Title

Pruning our expectations: The implications and applications of earth observation in developing urban forestry tools in Los Angeles County

Author

Ocón, Jonathan Pando

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Pruning our expectations: The implications and applications of earth observation in developing urban forestry tools in Los Angeles County

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Geography

by

Jonathan Pando Ocón

2023

ABSTRACT OF THE DISSERTATION

Pruning our expectations: The implications and applications
of earth observation in developing urban forestry tools
in Los Angeles County

by

Jonathan Pando Ocón
Doctor of Philosophy in Geography
University of California, Los Angeles, 2023
Professor Thomas W. Gillespie, Co-Chair
Professor Elsa M. Ordway, Co-Chair

This dissertation, largely funded by the County of Los Angeles to optimize its urban forest planning and management, conducts a critical examination of the implications, applications, and limitations of remote sensing technology in applied science. Leveraging technologies such as Light Detection and Ranging (LiDAR) and high-resolution optical data, remote sensing has offered unparalleled insights into the Earth's surface and atmosphere. Specific applications, such as tree crown segmentation and creation of precise 3D environmental models, have proven invaluable in fields including ecology and urban planning. However, this technology is not without its challenges. An over-reliance on remote sensing data, absent corroborative

ground-truthing, can lead to flawed conclusions. Furthermore, using low-resolution imagery for policy decisions concerning fine-scale, on-the-ground issues could create discrepancies between the data and actual conditions, undermining the effectiveness of interventions. In my first and second chapter of this dissertation, I demonstrate that remote sensing technology has become indispensable, particularly for monitoring individual species over time for urban forest management. By identifying and delineating individual tree crowns, this technology enables accurate estimation of species composition and condition. This detailed information facilitates species health monitoring and early disease detection, thereby aiding effective management interventions. However, the effectiveness of remote sensing in urban forestry necessitates continuous data update and integration of ground-based observations for validation. In my third chapter, I transition from the practical applications of remote sensing in urban forestry, and lean on previous research experience, including my collaboration with Los Angeles County, to delve into the social dimensions of this technology. The concept of "socializing the pixel" provides a critical framework for understanding how remote sensing operates within and is influenced by sociopolitical contexts. This perspective recognizes that remote sensing is not a purely objective tool but is embedded within social structures and power dynamics. Stakeholder engagement is thus crucial to consider diverse perspectives in data interpretation and application, enhancing the technology's effectiveness and local relevance. Conversely, power imbalances between remote sensing experts and non-experts, and denial of local knowledge can lead to biased results or misinterpretation of data. Hence, returning to the concept of "socializing the pixel" is critical to ensure effective and equitable use of remote sensing technology.

The dissertation of Jonathan Pando Ocón is approved.

Stephanie S. Pincetl

Michael E. Shin

Victoria Kelly Turner

Thomas W. Gillespie, Committee Co-Chair

Elsa M. Ordway, Committee Co-Chair

University of California, Los Angeles

2023

For Steph,

she knew even before I did.

Table of Contents

Vita

EDUCATION

University of California, Los Angeles, CA 2020
Master of Arts, Geography

University of Southern California, Los Angeles, CA 2013
Bachelor of Science, Policy, Planning, and Development

PEER-REVIEWED PUBLICATIONS

Ocón, J. P., Stavros, E. N., Steinberg, S. J., Robertson, J., & Gillespie, T. W. (2023) Using Multi-source Remote Sensing Imagery to Identify Urban Tree Species. Progress in Physical Geography. *Progress in Physical Geography. In Revisions.*

Madson, A., Dimson, M., Fortini, L. B., Kawelo, K., Ticktin, T., Keir, M., ... & Gillespie, T. W. (2022). A near four-decade time series shows the Hawaiian Islands have been browning since the 1980s. *Environmental Management*, 1-16.

Kelly Turner, V., Rogers, M. L., Zhang, Y., Middel, A., Schneider, F. A., Ocón, J. P., ... & Dialesandro, J. (2022). More than surface temperature: mitigating thermal exposure in hyper-local land system. *Journal of Land Use Science*, 17(1), 79-99.

Ordway, E. M., Elmore, A. J., Kolstoe, S., Quinn, J. E., Swanwick, R., Cattau, M., ... & Wilson, C. (2021). Leveraging the NEON Airborne Observation Platform for socio‐environmental systems research. *Ecosphere*, 12(6), e03640.

Ocón, J. P., Ibanez, T., Franklin, J., Pau, S., Keppel, G., Rivas-Torres, G., ... & Gillespie, T. W. (2021). Global tropical dry forest extent and cover: A comparative study of bioclimatic definitions using two climatic data sets. PloS one, 16(5), e0252063.

AWARDS & GRANTS

MOU Los Angeles County	2020 – 2023
K. Patricia Cross Future Leaders Award, Finalist, AAC&U	2023
Inaugural California Resident, Nature, Art, & Habitat Residency	2021
Graduate Dean’s Scholar Award, UCLA	2019 & 2020
Geospatial @ UCLA Summer Fellowship	2019

RESEARCH EXPERIENCE

Optimizing the Planning & Management of Los Angeles County’s Urban Forest | L. A. County
Aug 2020 – Aug 2023

Awarded 3-year research grant awarded by the Los Angeles County Department of Public Health to automate individual tree species identification using multisource airborne remote sensing (VHR, LiDAR, and imaging spectroscopy).

Individual Tree Segmentation Using LiDAR in Tahoe Basin | Dr. Sassan Saatchi, NASA JPL
Mar – Jun 2022

Segmented single tree point clouds using discrete-return LiDAR and an adaptive mean shift algorithm.

SHIFT Field Campaign | Dr. Dana Chadwick, NASA JPL & Dr. Elsa Ordway, UCLA
Jan & May 2022

Collected terrestrial plant samples during the NASA JPL SBG precursor SHIFT campaign to study plant trait changes during the transition from the wet to dry season in the Sedgwick and Dangermond Reserves near Santa Barbara, CA

Global Analysis of Tropical Dry Forest Extent | Dr. Thomas W. Gillespie, UCLA
Aug 2018 – Sep 2020

Compared climatic definitions of tropical dry forest extent for conservation assessment using remote sensing and geospatial programming.

Impact of Urban Form on Thermal Comfort | Dr. V. Kelly Turner, UCLA
Aug 2019 – Jul 2020

Modeled mean radiant and land surface temperatures at fine resolutions to understand thermal comfort in underrepresented neighborhoods in California, and a new urbanist development in Tucson, Arizona using remote sensing, geospatial programming, and microclimate modeling.

Housing Development and Design in Los Angeles | Liz Falleta, USC
Aug 2012 – May 2013

Ascertained historical records and tenant data on three multi-unit, architecturally significant housing projects in Los Angeles to highlight the importance of well-designed, multi-unit housing for future development. Falletta, L. (2019). By-right, By-design: Housing Development Versus Housing Design in Los Angeles. Routledge.

TEACHING EXPERIENCE

Computer Programming for GIS, Masters of GIS Program, Clark University | Oct – Dec 2023
Python Programming, Masters of GIS Program, Clark University | Aug – Oct 2023
Web Mapping & Open Source GIS, Masters of GIS Program, Clark University | Aug – Dec 2023
Introduction to GIS, Masters of GIS Program, Clark University | Aug – Dec 2023
Cartography, Department of Geography, UCLA | Jun – Jul 2021
GIS Programming and Development, Department of Geography, UCLA | Apr – Jun 2021
Cartography, Department of Geography, UCLA | Aug – Sep 2020
Introduction to GIS, Department of Geography, UCLA | Jan – Mar 2020

ACADEMIC SERVICE

Associate Editor | Journal of Environmental Studies and Sciences (JESS)

Springer, Apr 2022 – Present

Graduate Student Steering Committee Member | Landscape Exchange Network for Socio-environmental systems (LENS)

NSF: DEB-2054939, Mar 2020 – Present

President and Social Chair | Geography Graduate Students Association, UCLA

Sep 2019 – Jun 2020

Graduate Student Representative, Geography | Graduate Student Advisory Board, UCLA

Sep 2019 – Jun 2020

Chapter 1.

Overview of Socio-ecological Applications in Remote Sensing

Remote sensing is a powerful tool that has revolutionized the way we understand and interact with our environment. Dating back to aerial photography in the 19th century and evolving with advancements in satellite imaging in the 20th century, it involves the use of sensors to collect data from a distance. This data can then be analyzed to gain insights into different aspects of the Earth's surface and atmosphere. Today, remote sensing is an indispensable tool in fields such as ecology, geography, and urban planning. In particular, remote sensing data allows for the monitoring of urban, rural and agricultural expansion and provides a temporal understanding of changing conditions. Specifically, remote sensing imagery provides valuable information on ecosystem properties, and can be used alongside qualitative research, and as a complement to ethnographic studies (Dennis et al., 2005; Isager et al., 2007; Liverman et al., 2008; Niemiec et al., 2018). Through an exchange between remote sensing and social scientists, there exists a reciprocity in the information gained by both, and as remote sensing pertains to local knowledge, sometimes the best analysis lies in humans-as-sensors. This idea of validating remotely sensed imagery with local knowledge is a method already being deployed among indigenous peoples around the world. As indigenous peoples relearn and restore ancestral hunting and foraging techniques as a part of a larger struggle for decolonization, they engage with science to enhance traditional ways of knowing and the reciprocity of having this technology in-hand gives what one Alaskan subsistence hunter equated to an “elder in a box,” (Rammage et al., 2020; Rattling Leaf Sr. et al., 2020). Also, by including ground referenced data, especially as it pertains to stakeholders, remote sensing can provide an observational asset in relevant technical approaches such as agent-based modeling, cloud computing, community science, and humans-as-sensors to study social-ecological change (Pricope et al., 2019).

