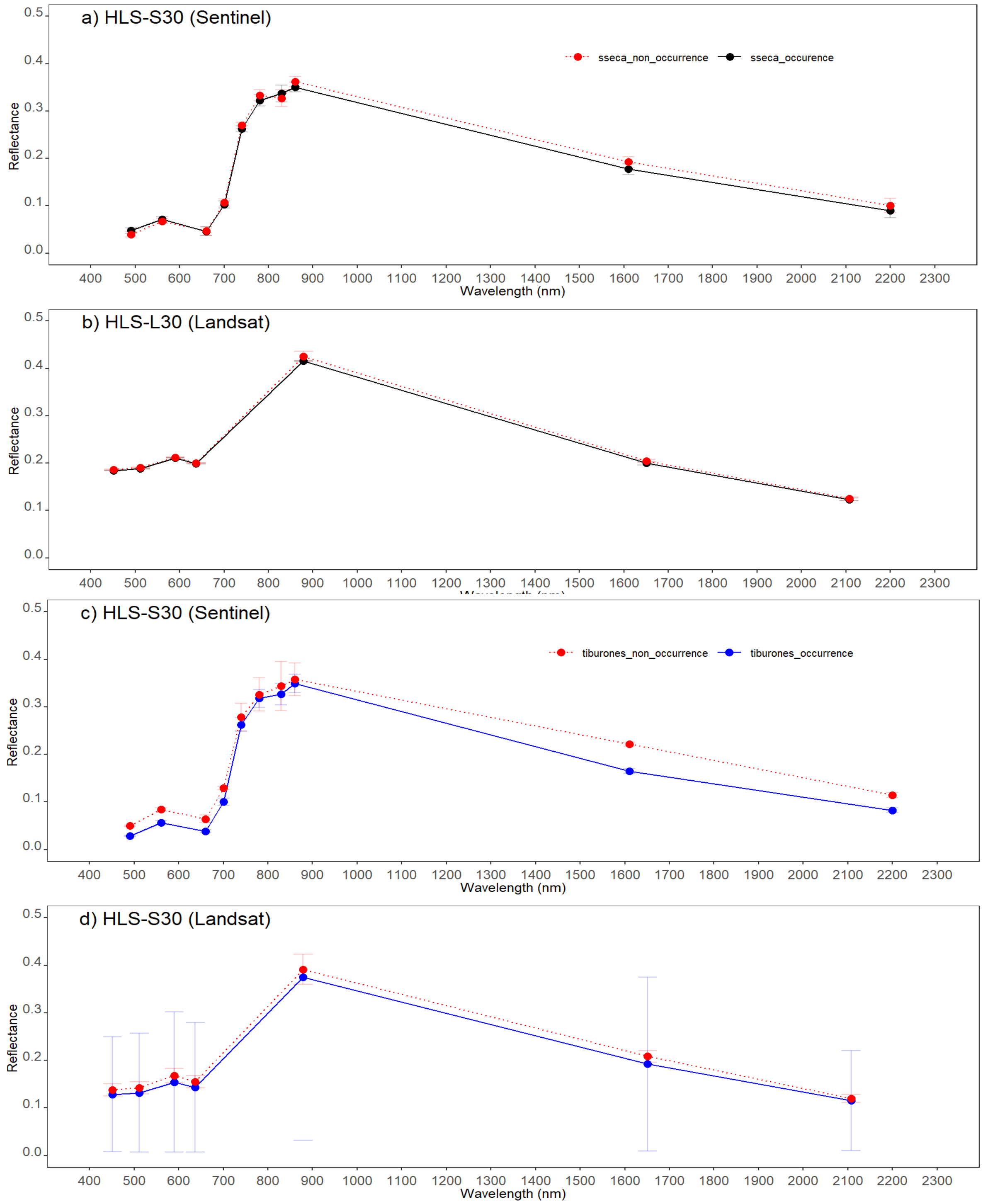

Figure 6. Mean spectral reflectance curves of *S. lancifolia* occurrence and non-occurrence from study sites in Puerto Rico from Sentinel-2 and Landsat 8 sensors.

2.3.3 Variance Importance plots

Considering the Random Forest approach to discrete vegetation mapping, understanding the impact that each variable has within the model is crucial. The variance importance plots in this section show the relative importance of variables in the RF classification. The Gini coefficient is a measure of inequality among values of a frequency distribution. Within the context of this study, the Gini index is used to measure the degree of difference among the band variables used in this Random Forest classification scheme of the HLS imagery and ancillary data. The Mean Decrease Gini (MDG) was plotted in R for all sensors in both regions, as well as specifically for model efficiency within the context of HLS-L30 (Landsat) and HLS-S30 (Sentinel-2) separately. The results of Mean Decrease Gini for each model are shown in Figures 7-11. In these figures, a higher 'mean decrease in Gini' MDG score indicates the higher importance of variables in the model. Across all data bands utilized in the random forest classification scheme, the most important variables were found to be elevation ('dem') and the saturated soil water content ('wcsat') with Mean Decrease Gini at > 80 for elevation and ~70 for soil porosity (Figure7).

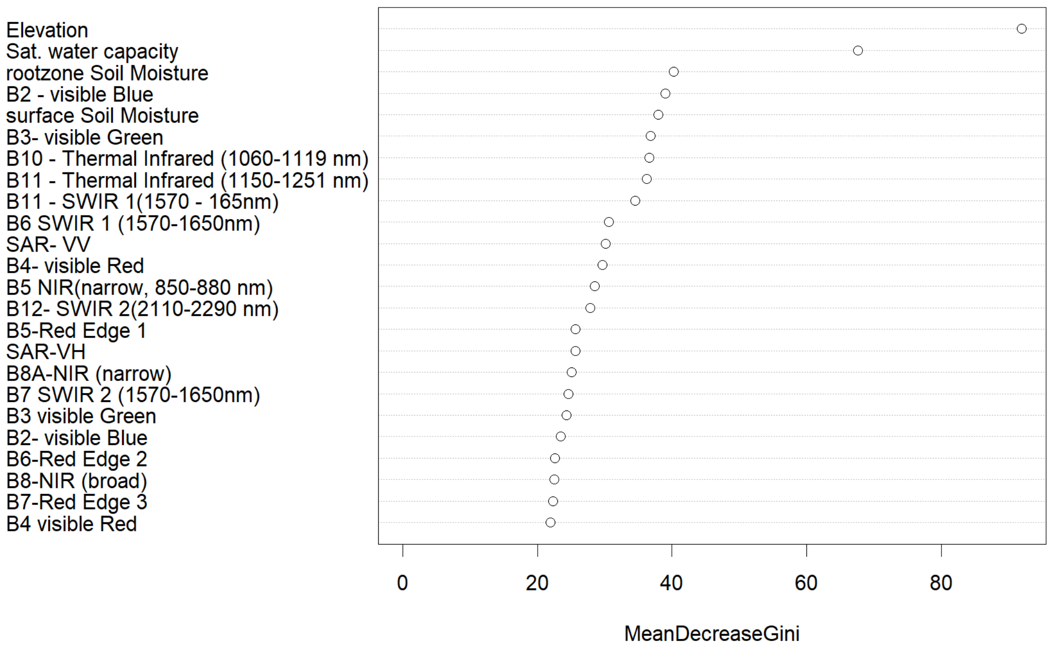

Figure 7. Variable importance plot from random forest classification for all data sensors and variables in South Florida.