Remote sensing also allows for longitudinal studies to measure change as it occurs. To fully comprehend the feedback mechanisms between human behavior and the environment—how the environment impacts human decision-making, and in turn, how human decision-making impacts the environment—data are required at appropriate spatial, temporal, and socioeconomic scales of interest (Cumming et al., 2006). For nearly 40 years, moderate (10 – 60 m) to coarse (> 60 m) spatial and spectral resolution remote sensing imagery has been publicly available to monitor and measure landscape changes such as deforestation or changes in land surface temperature, which have led to deeper understandings of socio-ecological phenomena like the impact of logging on fire regimes or the drivers of the urban heat island effect. LiDAR (Light Detection and Ranging) and optical data are the primary sources of remote sensing data used in ecological studies. LiDAR provides precise 3D representations of terrestrial environments, especially useful in urban contexts. Optical data captures reflected light across various wavelengths that correspond to different biophysical properties of the objects reflecting or emitting light, offering insights into vegetation health and composition. When combined, these data can be used to provide comprehensive insights, from 3D terrain models to vegetation maps. Fine resolution remotely sensed data (< 10 m) can further address challenges in socio-ecological systems by linking larger landscape scale drivers like drought to hyperlocal, leaf-level observations such as canopy water content (Ordway et al., 2021).

Chapters Two & Three: Local Applications in Los Angeles County

Remote sensing data allows for the monitoring of urban, rural, and agricultural expansion, offering a temporal understanding of changing conditions. In Los Angeles County, trees are a critical infrastructure as they help cool neighborhoods, clean air and water, and improve our communities' emotional and social health. Los Angeles County manages over 800,000 public trees, with inventory costs using traditional methods exceeding $5 million (Los Angeles County Department of Public Health, 2020). These approaches are also limited to trees in the public right-of-way. The ability to inventory public and private trees using remote sensing will provide the county with a faster, cheaper, and more efficient way to inventory trees throughout the region. Although they cannot fix our social ills, trees are still a valuable resource for the county's most vulnerable populations, including but not limited to elderly and economically disadvantaged communities. Redevelopment and land-use dynamics can disrupt urban tree species' natural life cycle. Regardless of their benefits to residents, urban trees are subject to how we reconfigure urban spaces (Avolio et al., 2015). Moreover, invasive species, diseases, and climate change are also posing additional threats, and to address some of these challenges Los Angeles County has partnered with the Department of Geography at UCLA to identify the best approach in tree species identification and health assessment to help ease the burden of managing local trees manually. Currently, no department within the county has a cost-effective or efficient way to know where and when to plant or manage trees. Those in charge of managing the urban forest are stretched thin and often have disparate tree inventory data, making hands-on management difficult to scale across the region.

Tree crown segmentation is a crucial aspect of ecological studies that utilize remote sensing data. It involves the identification and delineation of tree crowns from other features on the Earth's surface, such as buildings or roads. This process enables researchers to accurately estimate key metrics such as tree height, and important crown structure such as width, length, area, and perimeter. Tree crown segmentation also plays a vital role in species identification by implementing detailed crown structure and tree height information. 3D differentiation between species can help distinguish individual trees from each other within 2D optical data sets. Chapters two and three will cover LiDAR processing for individual tree crown segmentation, as well as classifying trees to species using a machine learning model, Random Forest, to optimize and automate Los Angeles County's urban forest management. The goal of both chapters is to alleviate the burden of local departments that currently conduct manual, visual assessments of local tree stock in one of the largest and most populous counties in the United States.

Chapter Four: Beyond Science Applications – Interrogating Sociopolitical Dynamics in Applied Remote Sensing

There is immense power and responsibility in the technology and programming conducted to quantify and model complex, multitemporal, and multiscalar socio-ecological change, like Los Angeles County’s urban forest. With these recent advancements and their impact on policy and planning, and contribution to ecosystem science, it is essential to acknowledge the actors involved in co-producing socio-ecological systems (SES) research, the agency afforded or removed from each actor, and the narratives at play in the decision-making process (Robbins, 2001; Robbins, 2003; Turner et al., 2016). Methods used to simplify and classify landscapes have a buried complexity. As classifications become institutionalized, their complexity becomes more widely accepted and unquestioned, making a case for ground truth observations and demonstrating that maps are representations that systemically establish their authority (Zubrow, 2003). It becomes imperative to critically analyze our methods, data collection, and models to understand better who and what our science impacts (Runk et al., 2010).

Ground truth observations are crucial, reminding us that maps are mere representations, yet powerful communication tools—or boundary objects—for individuals engaging across professional or academic boundaries. We must critically analyze our methods, data, and models to ensure our science benefits all. It is essential to note that remote sensing data can also be misused or misinterpreted. Scientists and stakeholders may rely solely on this technology and neglect crucial ground-truthing measures, leading to flawed conclusions and misguided decision-making. In chapter four, I use a combination of my own experience working as a remote sensing scientist along with surveying and interviewing participants across the geospatial industry to interrogate how politics, like the omission of key stakeholders, influences this all-too-common oversight.

Dissertation Road Map

The transformative power of remote sensing in urban forestry management and broader socio-ecological systems cannot be overstated. With rapidly advancing geospatial technologies, there arises an intricate web of data producers, scientists, users, and those directly impacted by the decisions made as a result of geospatial studies. This dissertation delves deep into these interconnections, specifically within the context of Los Angeles County's urban forestry. The first two chapters focus on more technical terrains. Chapter two and three cover my work developing urban forestry tools for Los Angeles County. Chapter two tested five existing tree segmentation algorithms for LiDAR processing and found that two, Dalponte2016 and AMS3D, outperformed the other three in terms of accuracy and promise of segmenting trees within closed canopy, but found that Dalponte2016 was faster and less computationally expensive than AMS3D, so was the better option for County managers to use when processing future LiDAR acquisitions.

In chapter three, I develop a Random Forest algorithm in Google Earth Engine using the tree height information gathered from the Dalponte2016 algorithm in chapter two, as well as spectral information from NAIP 2016 data at a 60 cm spatial resolution. A 14-band stack was created by processing spectral and structural information, as well as transforming RGB color space into HSV color space, shadow masking, and calculating vegetation indices. The model performed with relative success and correctly classified an overall 89% of individual tree species. This is another benefit for County managers as Random Forest is a simple and robust model to use for classification purposes, and performing the analysis in Google Earth Engine avoids heavy data downloads.

Chapter four seeks to interrogate the sociopolitical space of science-based solutions where remote sensing is involved. This includes policy and land management decision-making in collaboration with public and private stakeholders. I define the entire remote sensing data handling process as spanning four distinct groups that I define as Data Producers, Data Scientists, Data Users, and Impacted Persons. Through participant observation (including the collaborative process with Los Angeles County from chapters two and three), surveys with those familiar with geospatial technologies or use them in their work and decisions, and interviews with remote sensing experts, I come to the conclusion that Impacted Persons--or those that bear the brunt of decisions made without their involvement--should be included in the information exchange between Data Producers and Scientists and Data Users. To operationalize this involvement is to give recognition to local knowledge and enhance already robust remote sensing studies further with even more critical ground truthing.

References

Avolio, M., Pataki, D. E., Gillespie, T., Jenerette, G. D., McCarthy, H. R., Pincetl, S., & Weller-Clarke, L. (2015). Tree diversity in southern California's urban forest: the interacting roles of social and environmental variables. *Frontiers in Ecology and Evolution*, 3, 73.

Cumming, Graeme S., David HM Cumming, and Charles L. Redman. "Scale mismatches in social-ecological systems: causes, consequences, and solutions." *Ecology and Society* 11.1 (2006).

Dennis, Rona A., et al. "Fire, people and pixels: linking social science and remote sensing to understand underlying causes and impacts of fires in Indonesia." *Human Ecology* 33.4 (2005): 465-504.

Isager, L., & Broge, N. H. (2007). Combining remote sensing and anthropology to trace historical land-use changes and facilitate better landscape management in a sub-watershed in North Thailand. *Landscape Research*, 32(2), 147-169.

Liverman, D. M., & Cuesta, R. M. R. (2008). Human interactions with the Earth system: people and pixels revisited. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 33(9), 1458-1471.

Niemiec, R. M., Asner, G. P., Brodrick, P. G., Gaertner, J. A., & Ardoin, N. M. (2018). Scale-dependence of environmental and socioeconomic drivers of albizia invasion in Hawaii. *Landscape and Urban Planning*, 169, 70-80.

Ordway, E. M., Elmore, A. J., Kolstoe, S., Quinn, J. E., Swanwick, R., Cattau, M., ... & Wilson, C. (2021). Leveraging the NEON Airborne Observation Platform for socio‐environmental systems research. *Ecosphere*, 12(6), e03640.

G Pricope, N., L Mapes, K., & D Woodward, K. (2019). Remote sensing of human–environment interactions in global change research: A review of advances, challenges and future directions. *Remote Sensing*, 11(23), 2783.

Ramage, S., Carroll, S. R., Hudson, M., David-Chavez, D. M., Kenote, T., Anderson, J., ... & Shanley, L. (2020, December 7-9). Indigenous Data Sovereignty [Conference session]. GEO Indigenous Summit, Virtual. https://www.earthobservations.org/indigenoussummit2020.php

Rattling Leaf Sr., J., Lenkokwai, R., Sudlovenick, C., Schmidt, C., DeRoy, S., Zwartjes, L., ... & Prados, A. (2020, December 7-9). Education / Intergenerational Knowledge Transfer [Conference session]. GEO Indigenous Summit, Virtual. https://www.earthobservations.org/indigenoussummit2020.php

Robbins, P. (2001). Fixed categories in a portable landscape: the causes and consequences of land-cover categorization. *Environment and Planning* A, 33(1), 161-179.

Robbins, P. (2003). Beyond ground truth: GIS and the environmental knowledge of herders, professional foresters, and other traditional communities. *Human Ecology*, 31(2), 233-253.

Runk, J. V., Negría, G. O., Conquista, L. P., Peña, G. M., Cheucarama, F. P., & Chiripua, Y. C. (2010). Landscapes, legibility, and conservation planning: multiple representations of forest use in Panama. *Conservation Letters*, 3(3), 167-176.

Turner Ii, B. L., Esler, K. J., Bridgewater, P., Tewksbury, J., Sitas, N., Abrahams, B., ... & Mooney, H. (2016). Socio-Environmental Systems (SES) Research: what have we learned and how can we use this information in future research programs. *Current Opinion in Environmental Sustainability*, 19, 160-168.

Zubrow, A. A. (2003). Mapping tension: remote sensing and the production of a statewide land cover map. *Human Ecology*, 31(2), 281-307.

Chapter 2.