Landsat 8- OLI thermal Band 10 was the third most important variable with a Mean Decrease Gini of > 40. Soil moisture in the root zone was shown to be slightly more important than soil moisture at the surface. Notably, the SAR dual-polarization variable ‘vv’ was considerably more important than the cross-polarization variable ‘vh’, although previous studies (Henderson, 2008) have listed the cross-polarization variable ‘vh’ as more accurate when used in detecting wetland vegetation. Optical red and green bands (B4 and 3) from Sentinel 2 MSI were of greater importance than the optical red band (B4) from Landsat 8 OLI.

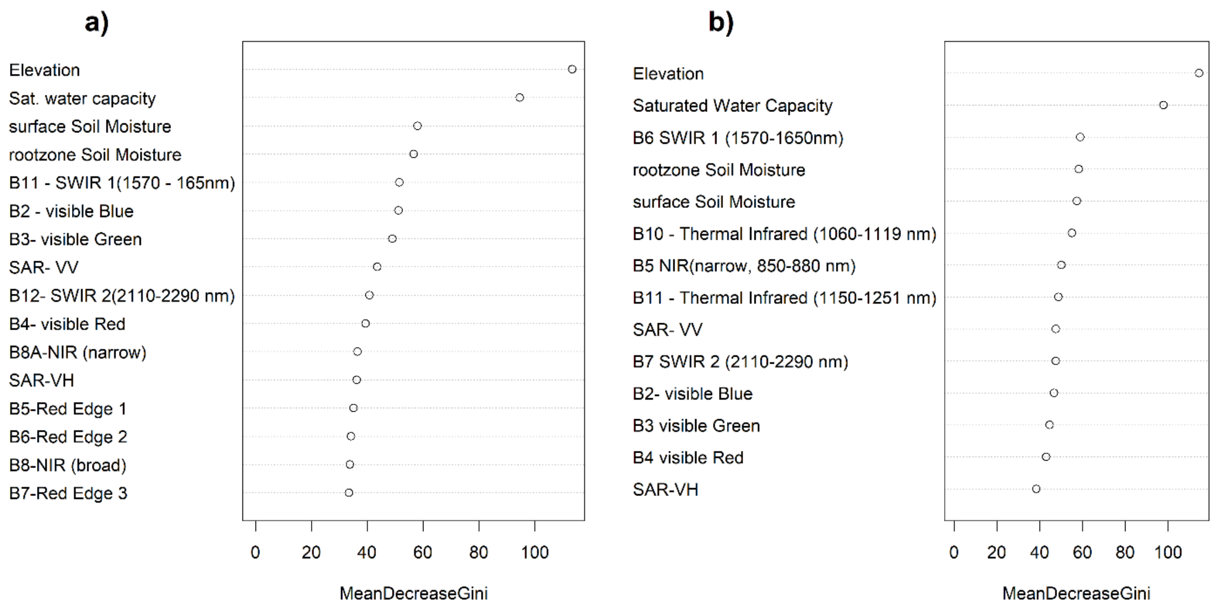

Figure 8. Variable importance plot from random forest classification for all soil moisture bands and a) HLS-S30 spectral and b) HLS-L30 spectral and thermal data, data in South Florida.

When considering the most important variables using HLS-L30 (Landsat 8/9) and HLS-S30 spectral bands in conjunction with additional elevation, saturated water content/soil porosity, and soil moisture data bands, elevation (>100 Mean Decrease Gini) and saturated water content/soil porosity (~100 Mean Decrease Gini) were, by far, the most important. Of secondary importance in this classification scheme are the shortwave infra-red (B6) bands, which showed comparable importance (Mean Decrease Gini between 60-70), and were the third and fourth most important variables in determining the occurrence of *S. lancifolia* in any given pixel (Figure 8).

Elevation was found to be the most important variable in differentiating the occurrence and non-occurrence of *S. lancifolia* in South Florida. On average, the species was found to be primarily distributed among zones of lower elevation, occurring at a mean elevation of 7 m, while the non-occurrence sites were mostly distributed in slightly higher elevations (mean elevation of 19m). The two sample t-test showed a significant difference in mean elevation from the occurrence and non-occurrence sites (t-stat = 16.913, p-value < 0.001, n = 100). After elevation, saturated water content ('wcsat'), which is equivalent to soil porosity, was the second most significant variable in classifying pixels according to the occurrence/non-occurrence parameters. *S. lancifolia* was also found to occur more in soil areas with higher saturated water content (soil porosity), where there are higher root zone soil moisture values, which was the second most important variable overall explaining the variation in occurrence and non-occurrence sites. Mean saturated water content in occurrence and non-occurrence sites were 0.52 m^3/m^3 and 0.49 m^3/m^3, respectively, and were significantly different ($p<0.001$ from the two-sample t-test).

Interestingly, in Puerto Rico, neither elevation nor saturated water content were as important as in South Florida. Instead, the HLS-L30 (Landsat) Band 6 - SWIR was shown to be the most important variable in the Puerto Rico trained RF model with a Mean Decrease Gini of close to 6, followed by HLS-L30(Landsat) Band 5- NIR. Additionally, despite the variable importance of similar bands (SWIR 1 & 2, NIR, and visible Blue) between both sensors HLS-L30 (Landsat) outperformed HLS-S30 (Sentinel-2) in the pixel-based RF classification, despite having a higher spatial, spectral, and temporal resolution. Elevation and saturated water content were of lower importance than all Sentinel-2 bands. Noteably, HLS-S30 Band 3 (visible Green), had a greater importance than HLS-L30 Band 3(visible Green). NASA surface moisture data ('sm_surface_mean', 'sm_rootzone_mean') were of high importance in classifying pixels over South Florida but were the least important random forest inputs for classifying *S. lancifolia* occurrence in pixels in Puerto Rico (Figure 9).

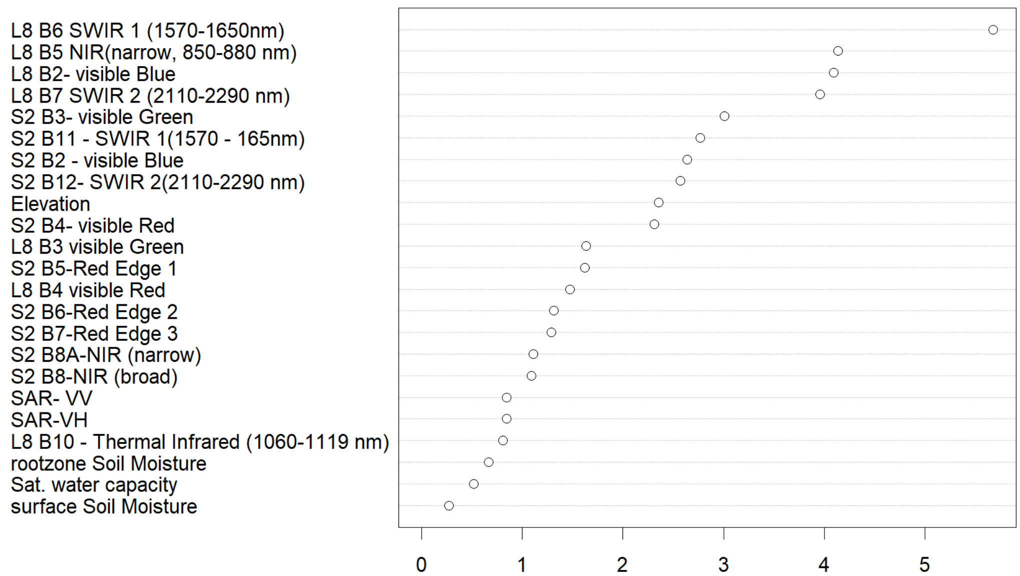

Figure 9. Variable importance plot from random forest classification for all soil moisture bands and HLS-S30 and HLS-L30 spectral data in Puerto Rico. HLS-L30 data inputs returned a greater Mean Decrease Gini than Sentinel, SAR, and SMAP variables by a wide margin, especially Band 6 (SWIR).

Upon running the RF classifier with only HLS-S30 (Sentinel-2) inputs in Puerto Rico, the inputs that returned the highest Mean Decrease Gini were the B11 (1570 – 1650 nm wavelength), by a large margin (Mean Decreased Gini > 6), and then the visible Blue band, followed by shortwave infrared B12 (2110 – 2290 nm wavelength), visible Green and Red bands, and elevation variables with Mean Decrease Gini values close to 5.0, respectively (Figure 10a). Overall, HLS-L30 returned greater Mean Decrease Gini values, indicating greater importance for distinguishing between the occurrence and non-occurrence of *S. lancifolia*.

When running the RF classifier in Puerto Rico with HLS-L30 (Landsat 8/9) spectral inputs, the most important variable was the SWIR1 and SWIR2, bands B6 (1570 – 1650 nm wavelength) and B7 (2110-2290 nm wavelength) which had a Mean Decrease Gini score over 8 (Figure 10 b). The subsequent next most important input variable was NIR (narrow) band B5(850- 880 nm wavelength) and B2- Visible Blue (450-510 nm wavelength) bands, returning a Mean Decrease Gini of 6 or greater. Elevation was shown to have a Mean Decrease Gini Value of 5, slightly less.

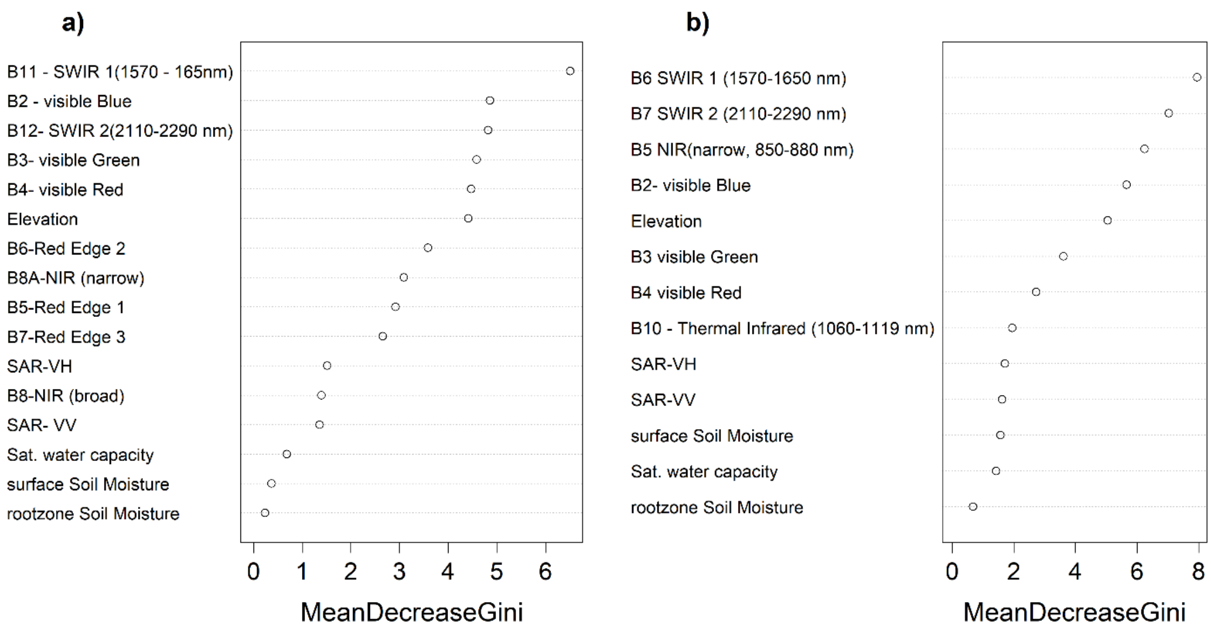

Figure 10. Variable importance plot from random forest classification for all soil moisture bands and a) HLS-S30 (Sentinel-2) spectral bands and b) HLS-L30 (Landsat) spectral and thermal bands in Puerto Rico.

Water content saturation (soil porosity) was found to have a value close to 1 but was the least important band for pixel-based classification in Puerto Rico, which was one of the most important inputs for the South Florida pixel-classification. This difference might be due to the homogeneity of the 'wcsat' data values among the sampled area in Puerto Rico (Table 3).

2.3.4 Random Forest Classification Accuracies

The Random Forest Classification using HLS-L30 which combines Landsat 8/9 data (combined with Sentinel 1-SAR, HiHydroSoil2.0 data, NASA SMAP level-4, and elevation variables) was found to be slightly more accurate than using the HLS-S30 (Sentinel-2) data in terms of classifying pixels for the occurrence of *S. lancifolia* in South Florida (Figure 11). While the overall accuracies were similar at 86-88%, the kappa accuracies were relatively lower at 60%-65% from both inputs.

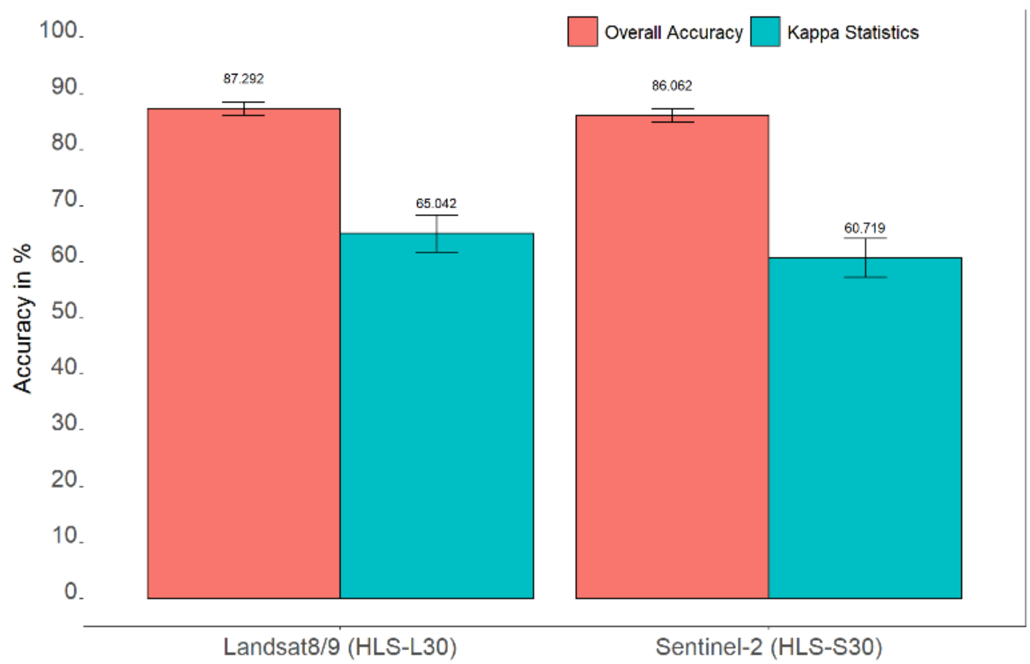

Figure11. Mean overall accuracy and Kappa statistic comparison between the performance of HLS-L30 data and HLS-S30 data from 100 iterations of random forest classification in South Florida. Error bars indicate the standard deviation of accuracy estimates from 100 random forest simulations.

Overall accuracy for 100 executions of the random forest classification using HLS-L30 and other ancillary data was 87% with a Kappa statistic of 65 %, slightly better performance than HLS-S30, indicating better applicability of HLS-L30 Landsat data for mapping *S. lancifolia* in South Florida. This is likely due to the presence of a thermal band (B10) in the HLS-L30 dataset. Notably, surface temperature was found to be the third most important variable in distinguishing *S. lancifolia* occurrence and non-occurrence points among all total variables in South Florida (Figure 7).