What's the point? A comparative analysis of individual tree segmentation algorithms using airborne LiDAR in urban Los Angeles County

Abstract

Urban forests significantly contribute to the well-being of residents by delivering vital ecosystem services and promoting biodiversity. Precise and efficient tree segmentation is crucial for effectively monitoring and managing these green spaces. This research investigates the performance of five unique tree segmentation algorithms to identify a suitable method for large-scale, automated urban tree detection across open and closed canopies. The criteria for determining a suitable model included a combination of computational run time, storage and memory needed to process input and output data, and accuracy of tree crown delineation when compared to manual tree crown segmentation. My findings demonstrate that an Adaptive Mean Shift technique generates the most accurate tree segments, albeit with a high computational cost and species specificity. Conversely, a less resource-intensive solution produced satisfactory tree segments for 20 species in Los Angeles County, California. Future research should consider employing the Adaptive Mean Shift method for structurally similar target species when computational time and resources are not a concern or opt for the less computationally expensive algorithm to attain suitable results for a diverse range of tree species with reduced computational demands.

Keywords

Airborne Remote Sensing, Individual Tree Segmentation, Light Detection and Ranging, Urban Forestry

1 Introduction

While trees indeed play a pivotal role in urban environments by contributing to local cooling and enhancing social and emotional well-being, their integration into urban spaces comes with multifaceted challenges and considerations, especially since some of the touted benefits, like energy savings or air quality management, may be more pronounced in cities with year-round warm climates (Mullaney et al., 2015; Roman et al., 2021). This is particularly true in Los Angeles County, California, where extreme heat events are on the rise, long term droughts are commonplace, tree canopy cover is unevenly distributed along wealth lines, and dangers associate to wildfires and high winds impact residential areas (Beuhler, 2003; McCarthy et al., 2010; Hulley 2012; Gago et al., 2013; Adelaine et al., 2017; Hulley et al., 2019; Grant and Hicks, 2020). Although they cannot fix our social ills, trees are still a valuable resource for the County's most vulnerable populations, including but not limited to older and economically disadvantaged communities (Dong et al., 2023). Redevelopment and land-use dynamics can disrupt urban tree species' natural life cycle. Regardless of their benefits to residents, urban trees are subject to how we reconfigure urban spaces (Avolio et al., 2015). Moreover, invasive species, diseases, and climate change pose additional threats. To address these challenges, a remote sensing approach to tree species identification and health assessment is likely to ease the burden of managing local trees manually.

Today, most urban tree inventories rely on time-intensive visual assessments deployed unevenly and are generally expensive to execute. Remote sensing imagery offers considerable potential as a source of information about urban forests and trees across large landscapes that would be unfeasible to survey on foot. With freely available and low-cost space-based and airborne datasets, remote sensing provides spatially explicit data over large geographic areas that can be—and often is—updated regularly, enabling scientists and stakeholders to monitor large urban areas more comprehensively, which can supplement and augment manual fieldwork. By remotely modeling the urban forests' current state, urban foresters can proactively identify and focus on trees that need immediate attention. In particular, the broadband, multi-spectral QuickBird, and Landsat sensors have led to advances in using spaceborne remote sensing for measuring vegetation cover, including urban tree canopy cover (McPherson et al., 2011) and tree cover change (Hansen et al., 2013).

Airborne LiDAR imagery dramatically enhances our ability to segment individual urban tree species, thanks to its remarkable increase in availability over the past two decades (Alonzo et al., 2013). This technology not only allows for species differentiation, but also furnishes critical crown metrics. It offers precise measurements of tree height, crown area and perimeter, and crown length and width (Lin & Hyyppä, 2016), enabling the assessment of structural characteristics for individual trees across the County's management area. These metrics then serve as powerful tools for informed decision-making in tree management strategies, effectively advancing urban forestry practices. With an inventory of tree-related data, such as location, species, size metrics, and health parameters collected from multiple sources (e.g., digital aerial imagery), local government agencies can better manage their trees for sustainable urban development.

Past urban forestry research has used several discrete return LiDAR sensors for tree identification among structurally unique species, such as palm trees. LiDAR sensors used in these applications have ranged from 2 to 66 points per square meter (Ocón et al., In Review). CAO, Riegl Q560, and Optech sensors were the most common LiDAR sensors. Airborne LiDAR alone identified 30% to 45% of the tree species (Alonzo et al., 2016; Liu et al., 2017; Yang et al., 2019). However, with the further development of LiDAR metrics and improved classification algorithms, it should be possible to improve the LiDAR classification of select native and non-native trees in urban areas. Furthermore, when spectral data and canopy structure from LiDAR are combined, data fusion provides horizontal and vertical information that has potential for improving tree species classifications (Zhang et al., 2018; Zhang et al., 2020).

By using LiDAR for tree segmentation, local government agencies can identify trees that are at risk of diseases or pests and prioritize their treatment accordingly. It also allows them to systematically assess the structural characteristics of individual trees, enabling more efficient scheduling of tree maintenance activities (e.g., pruning) that meet city safety standards and protect public health and welfare (Lin & Hyyppä, 2016). In addition, it gives them the ability to assess the health of trees across an urban landscape and develop appropriate strategies for their preservation. For example, LiDAR can be used to perform canopy change analysis for forest management (Tang et al., 2019) and identify areas with high risk of pests or diseases and prioritize their treatment (Omasa et al., 2006). By collecting these data more efficiently and accurately, local governments can better understand the health of their urban forest and develop effective strategies to protect it.

To address some of these challenges, the Los Angeles County Department of Public Health has teamed up with the Department of Geography at the University of California, Los Angeles, to identify the best approach in individual tree segmentation to support urban tree management. The County manages at least 800,000 public trees where inventory costs using traditional methods exceeded $5 million in 2020 (Los Angeles County Department of Public Health, 2020). Traditional, boots-on-the-ground approaches are also limited to assessments of trees in the public right-of-way. Currently, eight departments within the County have a cost-effective or efficient way to track and manage individual trees. However, those managing the urban forest are stretched thin and often have disparate tree inventory data, making hands-on management challenging to scale across the region. The ability to inventory public and private trees using remote sensing has the potential to provide the County with a faster, cheaper, and more efficient way to inventory trees throughout the region. Accurate and efficient individual tree segmentation can help alleviate the cost and effort burdens felt by Los Angeles County, which I explore further in this chapter. I compare the performance of individual tree segmentation algorithms using airborne discrete-return LiDAR data acquired in December-January of 2015-2016 by the Los Angeles Region Acquisition Consortium (LARIAC), and assess the suitability of applying each algorithm for management needs.

1.1 Advantages of algorithmic individual tree segmentation

Segmenting individual trees using discrete-return LiDAR point clouds provides valuable information for forest management, allowing researchers to measure essential crown metrics, such as crown area, perimeter, length, and width. These metrics are needed to understand critical structural characteristics across species and assist urban forest managers in tracking canopy changes over time (Lin & Hyyppä, 2016). Additionally, individual tree segmentation can provide insight into urban forest structure, including the number of trees, their size distribution, and species composition (Zhang et al., 2015).

Applying automated tree segmentation algorithms to LiDAR point clouds and aerial photography analysis can be much more efficient than manual field work. Algorithms can detect and delineate each tree in an urban canopy, providing a metric for comparing different areas without time-intensive physical measurements or visual assessment of each tree crown. In addition, algorithmic segmentation enables researchers to assess large areas encompassing multiple jurisdictions within a short period relative to manual methods.

1.2 Challenges Posed by urban environments

Urban environments are particularly challenging for automated individual tree segmentation algorithms due to their complex structure, such as the presence of buildings, roads, power lines, and high species richness. Furthermore, the varying densities of trees in urban forests further complicate the process – some trees have large open canopies, while others may be obscured under taller trees or other obstructions (Lin & Hyyppä, 2016). In addition, the size of individual tree crowns can vary significantly among species, which creates noise in the LiDAR data used by segmentation algorithms. Finally, noise from other sources often confounds algorithm performance (Zhang et al., 2015).

Additionally, many automated approaches to individual tree segmentation depend on species-specific crown metrics that quantitatively define the structure or shape of the tree crown. These metrics include the ratio of crown length to tree height and crown width to tree height, which can deviate severely from a normal distribution per species if tree pruning is present. This makes it challenging to set generic thresholds for segmentation, and in some cases, manual adjustment may be required (Liu et al., 2017).

1.3 Objectives

Those currently managing urban forests are stretched thin and often have disparate tree inventory data, making hands-on management challenging to scale across the region. I identify an approach best suited for urban forest segmentation with broader implications for urban tree management. In this chapter, I conduct a comparative analysis of five individual tree segmentation algorithms using aerial LiDAR point clouds in urban Los Angeles County. I identified five individual tree segmentation algorithms suitable for this study. These include: 'Find Trees' (Roussel et al., 2020), 'Watershed' (Roussel et al., 2020), 'Silva2016' (Silva et al. 2016), 'Dalponte2016' (Dalponte et al., 2016), 'Adaptive Mean Shift' (Ferraz et al., 2016), and 'Li2012' (Li et al., 2012). The performance of these algorithms was qualitatively assessed using ground-truth observations and compared to airborne imagery from the National Agriculture Image Program (NAIP) and high-resolution satellite imagery in Google Earth. This analysis will help assess the suitability of applying each algorithm for management needs and inform on the efficacy of automated methods for segmenting individual trees in urban L. A. County.

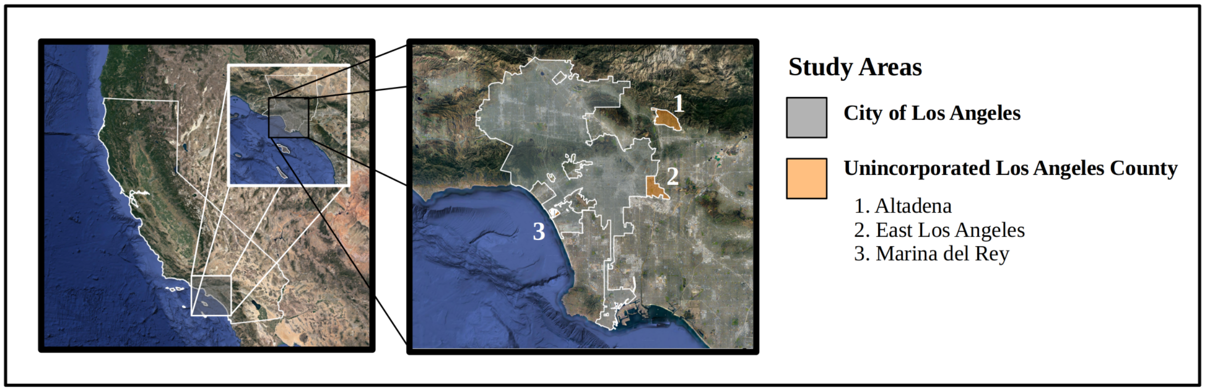
2 Methods

By comparing different approaches, this chapter aims to help users make an informed decision when choosing the most suitable individual tree segmentation algorithm for their urban data. I discuss the advantages and disadvantages of each approach and their suitability for different scenarios in urban forestry. In addition, I provide practical examples and recommendations on which type of algorithm is best suited for a given context.