Accuracy from the RF classifier for pixels in Puerto Rico produced even better results. HLS-S30 (Sentinel-2) and HLS-L30 (Landsat) data performed similarly with an accuracy of 95% and Kappa statistics of over 85% (Figure12). This discrepancy is likely due to the higher number of ground-verified occurrence and non-occurrence points (n = 30) collected as field data in Puerto Rico, and the limited number of confirmed GBIF database occurrence points (less than 30) of *S. lancifolia* within the period of the imagery used for modeling and mapping (2022-2023). Since this might potentially indicate an overfitting of the model, 100 x simulations were run to test the robustness of the RF model, which was found to be pretty robust given the small variations in training and testing accuracy across all 100 simulations (Figure 12).

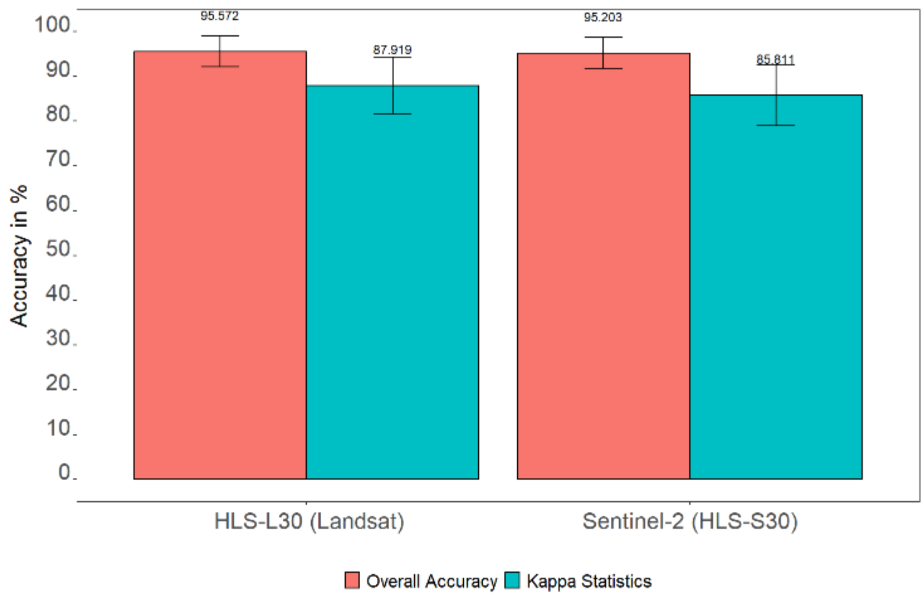

Figure 12. Accuracy and Kappa Statistic comparison between the performance of HLS-L30 data and HLS-S30 data in 100 iterations of random forest classification of a limited number of ground-verified *S. lancifolia* occurrence points in Puerto Rico. Error bars indicate the standard deviation of accuracy estimates from 100 random forest simulations.

Furthermore, the trained RF algorithm in South Florida was applied in Puerto Rico to test the replicability of the trained model in Florida, and its potential applications across Puerto Rico where occurrence datasets are limited. As expected, the accuracy of the model trained in South Florida and applied to Puerto Rico returned less robust results than models trained and tested within the same region. Nonetheless, the overall accuracy was still above 80% with Kappa statistics around 53-56% (Figure 13). These results indicate the potential applicability of the trained ML-based classification algorithm elsewhere in Puerto Rico.

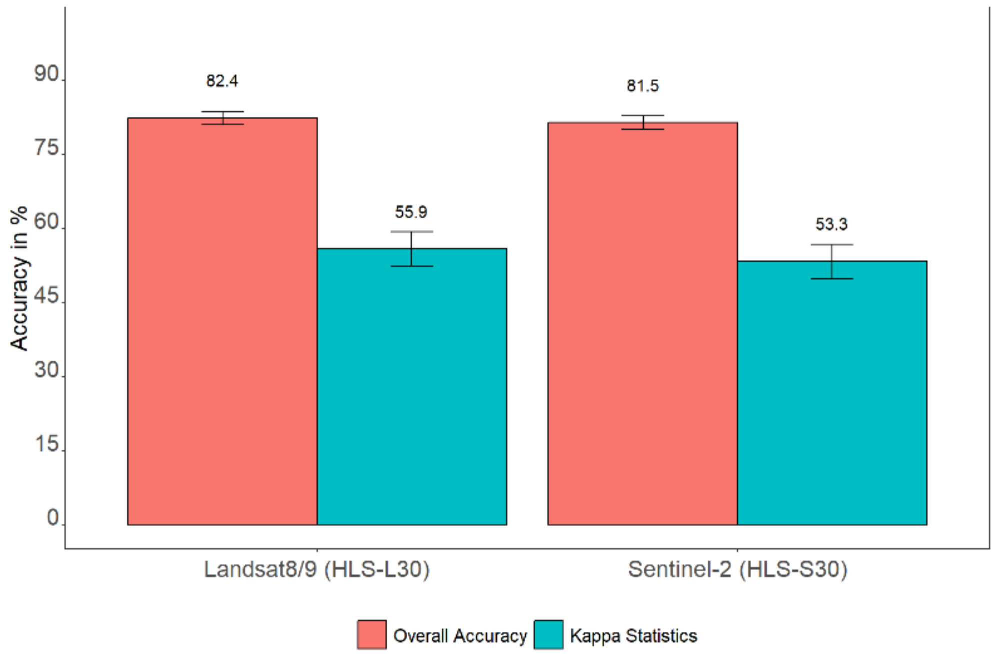

Figure 13. Accuracy and Kappa Statistic comparison between the performance of HLS-L30 data and HLS-S30 data in 100 iterations of random forest classification of *S. lancifolia* occurrence in South Florida, tested on points in Puerto Rico. Error bars indicate the standard deviation of accuracy estimates from 100 random forest simulations.

2.3.3. Mapping *S. lancifolia* Occurrence Prediction

For predictive mapping, raster layers in the exact extent of each study site were uploaded into R. Values of each pixel in each raster layer were extracted and converted into a Spatial Data Frame. For "NA" values in the HyHydrosoilv2.0 water content saturation layer, an Inverse Distance Weighting process was used to impute values and complete the numerical raster layer. These raster layers were then used as inputs to determine binary predictions (Figures 14 and 15) for each pixel based on the RF-trained model.