2.1 Study Area

Three unincorporated areas within central Los Angeles County were chosen as project pilot sites (Fig. 1). These locations were chosen as representative examples of areas managed by the County. These areas were also chosen because they represent diverse socioeconomic backgrounds, tree species, land-use patterns, and management approaches. The three unincorporated areas, include Altadena (1), which sits on the southwestern foothills of the San Gabriel Mountains; East Los Angeles (2), which is located east of downtown Los Angeles and is surrounded by three of the busiest freeways in the region; and the Marina (3), located on the coast just north of the LAX international Airport (Fig 1).

Figure 1. Three unincorporated pilot sites in relation to the City of Los Angeles.



2.2 Datasets

I compared five individual tree segmentation algorithms using aerial point clouds in urban Los Angeles County. The five methods can be divided into point cloud-based and raster-based, each with two subfamilies - algorithms that work in two steps (Individual tree detection followed by segmentation) and all-in-one algorithms.

2.2.1 Los Angeles Region Image Acquisition Consortium

The LARIAC dataset is proprietary aerial imagery commissioned by Los Angeles County in contract with the Pictometry International Corporation. Flights for the County have been flown every 2 to 4 years starting in 2002. The most recent flight was in 2020 using the EagleView-6 camera sensor suite. The full dataset covers fine-resolution multispectral (three- and four-band) imagery and discrete-return LiDAR point clouds. The four-band multispectral Red-Green-Blue

(RGB) and near-infrared (NIR) imagery are captured at a 10 cm spatial resolution for urban areas and a 30 cm spatial resolution for national forest within Los Angeles County.

Additionally, oblique imagery is provided at the same resolutions for the urban and community scales. The LiDAR imagery has a spatial resolution of 10.2 cm with vertical accuracy of 27.7 cm at a 95% confidence level and was used to derive several structural datasets, including building footprints, a digital elevation model, and a vegetation height dataset (LARIAC, 2016). Derived datasets from the multispectral imagery include the Normalized Difference Vegetation Index (NDVI), land cover, and tree canopy cover. The NDVI dataset was calculated using the 2006 imagery. The land cover dataset has eight land cover classes that were created using an object-oriented image analysis software called eCognition (Locke et al., 2017). Data used to create the land cover layer included 2014 LARIAC, NAIP imagery, 2016 LiDAR, CAMS data, 2014 building outlines, and land type data. The software was first trained to recognize the different land cover classes and then automated the feature extraction process. The reported accuracy for the tree class feature is 95% (Locke et al., 2017). The advantage of using very fine multispectral data is higher accuracy when performing land cover classification. With imagery at 10 cm spatial resolution, almost all land cover labels of interest can be classified. However, the issue of distinguishing between grass, shrubs, and trees remains, which multispectral data can be ill-suited to address. LiDAR data has an advantage by providing structural information and three-dimensionality to multispectral images. These structural parameters can distinguish between different types of vegetation and other classes, like buildings, and help researchers derive tree canopy layers that are highly accurate.

2.2.2 Ground-truth data

Georeferenced tree locations were collected across all four study areas and served as ground truth data to evaluate the performance of all five algorithms. The data were collected by contracted arborists by Los Angeles County departments (Los Angeles County Public Works, Los Angeles County Beaches and Harbors, Los Angeles County Parks and Recreation). A total of 16,947 tree point locations were collected from 11 target genera and species within our three study sites, which provided sufficient information to evaluate the accuracy of each algorithm. 423 tree point locations were logged in 2013, three in 2014, six in 2015, one in 2016, three in 2017, 14 in 2018, 562 in 2109, 10 in 2020, and 9,468 in 2021. The remaining trees do not have a collection date associated with the tree point data. The data were imported into a Geographic Information System (GIS) software program (state software used ARC or QGIS) for further analysis. The compiled dataset was used as the reference dataset to compare with the results generated by each algorithm evaluated in this study. Field crews used a combination of Global Positioning System (GPS) coordinates, canopy cover measurements, and species identification to verify each point as a distinct tree species or belonging to a specific genera if identification to species was impossible. These 11 target genera and species were chosen based on regional conservation goals related to California native trees (*Juglans californica*, *Platanus racemosa*, *Quercus agrifolia*, and *Sequoia sempervirens*), culturally significant species (*Citrus* spp. and *Washingtonia* spp.), hazard species unsuited to L.A.'s changing climate (*Eucalyptus* spp. and *Pinus* spp.), species responsible for infrastructure damage (*Ficus* spp.), and species identified as suitable for the region's need for drought tolerance (*Lagerstroemia* spp., and *Jacaranda mimosifolia*) (Table 1).

Table 1. Breakdown of target species and genera with a corresponding rationale for Los Angeles County interest and total number of ground points.

Species	Target Group	Reasoning	Altadena Support	East Los Angeles Support	Marina Support	Total Support
Citrus limon	*Citrus* spp.	An agricultural and historical Genera to Los Angeles.	113	-	-	**113**
Citrus sinensis			83	-	-	**83**
Eucalyptus citriodora	*Eucalyptus* spp.	Genera recognized regionally as a hazard.	-	-	132	**132**
Eucalyptus cladocalyx			99	-	-	**99**
Eucalyptus globulus			77	-	146	**223**
Eucalyptus polyanthemos			113	-	-	**113**
Ficus benjamina	*Ficus spp.*	Genera is responsible for high rates of infrastructure damage to sidewalks and pipes.	91	247	-	**338**
Ficus carica			105	-	-	**105**
Ficus microcarpa nitida			1,982	214	-	**2,196**
Ficus rubiginosa			-	-	45	**45**
Jacaranda mimosifolia	*Jacaranda mimosifolia*	Preferred street tree, and measuring the phenology may improve the model.	432	375	-	**807**
Lagerstroemia indica	*Lagerstroemia indica*	Preferred street tree in regards to appearance, maintenance, and water stress.	633	207	-	**840**
Pinus canariensis	*Pinus spp.*	Genera recognized as a hazard species	561	150	-	**711**

Pinus contorta		that will not respond well to climate change.	135	-	-	**135**
Pinus coulteri			46	-	-	**46**
Pinus halepensis			391	66	-	**457**
Pinus pinea			42	-	-	**42**
Platanus racemosa	*Platanus racemosa*	Drought intolerant species requiring high water use, and susceptible to shot hole borer.	171	222	-	**393**
Quercus agrifolia	*Quercus agrifolia*	Native, protected species. Symbolized species of Los Angeles biodiversity.	4,046	36	-	**4,082**
Sequoia sempervirens	*Sequoia sempervirens*	Native California species, Recognized as a hazard tree with high mortality rates.	89	-	-	**89**
Washingtonia filifera	*Washingtonia spp.*	An iconic and symbolized Genera for Los Angeles culture. These are in the process of dying out regionally.	255	-	-	**255**
Washingtonia robusta			2,596	313	468	**3,377**
Non Target Species	Non Target Species	To introduce a variety of species not chosen as target species to improve model performance.	742	1,032	492	**2,266**
Total Support	**-**	**-**	**12,802**	**2,862**	**1,283**	**16,947**

2.3 Individual tree segmentation algorithms

Individual tree segmentation is an essential part of urban forest inventory and management, providing helpful information on the health of urban ecosystems. The five individual tree segmentation algorithms applied to aerial point clouds in urban L. A. County in this study can be divided into two broad categories: point cloud-based algorithms, which do not require a canopy height model (CHM), and raster-based algorithms, which use a CHM. To compute a CHM, I normalized the raw point cloud by setting ground points (classified as such in the original dataset provided by LARIAC) to an elevation of zero. This recalculates all heights for each point relative to the ground, and makes noise classification easier (Roussel et al., 2020; Roussel et al., 2023). I then filtered for all above ground points (removing noisy points with an elevation less than zero), and subsetted outliers with an elevation greater than 95% of Z values for the remaining points. Lastly, I implemented a point-to-raster algorithm to calculate a CHM for each pilot site (Fig. 2-4).

Figure 2. Canopy Height Model for Altadena, CA at a 1 m spatial resolution and processed using a point-to-raster, pitfree algorithm in R with LARIAC 4 discrete-return point-clouds collected in December, 2015 to January, 2016.

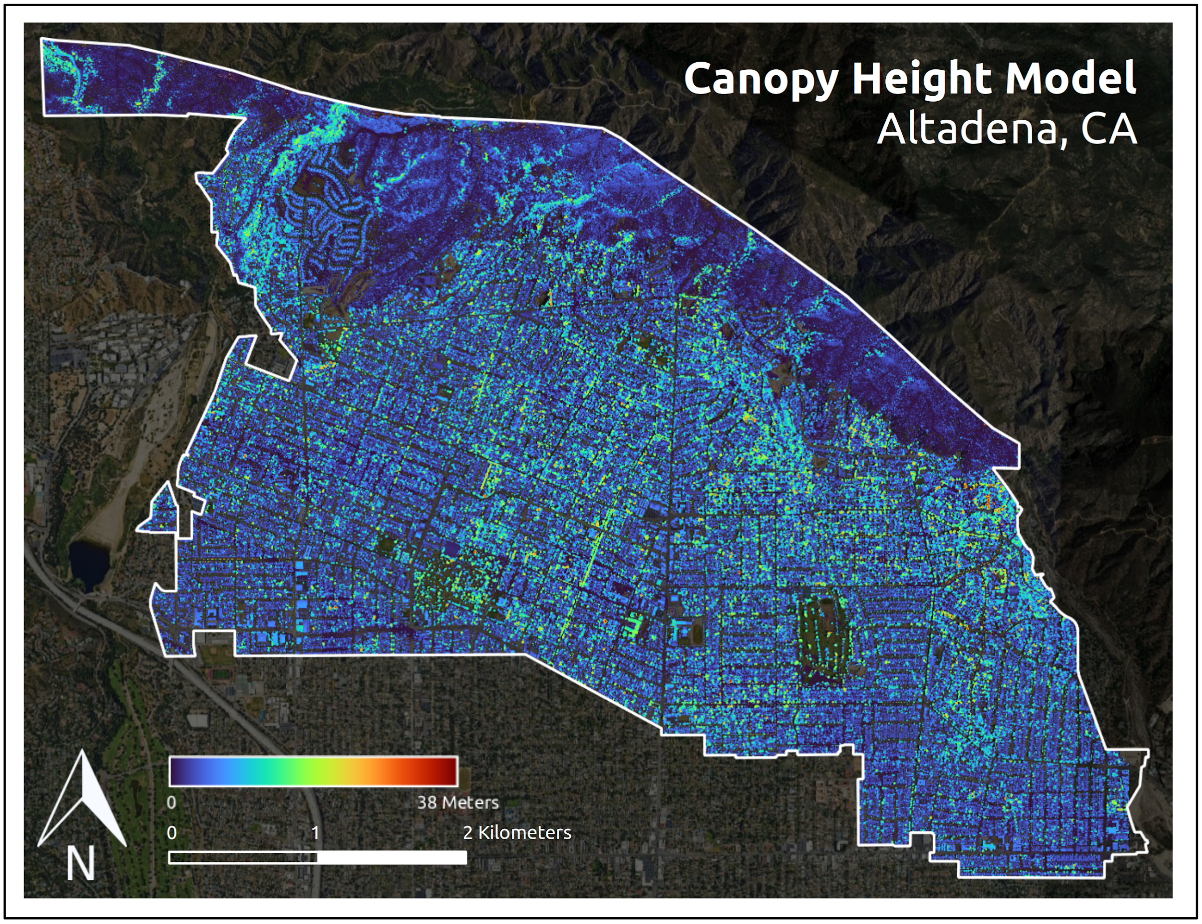


Figure 3. Canopy Height Model for East Los Angeles, CA at a 1 m spatial resolution and processed using a point-to-raster, pitfree algorithm in R with LARIAC 4 discrete-return point-clouds collected in December, 2015 to January, 2016.

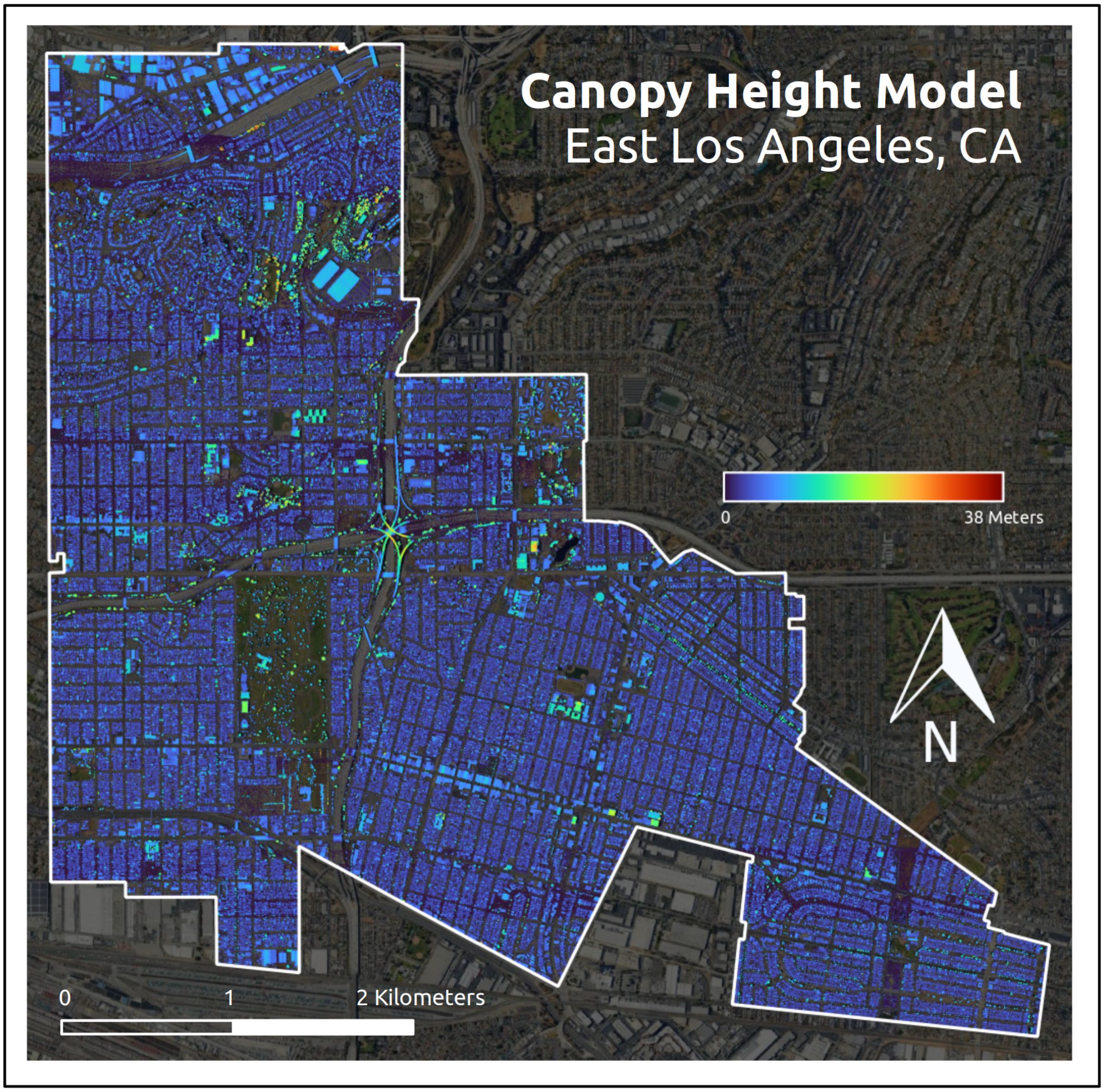
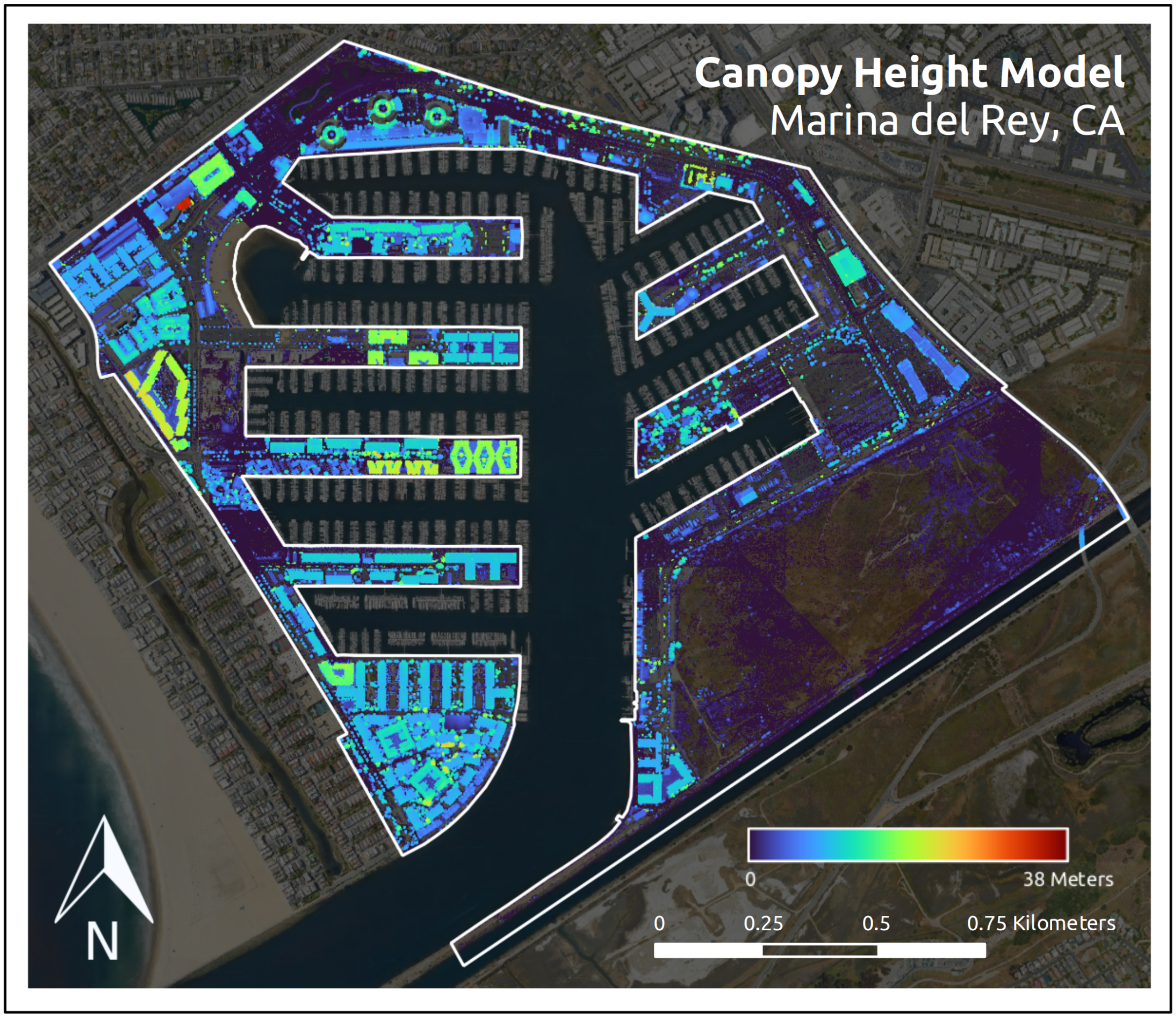
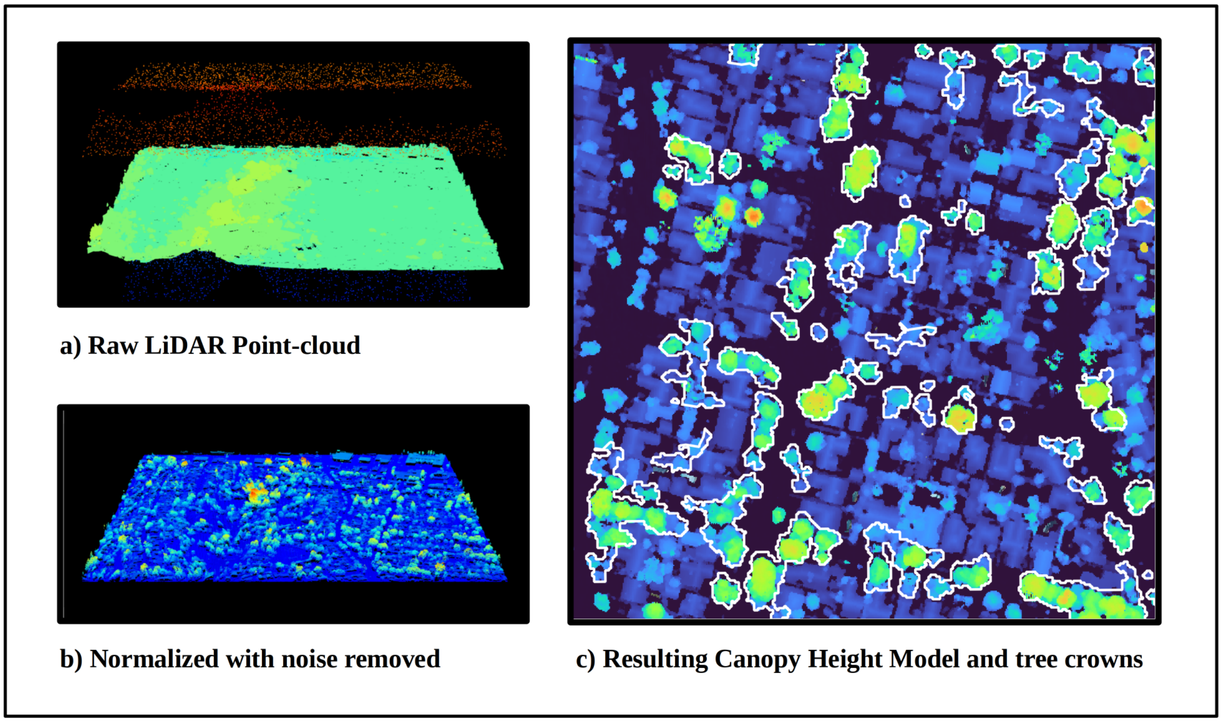