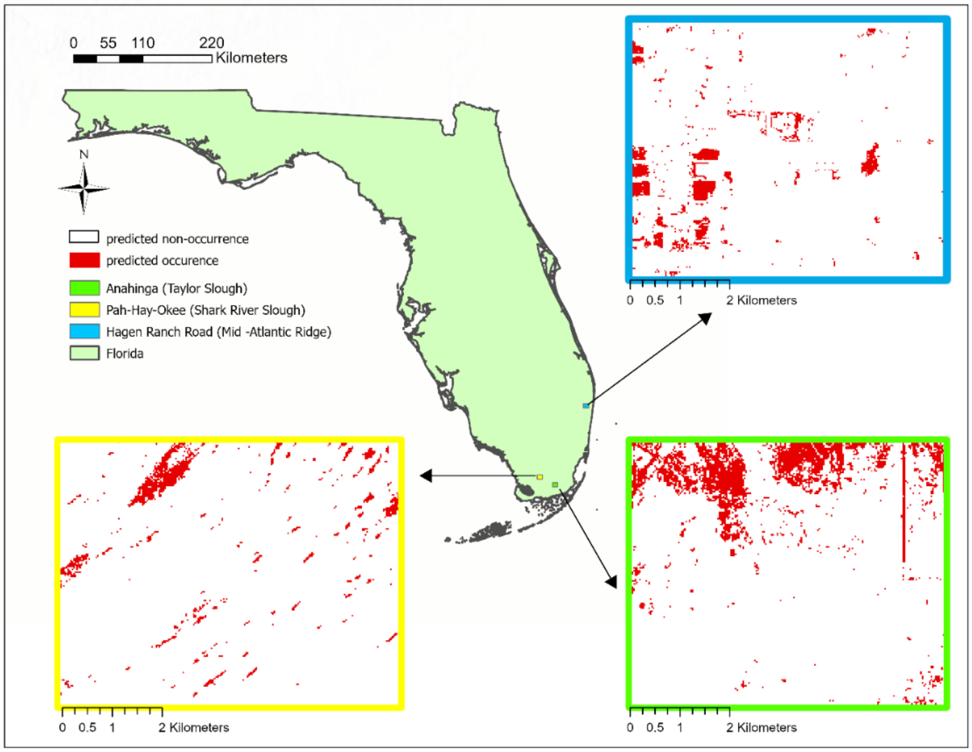

Figure 14. Binary prediction map of South Florida Study sites based on point-pixel Random Forest classification of HLS-L30 combined with SAR-1, and HyHydroSoilv2.0, then filtered through landcover type (NLCD 2021).

Sites in Peninsular Florida are situated in different watersheds with different characteristics and land cover compositions, which may potentially impact the distribution of *S. lancifolia*, but was not examined extensively in this study. The Random Forest algorithm trained on South Florida points was applied to each of the study sites and returned the probability of occurrence vs non-occurrence (Figure 14). The Anahinga study site, situated on the Taylor Slough returned a mean occurrence value of 19.75%, suggesting about 20% of this site is potentially occupied by *S. lancifolia*. This is nearly double the more urban site, Hagen Ranch Road, which had a mean occurrence value of

10%. The Pah-Hay-Okee study site, located within the Shark River Slough, showed the smallest mean occurrence of 6.33%.

Both study sites in Borikén (Puerto Rico) are located on the Northern Karst, within distinct watersheds in the Central Western (Caño Tiburonés) and Central Eastern (Sabana Seca) portions. A Random Forest algorithm which was trained on ground-verified collected field data points, as well as GBIF occurrences of record in Puerto Rico, was applied to each of the study sites and returned a binary probability of occurrence vs non-occurrence of *S. lancifolia* (Figure 15). The Caño Tiburonés returned a mean occurrence value of 13.26%, much higher than in the Sabana Seca site, where about 3.44% of the regions were predicted to be occupied by *S. lancifolia*.

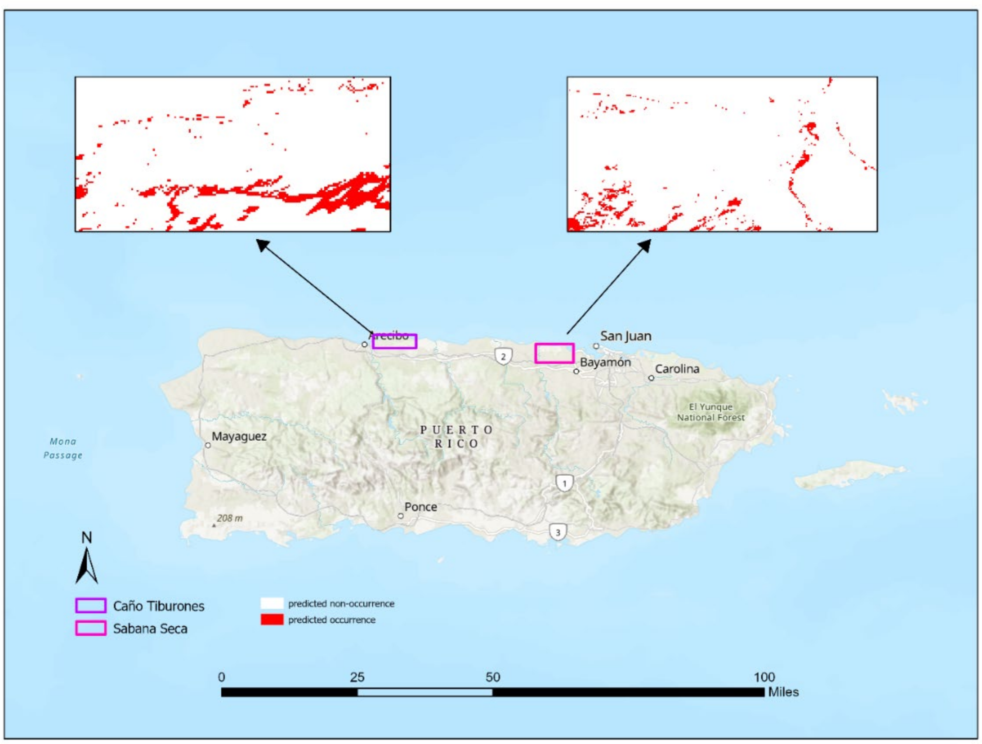

Figure 15. Binary prediction map of Puerto Rico sites, Sabana Seca and Caño Tiburonés based on RF trained in Florida using HLS-L30 combined with SAR-1, and HyHydroSoilv2.0, then filtered by landcover type from USFS 2020 landcover map.

2.4 Conclusion

Wetlands are highly complex and globally significant ecosystems that play significant roles in maintaining surface energy balance, influencing regional weather and climate regimes, nutrient cycling, methane sequestration, severe weather mitigation, preserving biodiversity, and redeveloping economic systems through applying policies related to the protection of blue carbon assets. Increasingly, the invaluable ecosystems face degradation, stress, and loss. A robust method for remotely mapping and monitoring wetland vegetation is a valuable tool for stakeholders and decision makers. The methods used in this study were found to be robust in terms of mapping *S. lancifolia* across Florida and Puerto Rico (>85%). In all 100 RF iterations, HLS-L30 data outperformed HLS-S30 data by slight margins. Differences in classification accuracy between sensors might be attributed to variations in surface conditions, elevation, and the spectral resolutions of each sensor. The method proposed in this study can be extended to other vegetation species, land surfaces, or trace gases in the atmosphere (such as methane), and be applied in the longer term to carbon accounting for blue carbon mapping, monitoring, and economic valuation.

Additional field campaigns to validate occurrence/non-occurrence points within specific timeframes of study would add to the robustness of the developed methodology. Further future work to increase the robustness of this methodology might also combine planar data with the existing cross-sectional data, examine spectral profiles of target vegetation *S. lancifolia* at higher spectral resolution (hyperspectral), and examine the thermal profiles at sites of occurrences/non-occurrences (at <0.5m, 0.5m, 1.0m,2m, and

10m) and changes in hydrologic sedimentation/sediment flow of the region. Since *S. lancifolia* is flood- and salt-tolerant, occurrences and non-occurrences through periods of drought and/or extreme rainfall would also be worth exploring through a quantitative analysis of local hydroperiods and variation in vegetative responses across the various communities where *S. lancifolia* is found.

The model trained in South Florida and applied to Borikén (Puerto Rico) underperformed both other models, while the model trained with mostly ground-verified points (Borikén) returned the highest accuracy and Kappa values, highlighting the intrinsic value of local data for robust performance of classification algorithms.

Chapter 3: Discussion, Limitations, and Future Recommendations

3.1 Extended Discussion and Limitations

When the RF classification method was applied on the combined datasets, accuracies were well above 80% in most cases, suggesting sufficient capabilities to detect *S. lancifolia* using freely available datasets. However, these accuracies were only achieved with the addition of ancillary soil moisture data, as well as Sentinel-1 SAR dual/cross-polarization mean values for each study site. This outcome suggests the need for ancillary data to improve model accuracy. Notably, the combination of spectral and thermal data from the HLS-L30 allowed for the HLS-L30 data to perform a random forest classification of pixels with higher accuracy than with just spectral data from the HLS-S30 (Harmonized Sentinel-2), suggesting a potential need for greater spectral resolution to detect *S. lancifolia*. Thus, the use of hyperspectral data would benefit the remote detection of *S. lancifolia* and its vegetative communities more effectively. Additionally, a number of studies demonstrate that longer wavelengths are better suited for forested wetlands due to their higher penetration capability (Li and Chen, 2005, Lang et al., 2008) such as L-band (15- 30 cm in wavelength), which have been shown to be useful in penetrating the canopy and discriminating understory vegetation from water and are sensitive to soil moisture, hence future studies might employ L-band time-series data to improve outcomes.

Some potential sources of error that might have impacted model accuracy include the number of points used (resulting in overfitting of the trained model), temporal variability, and missing values (NA') within the original data sets. Additionally, the inherent temporal variability (2-years) of HLS-L30/-S30 composites used for classification might potentially have impeded the efficiency of the classification model by theoretically homogenizing the phenology data; however, in early stages of this study, growing season reflectance values versus full-year reflectance values did not appear to have a significant impact. The variability in spatial resolutions among datasets might also be a source of error or reduced accuracy. While HLS-S30 Sentinel data is retrieved at 10 m, 20m, and 60m, HLS-L30 Landsat data which returned higher accuracy is retrieved at 30m. The HiHydroSoil v2.0 is at 250m spatial resolution, while NASA SMAP data is at 9 km spatial resolution. These inconsistencies in spatial resolution across datasets utilized might have resulted in reduced model accuracy.

3.2 Future Recommendations and Potential Applications

The results of this study showed our methodology for mapping *S. lancifolia to be* robust, however, mapping was limited to small subsets (~50 km^2) that were identified as potential sites through examining the GBIF occurrence record distribution density (in South Florida). Sites in Puerto Rico are scientific experimental sites of controlled access jointly managed largely by the USFS/USGS and the University of Puerto Rico. Primary recommendations following the results of this study are to allocate larger amounts of computational power and memory to be able to expand the extent of this mapping methodology to entire watersheds and/or state, county, or municipal bounds and compare accuracies according to this methodology. Another perspective that might potentially be of interest would be to test more advanced ML algorithms like deep learning, and artificial neural networks (ANNs), to more accurately model complex wetland features but also retain some level of stability with decision trees to prevent overfitting of data. A third critical recommendation is a field survey for *S. lancifolia* point-collection over selected sites, during an extremely wet/dry season, which would allow us to gain further insight into microclimate dynamics and vegetation responses of communities housing *S. lancifolia* (and potentially *Eleutherodactylus juanariveroii)* during flood or drought periods.

The fact remains, precisely assessing wetlands distributions, and net ecosystem productivity (NEP) remain critical for blue carbon accounting within local and global carbon budgets. Chen et. al (2025), details a methodology for calculating NEP of remotely sensed wetland areas in China, and illuminates trends from 1982-2020, which showed potential evapotranspiration (PET) as the dominant driver in dynamics of those wetlands. ANNs have been reported to perform more accurately than other remotely sensed image classification techniques (Goel et al. 2003; Mas and Flores 2008), especially when the feature space is complex, or the datasets have different statistical distributions which would potentially improve the results of this project. When ancillary data in addition to that of the spectral bands is incorporated (Benediktsson et al. 1993), artificial neural networks (ANN) might perform better than statistical classification techniques. Studies have shown than machine learning classification algorithms such as maximum likelihood (ML), support vector machine (SVM), decision trees (DTs), and artificial neural networks (ANNs) (Baker et al. 2006; Belluco et al. 2006; Sesnie et al. 2008; Szantoi et al. 2013; Wang et al. 2004; Wright and Gallant 2007) are a promising alternative to photointerpretation for mapping wetlands. (Santozi, Zoltan, et al., 2015). Baker et al. (2006) conversely found that multispectral image classification of wetlands has similar accuracy and is more repeatable relative to human photointerpretation.

Future studies for similar purposes in training machine learning algorithms to detect and differentiate vegetation or changes in atmosphere-surface chemistry would benefit from field campaigns to gather more ground-verified occurrence and non-occurrence points within a specific timeframe, as to better understand both the agreement between temporal data between spectral data, instead of utilizing a 2-year mean reflectance. This would be most effective as an airborne field survey with hyperspectral and thermal instrumentation, or an airboat campaign with field spectrometers.

Another potentially effective solution to further increase the resolution of the data and the robustness of the machine learning model and real-time processing applications would be to include sub-seasonal to seasonal precipitation amounts as variables. This would give more specific insights into the overall spectral response of *S. lancifolia* to variations throughout specific hydroperiods. At sites in Puerto Rico, *S. lancifolia* is embedded within elephant grass, a drought and flood tolerant prairie-grass native to Africa. Over larger areas, and especially when *S. lancifolia* is found within the understory, understanding the true distribution of *S. lancifolia* gives stakeholders, decision makers, and resource managers a better idea of wetland composition, and might be useful in understanding future landscape level changes, as well as preserving and protecting cultural assets such as the coquí.

Monitoring the changes of vegetation and hydrologic process and cycles over wetlands gives us deeper insight into regional land-atmosphere dynamics that are critical for understanding as conditions within the earth systems rapidly change, particularly in the tropics and subtropics. According to Rios-López (2007), microclimates are important controllers for the development of the species assembly. By including microclimate in ecosystem assembly rules, we broaden our understanding of factors that determine vertebrate succession in terrestrial habitats. Specifically, these data would allow for further understanding of thermal micro-profiles, micro-climates, and micro-habitats at the surface level, and as Rios-López (2007) suggests, would provide additional insight into the distributions of *Eleutherodactylus juanariveroii,* would allow for mapping complex ecosystems such as emergent wetlands at a finer scale, and perhaps be able to identify and protect additional potential locations of populations of *coqui llanero* and other endangered species.

In the current day, where local governments in both Peninsular Florida and Puerto Rico often favor short-sighted development 'opportunities' over the long-term stability of ecosystems and preserving local culture, it is imperative that policymakers and legislators utilize the most recent and accurate scientific and cultural knowledge. Recently, a Reddit thread where tourists discuss spraying questionable deterrents (possibly toxic chemicals) to keep *coqui* frogs away from their vacation rentals went viral within the local community, highlighting the lack of knowledge and respect that visitors and tourists have for the local ecosystems and the cultural values anchored in the ancestral history of the people that live on the island and among the diaspora. In Florida, developers often attempt to expand the urban boundaries into sensitive wetland ecosystems and relations such as Kahayatle, also known in English as the one and only Florida Everglades.

In these remote areas selected for this study, indigenous cultures (Miccosukee, Seminole, and Jibaro-Boricua/Arawak-Taino) maintain significant cultural ties to the lands, waters, and air flows. This type of *embodied* remote sensing is codified within cultural artifacts such as petroglyphs, celestial clocks, music and dance, storytelling, foodways and customs. Such embodied knowledge is passed along, anchored through time, and made continually available via culture. Hence, mapping the culturally significant *S lancifolia,* using open-source data can help better understand the *cultural resolution* of these remote areas, historically under indigenous stewardship. Cultural resolution can be defined as the breadth and depth of an area's significance within the longstanding cultural associations to that area or the ecosystem services that an area might provide. This study successfully utilizes open-source technology to map and protect valuable indigenous cultural, ecological, and economical assets, *Sagittaria lancifolia* and *Eleutherodactylus juanariveroii*, while demonstrating the utility of basic machine learning tools on readily available satellite datasets for monitoring specific wetland vegetation and can be easily modified and replicated for situational suitability. Since the datasets performed differently in different locations, with highest accuracy outcomes coming from the RF model trained on a majority of ground verified occurrence and non-occurrence points, this study also emphasizes the idea that local, ground-verified data is observably useful in increasing accuracy of classification of wetland vegetation communities in satellite imagery. Additionally, since the model trained in South Florida, and applied to Borikén (Puerto Rico) underperformed both models, the idea that local data is necessary for robust classification is further supported. Both of these outcomes highlight the notion that applications of and for the use of technology should be place-based, and integrate local knowledge, history, and culture for optimal resolution in decision and policy making.