Figure 4. Canopy Height Model for Marina del Rey, CA at a 1 m spatial resolution and processed using a point-to-raster, pitfree algorithm in R with LARIAC 4 discrete-return point-clouds collected in December, 2015 to January, 2016.



There are a number of methods to calculate a CHM including subtracting a digital surface model from a digital terrain model, which are common datasets included with LiDAR data. However, LARIAC does not provide either dataset, and by normalizing my data, I was able to implement the point-to-raster method instead, which rasterizes point Z values by mapping to a user defined grid (Roussel et al., 2020; Roussel et al., 2023). I chose a grid resolution of 0.5 m and a triangular irregular network algorithm to compute an elevation value for each pixel in the grid. I also applied a pitfree algorithm as suggested by Roussel et al. (2023) to avoid creating empty pixel values during the triangulation (Fig. 5). Within these categories, the algorithms are further divided into two subgroups: those that work in two steps—individual tree detection followed by segmentation, and all-in-one algorithms which complete both processes in a single execution (Roussel et al., 2020; Roussel et al., 2023). In order to determine the relative effectiveness of each algorithm, performance metrics such as accuracy, precision, recall, and computation time were employed to evaluate their relative merits for use in urban forest management. Four of the five algorithms tested include the LiDAR processing R package 'lidR' (Roussel et al., 2020; Roussel et al., 2023, R Core Team, 2023). The Adaptive Mean Shift algorithm (Ferraz et al., 2016) was instead deployed using a forked version of the 'MeanShiftR' R package (Knapp, 2023, R Core Team, 2023).

Figure 5. LiDAR point-cloud processing needed to produce a CHM. a) Visualization of raw point cloud data, followed by b) normalization (setting the ground to '0'), and c) rasterization (2-D) representation of the 3-D point cloud data.



2.3.1 Point cloud-based algorithms

Point cloud-based algorithms are a popular approach for individual tree segmentation due to their ability to process large volumes of data quickly. While these algorithms can be computationally expensive, they offer the potential for high accuracy and resolution.

1. The Adaptive Mean Shift algorithm (Ferraz et al., 2016) uses an iterative approach to segment individual trees from LiDAR point clouds. The algorithm begins by choosing an initial seed point that identifies the tree's center. From this point, it then calculates the probability profile of each neighboring pixel that could be part of the tree and segments them sequentially based on their probability values. When complete, the algorithm has

successfully segmented all points within the LiDAR data that belong to one individual tree. The process begins with a region-based search for candidate pixels most likely part of the given tree. This is done by first computing a set of weights associated with each pixel, where higher weights indicate higher probabilities that it belongs to the same tree being segmented. Then, high-probability pixels are added to a list of valid candidate pixels. Pertinent parameters for determining these weights include the ratio of tree crown length to tree height, as well as tree crown width to tree height. Once a suitable clustering is obtained, all pixels belonging to each cluster are assigned a unique label representing its identity as belonging to one specific tree and stored within an image mask to assist subsequent image processing operations such as multiple object tracking and branch extraction from individual trees. Finally, all points belonging to each identified cluster can be filtered out from an original cloud data set to obtain only points constituting individual trees that now reside in separate containers for further analysis and interpretation. Parameter fine-tuning for Adaptive Mean Shift can be found in Appendices D and E.

2. The Li2012 algorithm is an object-based tree segmentation algorithm based on the work of Li et al. (2012). It works by creating objects representing a group of points with similar local features and then merging them into segments. This algorithm is beneficial when dealing with data collected from terrestrial-based LiDAR scanners, as it can create segments with fewer gaps than other methods. However, this algorithm is known to be computationally time intensive. It has an algorithmic complexity worse than $O(n^2)$, meaning there is exponential growth in processing time associated with this approach (Roussel et al., 2020).

2.3.2 Point cloud-derived raster-based algorithms

Point cloud-derived raster-based algorithms combine traditional aerial imagery and point clouds to produce high-resolution results. These algorithms are practical for large datasets, as they are less computationally expensive and quicker to execute. However, the accuracy of these algorithms can be affected by the resolution and quality of the aerial imagery used.

1. The Dalponte2016 algorithm is a tree segmentation approach developed by Dalponte and Coomes (2016). This method uses smooth clustering, or "superpixeling," to create data segments from airborne LiDAR scanners. The superpixels process creates small uniform data patches that can then be merged into more significant segments representing individual trees. It offers a balanced trade-off between speed and accuracy, making it an ideal choice for urban forest mapping applications in Los Angeles. This method has been found to produce reliable results when used with airborne data from LiDAR sensors (Roussel et al., 2020). The inputs for the Dalponte algorithm include a GIS layer of tree tops and a CHM. I calculated tree top locations programmatically using the 'lidR' R package by designing a function that computes a window size as a function of pixel (CHM) height (Roussel et al., 2020).

2. The Silva2016 (Silva et al., 2016) algorithm is based on adaptive region growth. It uses a two-dimensional surface fitting method to identify potential trees within a point cloud and then classifies individual segments based on their characteristics. This algorithm is more accurate than other tree segmentation methods, mainly when dealing with large, complex datasets. Additionally, it has been shown to work better with high noise and outliers.

3. The Watershed function (Roussel et al., 2020) is a tree segmentation algorithm that uses a watershed-based approach to group similar points together. It divides the point cloud into catchment basins and then classifies individual trees based on the boundaries between them. This algorithm is more accurate than other tree segmentation algorithms, as it can separate closely grouped trees better than the traditional growing region methods (Roussel et al., 2020; Roussel et al., 2023). Another version of this algorithm is the Marker-Controlled Watershed function that combines the watershed-based approach with marker-controlled region growth. This allows it to identify discontinuities between segments and filter out noise not part of any tree structure.

2.4 Conducting a visual evaluation of algorithmic performance

A visual evaluation of the tree segmentation algorithms can be conducted manually (Murtha and Fournier, 2014; Yang et al., 2022). This evaluation allows for an accuracy assessment by comparing the generated segmentation results with reference data or aerial imagery. I conducted a visual evaluation by preparing a dataset to use for comparison. This dataset included high-resolution aerial imagery or reference data representing ground truth data from the study area. Additionally, the dataset covered a range of urban settings to comprehensively assess the algorithms' performance in different environments. The selected study areas included different tree species, densities, canopy structures, and proximity to other urban features such as buildings and roads.

Each algorithm generated individual tree segments for the second step using the same input data. I ensured that the optimal parameters for each algorithm were used to achieve the best possible results. The generated outputs were georeferenced and overlaid on the ground points, as well as temporally related (within six months) aerial imagery (NAIP and Google Satellite Imagery) for quick comparison. To evaluate how well the algorithms segmented 16,947 ground points, I randomly selected a single LARIAC tile from the pilot sites, then randomly selected 50 segmented tree crowns from the results. The results were then compared using the following criteria:

- Correct segmentation: The percent of individual trees accurately segmented and separated from adjacent trees and other urban features.
- Under-segmentation: The percent of instances where multiple trees are grouped as one segment causes an underestimation of the total number of trees.
- Over-segmentation: The percent of instances where a single tree is divided into multiple segments, leading to an overestimation of the total number of trees.
- Misclassification: The percent of instances where non-tree objects, such as buildings or vehicles, are mistakenly identified as trees.

This qualitative assessment allows for a more nuanced understanding of each algorithm's challenges and complexities in different urban settings, ultimately leading to better decision-making when selecting the most suitable algorithm for a particular project.

3 Results

The comparative analysis highlighted differences between the performance levels of the five algorithms tested. It was found that point cloud-based algorithms produced more accurate results (range of 10% in accuracy) than raster-based algorithms (range of 30%) (Table 2). However, point cloud algorithms took significantly longer to produce a result than raster-based algorithms with point cloud-based algorithms needing up to 90 minutes for tiles with the presence of dense canopy (45-60 minutes for tiles with less dense or open canopy) versus ~30 seconds for raster-based approaches regardless of canopy density.