References

Afifudin, Alfin Fatwa Mei, and Rony Irawanto. “Estimating the Ability of Lanceleaf Arrowhead (Sagittaria Lancifolia) in Phytoremediation of Heavy Metal Copper (Cu).” *Berkala Sainstek*, vol. 9, no. 3, 2021, pp. 125–30. https://doi.org/10.19184/bst.v9i3.26667.

Álvarez Ruiz, Migdalia, and Ariel E. Lugo. “Landscape Effects on Structure and Species Composition of Tabonuco Forests in Puerto Rico: Implications for Conservation.” *Forest Ecology and Management*, vol. 266, Feb. 2012, pp. 138–47. *DOI.org (Crossref)*, https://doi.org/10.1016/j.foreco.2011.10.044.

Aytekin, Örsan, Mehmet Koç, and Ilkay Ulusoy. "Local primitive pattern for the classification of SAR images." *IEEE transactions on geoscience and remote sensing* 51.4 (2012): 2431-2441.

Barbier, Edward B., et al. "The Value of Wetlands: Ecosystem Services and Economic Valuation." *Ecological Economics*, vol. 69, no. 11, 2011, pp. 269-280. https://doi.org/10.1016/j.ecolecon.2009.11.014.

Breiman, Leo. "Random Forests." *Machine Learning*, vol. 45, no. 1, 2001, pp. 5-32.

Brisco, B., et al. "A semi-automated tool for surface water mapping with RADARSAT-1." *Canadian Journal of Remote Sensing* 35.4 (2009): 336-344.

Ceriani, Lidia, and Paolo Verme. “The Origins of the Gini Index: Extracts from Variabilità e Mutabilità (1912) by Corrado Gini.” *Journal of Economic Inequality*, vol. 10, no. 3, 2012, pp. 421–43, doi:10.1007/s10888-011-9188-x.

Chen J, Tang Z, Kang X, He N, Li M. Rise in wetland carbon uptake linked to increased potential evapotranspiration. Environ Res. 2025 Aug 15;279(Pt 1):121778. doi: 10.1016/j.envres.2025.121778. Epub 2025 May 3. PMID: 40324617.

Costanza, Robert, et al. "The Value of the World's Ecosystem Services and Natural Capital." *Nature*, vol. 387, no. 6630, 1997, pp. 253-260. https://doi.org/10.1038/387253a0.

Claverie, Martin, et al. "The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set." *Remote Sensing of Environment*, vol. 219, 2018, pp. 145–161.

Cutler, David R., et al. "Random forests for classification in ecology." *Ecology*, vol. 88, no. 11, 2007, pp. 2783-2792.

David-Chavez, D.M. and Gavin, M.C. 2018. A global assessment of Indigenous community engagement in climate research. Environmental Research Letters 13(12): 123005, https://doi.org/10.1088/1748-9326/aaf300

Day, J. W., et al. "The Impacts of the Mississippi River on the Estuarine and Coastal Ecosystems of the Gulf of Mexico." *Estuaries*, vol. 23, no. 1, 2000, pp. 1-12. https://doi.org/10.2307/1352788.

Dewitz, J., 2023, National Land Cover Database (NLCD) 2021 Products: U.S. Geological Survey data release, https://doi.org/10.5066/P9JZ7AO3.

Duffy, J. E. "Biodiversity and the Functioning of Ecosystems: The Primacy of Species Interactions." *Ecology Letters*, vol. 9, no. 6, 2006, pp. 560-568. https://doi.org/10.1111/j.1461-0248.2006.00956.x.

Fan, Chang, et al. "Mapping Phenology of Complicated Wetland Landscapes through Harmonizing Landsat and Sentinel-2 Imagery." *Remote Sensing (Basel, Switzerland)*, vol. 15, no. 9, 2023, p. 2413, doi:10.3390/rs15092413.

Fitrihidajati, H., et al. "The Ability of Sagittaria Lancifolia as Phytoremediator on Detergent Solution." *Journal of Physics. Conference Series*, vol. 1569, no. 4, 2020, p. 42072. https://doi.org/10.1088/1742-6596/1569/4/042072.

Fu, Bolin, et al. "Comparison of Object-Based and Pixel-Based Random Forest Algorithm for Wetland Vegetation Mapping Using High Spatial Resolution GF-1 and SAR Data." *Ecological Indicators*, vol. 73, 2017, pp. 105–17, doi:10.1016/j.ecolind.2016.09.029.

Gao, Q., and M. Yu. "Discerning Fragmentation Dynamics of Tropical Forest and Wetland during Reforestation, Urban Sprawl, and Policy Shifts." *PLOS ONE*, vol. 9, no. 11, 2014, e113140. https://doi.org/10.1371/journal.pone.0113140.

Giardina, C., and T. Wood. *Tropical Island Forests and Climate Change*. U.S. Department of Agriculture, Forest Service, Climate Change Resource Center, 2015.

Gini, C (1912). Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche. Bologna: C. Cuppini.

Hong, Changqiao, et al. "The Role of Harmonized Landsat Sentinel-2 (HLS) Products to Reveal Multiple Trajectories and Determinants of Cropland Abandonment in Subtropical Mountainous Areas." *Journal of Environmental Management*, vol. 336, June 2023, p. 117621. *DOI.org (Crossref)*, https://doi.org/10.1016/j.jenvman.2023.117621.

Henderson, Floyd M., and Anthony J. Lewis. "Radar detection of wetland ecosystems: a review." *International journal of remote sensing* 29.20 (2008): 5809-5835.

ICF International, in association with R2 Consultants. 2012. Wapato Reach Assessment Report. April 26. (ICF 00703.11.) Seattle, WA. Prepared for Yakama Nation, Toppenish, WA.

Jin, Huiran, Giorgos Mountrakis, and Stephen V. Stehman. "Assessing integration of intensity, polarimetric scattering, interferometric coherence and spatial texture metrics in PALSAR-derived land cover classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 98 (2014): 70-84.

Koma, Zsófia, et al. “Classifying Wetland-Related Land Cover Types and Habitats Using Fine-Scale Lidar Metrics Derived from Country-Wide Airborne Laser Scanning.” *Remote Sensing in Ecology and Conservation*, edited by Nathalie Pettorelli and Mat Disney, vol. 7, no. 1, Mar. 2021, pp. 80–96. https://doi.org/10.1002/rse2.170.

Kokaly, Raymond F., et al. "Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data." *Remote sensing of environment* 84.3 (2003): 437-456.

Kumar, Pushpam, editor. *The Economics of Ecosystems and Biodiversity Ecological and Economic Foundations*. Earthscan, 2010.

Li, Junhua, and Wenjun Chen. "A rule-based method for mapping Canada's wetlands using optical, radar and DEM data." *International Journal of Remote Sensing* 26.22 (2005): 5051-5069.

Lievens, H., et al. “Joint Sentinel-1 and SMAP Data Assimilation to Improve Soil Moisture Estimates.” *Geophysical Research Letters*, vol. 44, no. 12, 2017, pp. 6145–53. https://doi.org/10.1002/2017gl073904.