Table 2. Comparing the performance between Adaptive Mean Shift (AMS3D), Dalponte2016, Silva2016, Watershed, and Li2012 on 50 randomly selected tree crowns in Altadena. The cells with the best performing metric for each category are highlighted in dark green (second best in light green), and the worst performing in dark red (second worst in light red).

Criteria	Dalponte2016	AMS3D	Silva2016	Watershed	Li2012
Correct	0.50	0.42	0.20	0.40	0.32
Under	0.12	0.28	0.42	0.30	0.58
Over	0.38	0.28	0.14	0.10	0.00
Misclassified	0.00	0.04	0.24	0.20	0.10

Overall, the analysis showed that point cloud-based algorithms are better suited for applications where detailed accuracy is needed. In contrast, raster-based algorithms are more suitable for applications requiring a rapid result, with Dalponte2016 providing both a rapid result and the best overall accuracy of 50% correctly segmented trees. The Watershed (30%) and Silva2016 (42%) algorithms had an unfortunate propensity for both under-segmentation of the target species at a rate much higher than that of Dalponte2016 (12%). As for Li2012, it failed to execute the task in a timely manner after each attempt, making its slow processing speed a clear issue, as well as having the worst performing under-segmentation of all five algorithms (58%), but ranked best in over-segmentation (0%). In addition, further research is needed to fully assess the performance of these algorithms in different urban environments around the world. This will help ensure they can be used effectively and accurately in all urban forestry management scenarios.

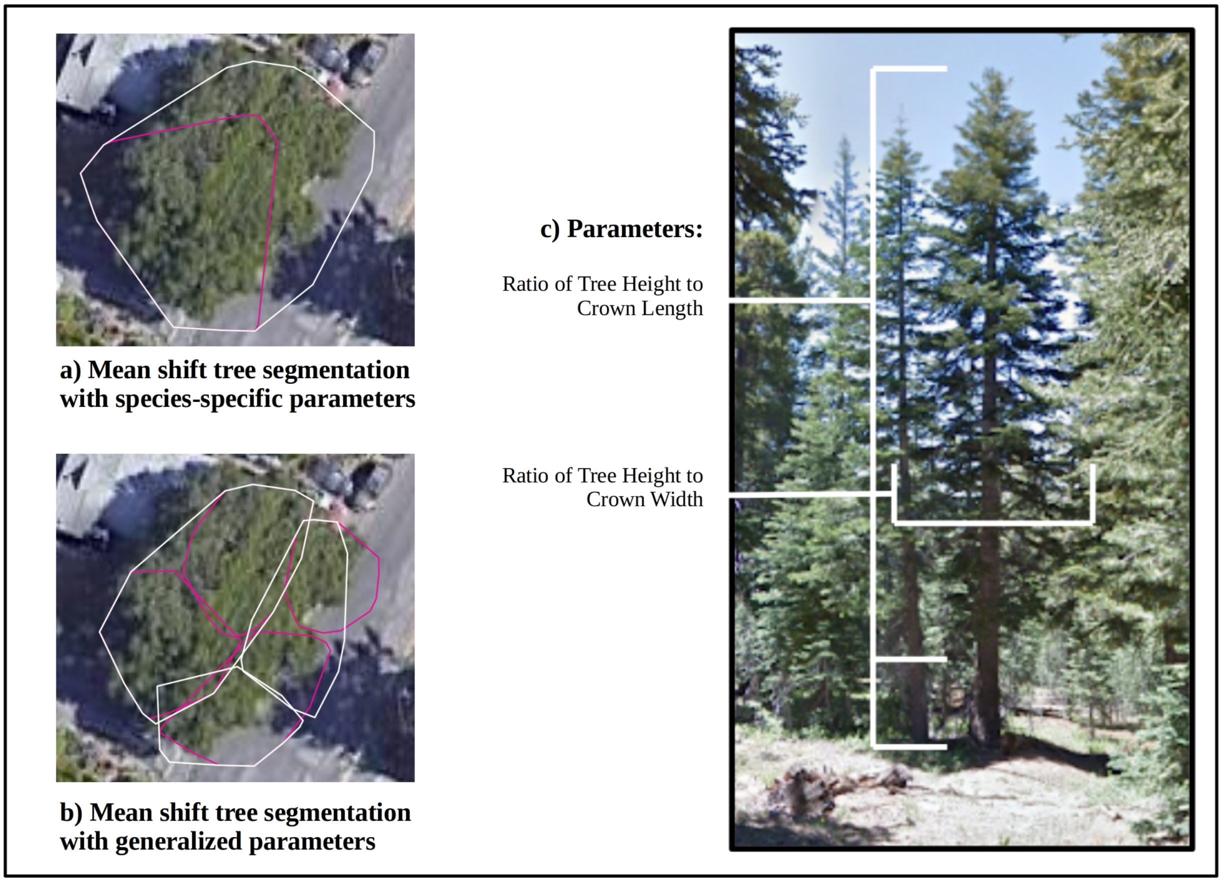
3.1 Adaptive mean shift

Overall, Adaptive Mean Shift performed well when given species-specific parameters describing crown shape and demonstrated its potential as a powerful tool for segmenting individual trees from LiDAR point clouds. After visually assessing the algorithm's performance across 50 randomly selected crowns, Adaptive Mean Shift correctly segmented 42% of the crowns (Fig. 8-10). However, there were also limitations associated with applying the algorithm to heterogeneous sample sites–namely, poor performance when trying to generalize across structurally dissimilar species—and the algorithm was not tested for an all-too-common structural change for many street trees, which is pruning. Adaptive Mean Shift under and over segmented tree crowns by 28% each, and also misclassified 4% of crowns. Additionally, the Adaptive Mean Shift algorithm is computationally expensive, taking upwards of one hour to run on a single, square LARIAC tile with a length of two-fifths of a kilometer (quarter-mile—raw data is provided individually as quarter-mile tiles projected in 'NAD83 / California zone 5' with units in U.S. feet).

With over 4,000 tiles in Los Angeles County, this method would take 166 days to complete using a single-core computer with 16-GB of RAM. The number of days needed to process reduces significantly with the use of high performance computing, but increasing the number of computer cores and memory to process the data is not a linear relationship. The amount of data in need of processing for our area would take an estimated two weeks (13-15 days) to run using a 64-core machine and over 32-GB of RAM. Moreover, this is the minimum amount of processing needed to segment one structural type among the dozen or so found across the project's target species.

This approach is best suited for application in structurally homogeneous forests, or if the target species spanned a small number of structural types. For the purposes of the County's urban forest program, implementing Adaptive Mean Shift becomes counterproductive when trying to determine a workflow that reduces the time and effort to inventory individual trees across the study area (Fig. 6).

Figure 6. Visualization of the performance metrics of the Adaptive Mean Shift tree segmentation algorithm. a) General performance when species-specific crown shape parameters are used to delineate an individual *Quercus agrifolia.* b) Performance depicting individual tree segmentation when generalized crown shape parameters are used for *Quercus agrifolia*. c) Visualization of the crown shape parameters, ratio of tree height to crown length and width, used in the Adaptive Mean Shift algorithm.



3.2 Canopy height model and ground points

The raster-based Dalponte2016 algorithm was the second best performing algorithm for tree segmentation in this study despite correctly segmenting more tree crowns than Adaptive Mean Shift (50% versus 42%) and having no instances of misclassification (0% versus 4%). It is considered second best due to the high instances of over classification (38%) compared to Adaptive Mean Shift (28%) (Table 2).

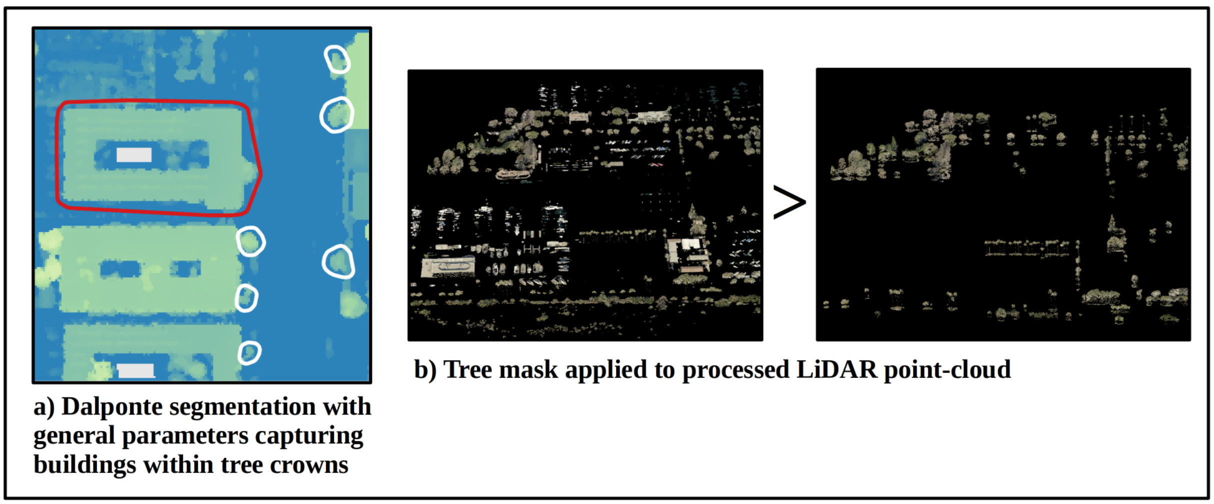
Unlike the Adaptive Mean Shift algorithm, which required an individual calibration on each target species, Dalponte2016 provided a generalized set of parameters that were suitable for all target species. However, over segmentation leads to dissimilar tree species in closed canopy being included in a single tree crown, which can throw off crown metrics and overestimate the crown area, perimeter, and height of any given species (Table 3).

Table 3. Comparison of crown metrics between Dalponte2016 and Adaptive Mean Shift algorithm (AMS3D) in Altadena.