Liaw, Andy, and Matthew Wiener. "randomForest: Breiman and Cutler's Random Forests for Classification and Regression." *R Package Documentation*, 2007.

Masek, J., Ju, J., Roger, J., Skakun, S., Vermote, E., Claverie, M., Dungan, J., Yin, Z., Freitag, B., Justice, C. (2021). HLS Operational Land Imager Surface Reflectance and TOA Brightness Daily Global 30m v2.0 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2023-09-12 from https://doi.org/10.5067/HLS/HLSL30.002

Martin, Shannon B., and Gary P. Shaffer. “Sagittaria Biomass Partitioning Relative to Salinity, Hydrologic Regime, and Substrate Type: Implications for Plant Distribution Patterns in Coastal Louisiana, United States.” *Journal of Coastal Research*, vol. 21, Jan. 2005, pp. 167–74. https://doi.org/10.2112/02110.1.

Mahdianpari, Masoud, et al. "Random Forest Wetland Classification Using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, 2017, pp. 13-31. https://doi.org/10.1016/j.isprsjprs.2017.05.010.

Mitsch, William J., and James G. Gosselink. *Wetlands*. 5th ed., Wiley, 2015.

Mohammadimanesh, Fariba, et al. "An Efficient Feature Optimization for Wetland Mapping by Synergistic Use of SAR Intensity, Interferometry, and Polarimetry Data." *International Journal of Applied Earth Observation and Geoinformation*, vol. 73, 2018, pp. 450-462.

Monroe, Kelen D., et al. "Occupancy and Abundance of Eleutherodactylus Frogs in Coffee Plantations in Puerto Rico." *Herpetologica*, vol. 73, no. 4, Dec. 2017, p. 297. *DOI.org (Crossref)*, https://doi.org/10.1655/Herpetologica-D-16-00089.

Murray, B.C. et al. (2011). Green Payments for Blue Carbon

Nellemann, Christian, and Emily Corcoran, eds. *Blue carbon: the role of healthy oceans in binding carbon: a rapid response assessment*. UNEP/Earthprint, 2009.

Niculescu, Simona, et al. "Synergy of High-Resolution Radar and Optical Satellite Images for Identification and Mapping of Wetland Macrophytes on the Danube Delta." *Remote Sensing*, vol. 12, no. 14, 2020, p. 2188. https://doi.org/10.3390/rs12142188.

Pendleton, Linwood, et al. "Estimating the Value of Coastal Ecosystem Services in the U.S." *PLOS ONE*, vol. 7, no. 1, 2012, e43593. https://doi.org/10.1371/journal.pone.0043593.

Reichle, R., et al. *SMAP L4 Global 3-Hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data*. SPL4SMGP, Version 7, NASA National Snow and Ice Data Center Distributed Active Archive Center, 2022. https://doi.org/10.5067/EVKPQZ4AFC4D.

Rios-López, Neftalí, and T. Mitchell Aide. "Herpetofaunal Dynamics During Secondary Succession." *Herpetologica*, vol. 63, no. 1, 2007, pp. 35–50. https://doi.org/10.1655/0018-0831(2007)63[35:HDDSS]2.0.CO;2.

Rivera-Burgos, Ana C. and Martinez, Eloy and Chaparro, Rafael and Terando, Adam J. and Majumder, Reetam and Collazo, Jaime Agustin, Critical Thermal Maximum of Six Species of Eleutherodactylus Frogs in Puerto Rico: Implications for Conservation in a Warming Environment. Available at SSRN: https://ssrn.com/abstract=4556214 or http://dx.doi.org/10.2139/ssrn.4556214

Schedlbauer, Jessica L., et al. "Controls on Sensible Heat and Latent Energy Fluxes from a Short-Hydroperiod Florida Everglades Marsh." *Journal of Hydrology*, vol. 411, no. 3–4, Dec. 2011, pp. 331–41. https://doi.org/10.1016/j.jhydrol.2011.10.014.

Sheehy, Jack, et al. "Redefining Blue Carbon with Adaptive Valuation for Global Policy." *The Science of the Total Environment*, vol. 908, 2024, pp. 168253–168253, doi:10.1016/j.scitotenv.2023.168253.

Simons, G. W. H., et al. *HiHydroSoil v2.0 - A High Resolution Soil Map of Global Hydraulic Properties*. FutureWater Report 213, 2020.

Soylu, Mehmet Evren, and Rafael L. Bras. “Global Shallow Groundwater Patterns From Soil Moisture Satellite Retrievals.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, 2022, pp. 89–101, doi:10.1109/JSTARS.2021.3124892.

Szantoi, Zoltan, et al. "Classifying Spatially Heterogeneous Wetland Communities Using Machine Learning Algorithms and Spectral and Textural Features." *Environmental Monitoring and Assessment*, vol. 187, no. 5, 2015, pp. 1-15. https://doi.org/10.1007/s10661-015-4426-x.

Taylor, Christopher M., et al. “Mesoscale Rainfall Patterns Observed around Wetlands in sub-Saharan Africa.” *Quarterly Journal of the Royal Meteorological Society*, vol. 144, no. 716, Oct. 2018, pp. 2118–32. *DOI.org (Crossref)*, https://doi.org/10.1002/qj.3311.

Temmerman, Stijn, et al. "Vegetation Causes Changes in Wave Regime in Coastal Wetlands: Implications for Wave-Based Erosion and Sedimentation." *Geology*, vol. 41, no. 3, 2013, pp. 227-230. https://doi.org/10.1130/G33860.1.

Wang, Yongjun, Shuanggen Jin, and Gino Dardanelli. "Vegetation Classification and Evaluation of Yancheng Coastal Wetlands Based on Random Forest Algorithm from Sentinel-2 Images." *Remote Sensing*, vol. 16, no. 7, 2024, p. 1124. https://doi.org/10.3390/rs16071124.

Washines, Emily, and Jerry Peltier. *Natural Restoration and Cultural Knowledge of the Yakama Nation*. The Evergreen State College, 2010. PDF file. “Natural Restoration and Cultural Knowledge of the Yakama Nation,” 2010 copyright © The Evergreen State College. “washinescasestudy.pdf,” Evergreen State College website, accessed July 21, 2025

Wu Y, Xi Y, Feng M, Peng S. Wetlands Cool Land Surface Temperature in Tropical Regions but Warm in Boreal Regions. *Remote Sensing*. 2021; 13(8):1439. https://doi.org/10.3390/rs13081439

Yang, H., et al. "Mapping Dongting Lake Wetland Utilizing Time Series Similarity, Statistical Texture, and Superpixels With Sentinel-1 SAR Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, 2022, pp. 8235-8244. https://doi.org/10.1109/JSTARS.2022.3206246

Zhang, Boya, et al. "Spatiotemporal Variations of Wetland Backscatter: The Role of Water Depth and Vegetation Characteristics in Sentinel-1 Dual-Polarization SAR Observations." *Remote Sensing of Environment*, vol. 270, 2022, p. 112864.

Zhang, Xiaotong, et al. "Coastal Wetland Classification with GF-3 Polarimetric SAR Imagery by Using Object-Oriented Random Forest Algorithm." *Sensors*, vol. 21, no. 10, 2021, p. 3395. https://doi.org/10.3390/s21103395.

Zhao, Xiaochen, et al. “Temperature Across Vegetation Canopy-Water-Soil Interfaces Is Modulated by Hydroperiod and Extreme Weather in Coastal Wetlands.” *Frontiers in Marine Science*, vol. 9, May 2022, p. 852901. *DOI.org (Crossref)*, https://doi.org/10.3389/fmars.2022.852901.