Species	Mean Height (m) AMS3D	Mean Height (m) Dalponte	Mean Points (#) AMS3D	Mean Points (#) Dalponte	Mean Area (m^2) AMS3D	Mean Area (m^2) Dalponte
Citrus limon	11.67	12.94	527	1,114	61.32	103.18
Eucalyptus cladocalyx	16.43	16.74	833	1,238	99.98	131.32
Eucalyptus globulus	15.78	16.11	787	830	95.31	72.76
Ficus benjamina	10.91	10.56	322	425	55.13	46.20
Ficus carica	14.09	16.22	601	853	98.63	73.48
Ficus microcarpa nitida	10.97	10.99	432	469	62.98	56.98
Jacaranda mimosifolia	10.77	10.91	404	551	60.39	73.76
Lagerstroemia indica	8.60	8.87	268	370	41.86	49.69
Pinus canariensis	21.21	22.92	770	907	118.79	115.02
Pinus contorta	10.60	10.81	430	520	56.19	66.74
Pinus coulteri	19.23	20.01	1,827	2,546	144.3	165.79
Pinus halepensis	15.82	17.19	828	1,263	101.84	125.25
Pinus pinea	13.20	13.30	1,113	1,377	137.56	157.39
Platanus racemosa	13.15	14.93	753	1,084	92.78	111.19
Quercus agrifolia	11.96	12.78	523	766	82.36	90.72
Sequoia sempervirens	13.98	15.38	347	518	52.94	59.66
Washingtonia filifera	15.87	17.07	359	434	44.67	48.45
Washingtonia robusta	18.22	21.52	415	323	58.24	40.78

Furthermore, misclassification was initially a limitation for all algorithms when trees were near large structures such as buildings and towers. Fortunately, this could be mitigated by applying a canopy mask to remove any points from the LiDAR data not overlapping with tree canopy (Fig. 7). These adjustments improved overall accuracy and allowed for more precise segmentation of all target species. I also map Dalponte2016 tree crowns for all three study sites (Fig. 8-10).

Figure 7. Visualization of erroneous tree segmentation using Dalponte2016. a) Misclassification, and b) Applying a tree mask to point clouds to avoid misclassifications.



To provide an overview of the tree crown metrics from this study's target species, Table 4 was created utilizing the Dalponte2016 approach. This table offers detailed measurements on various aspects, including the number of crowns (Count), average number of points per species in the LiDAR data (Mean Points), as well as the average tree height (Mean Height), crown area (Mean Area) and perimeter (Mean Perimeter). Figures 8 – 10 visualize the tree crowns produced by Dalponte2016 for all public street trees across all three pilot sites.

Table 4. Aggregate tree crown statistics utilizing the Dalponte2016 approach based on the target species/genera.

Altadena	Mean Height (m)	Mean Area (m^2)	Mean Perimeter (m)	Mean Points (#)	Count
Citrus limon	12.94	103.18	42.37	1114	113
Citrus sinensis	13.73	78.22	34.83	946	83
Eucalyptus cladocalyx	16.74	131.32	55.51	1238	99
Eucalyptus globulus	16.11	72.76	34.42	830	77
Eucalyptus polyanthemos	12.11	28.72	19.64	382	113
Ficus benjamina	10.56	46.20	29.20	425	91
Ficus carica	16.22	73.48	29.60	853	105
Ficus microcarpa/nitida	10.99	56.98	32.54	469	1982
Jacaranda mimosifolia	10.91	73.76	38.43	551	432
Lagerstroemia indica	8.87	49.69	29.26	370	633
Pinus canariensis	22.92	115.02	45.86	907	561
Pinus contorta	10.81	66.74	36.72	520	135
Pinus coulteri	20.01	165.79	54.03	2546	46
Pinus halepensis	17.19	125.25	46.64	1263	391
Pinus pinea	13.30	157.39	53.04	1377	42
Platanus racemosa	14.93	111.19	46.03	1084	171
Quercus agrifolia	12.78	90.72	38.02	766	4046
Sequoia sempervirens	15.38	59.66	31.35	518	89
Washingtonia filifera	17.07	48.45	29.41	434	255
Washingtonia robusta	21.52	40.78	25.79	323	2596
East Los Angeles	**Mean Height (m)**	**Mean Area (m^2)**	**Mean Perimeter (m)**	**Mean Points (#)**	**Count**
Ficus benjamina	8.62	53.34	31.56	262	247
Ficus	9.09	55.81	32.33	265	214

microcarpa/nitida					
Jacaranda mimosifolia	8.62	48.19	31.02	238	375
Lagerstroemia indica	6.59	26.14	21.91	108	207
Pinus canariensis	20.03	63.92	34.69	345	150
Pinus halepensis	14.95	86.47	41.99	409	66
Platanus racemosa	11.34	42.61	29.70	231	222
Quercus agrifolia	9.69	58.11	33.95	229	36
Washingtonia robusta	14.26	20.42	20.51	100	313
Marina del Rey	**Mean Height (m)**	**Mean Area (m^2)**	**Mean Perimeter (m)**	**Mean Points (#)**	**Count**
Eucalyptus citriodora	17.09	94.49	41.73	1408	132
Eucalyptus globulus	18.23	167.07	55.83	2778	146
Ficus rubiginosa	10.16	79.3	35.46	1385	45
Washingtonia robusta	16.19	19.45	17.09	351	468

Figure 8. Individual tree crowns for Altadena, CA processed using the raster-based Dalponte2016 tree crown delineation algorithm with CHM inputs derived from LARIAC 4 LiDAR data.

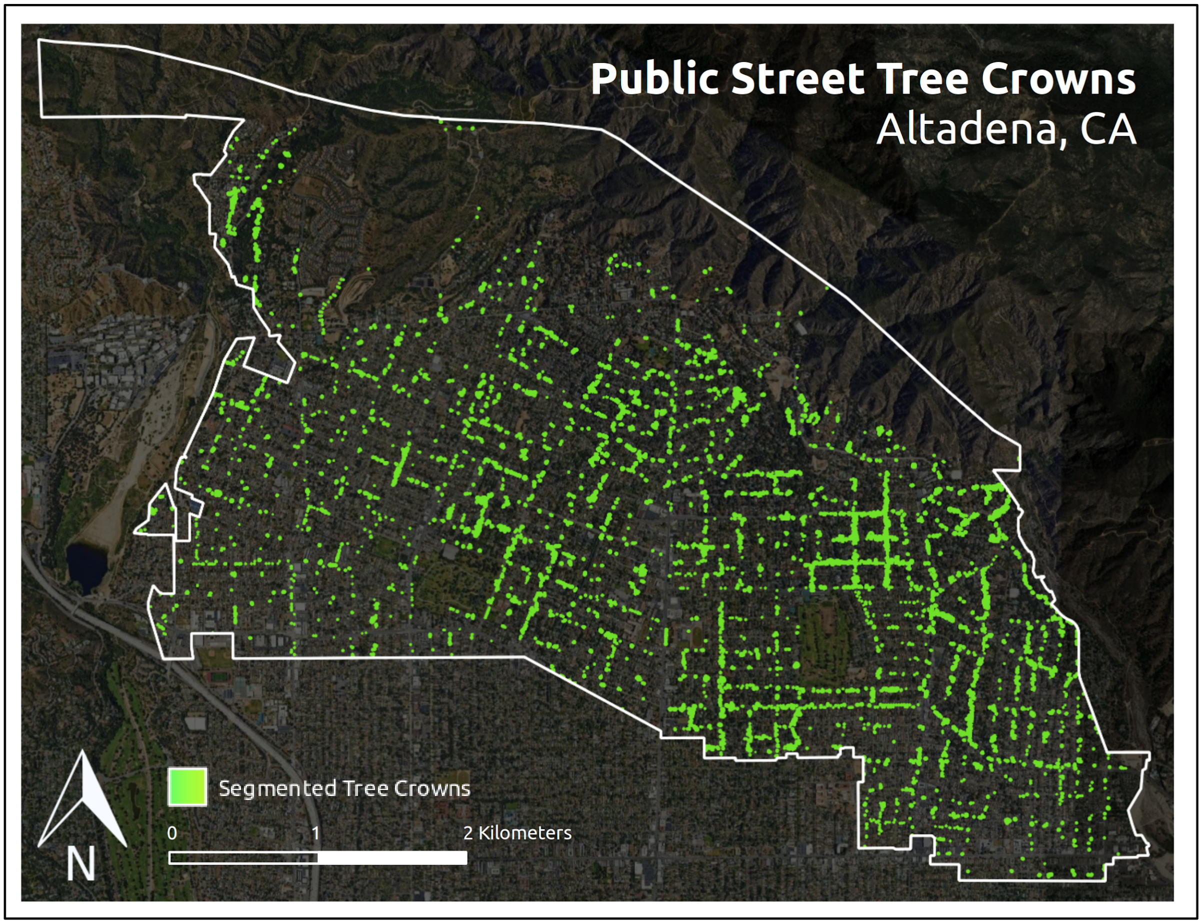


Figure 9. Individual tree crowns for East Los Angeles, CA processed using the raster-based Dalponte2016 tree crown delineation algorithm with CHM inputs derived from LARIAC 4 LiDAR data.

Figure 10. Individual tree crowns for Marina del Rey, CA processed using the raster-based Dalponte2016 tree crown delineation algorithm with CHM inputs derived from LARIAC 4 LiDAR data.

4 Discussion

While trees are integral to urban infrastructure, offering cooling effects, air and water purification, and enhancing community well-being (Mullaney et al., 2015), their effective integration requires navigating challenges such as equitable distribution and misallocation of ecosystem services (i.e., planting *L. indica* or *W. robusta* for shade purposes) (Roman et al., 2021). Moreover, the benefits can vary significantly based on regional climates and socio-economic or socio-ecological contexts.. Therefore, accurate tree segmentation is a vital component of urban forest inventory and management, as it enables the monitoring of individual trees and the overall health of urban ecosystems. This section discusses the importance of implementing accurate tree segmentation for urban forest inventory and its potential application in species classification models. Specifically, accurate tree segmentation allows for the identification and monitoring of individual trees, facilitating the assessment of tree health and identifying individual trees to species within an urban environment, which provides valuable information about tree species composition and distribution and can guide conservation efforts for California natives and inform urban planning decisions to enhance and protect biodiversity by using urban areas as a conservation space. By identifying individual trees and their locations, accurate tree segmentation supports the integration of green infrastructure into urban planning processes. As a result, urban forests can be strategically managed to maximize benefits such as heat island reduction by identifying heat-prone neighborhoods with the necessary sidewalk space fit for planting large canopy trees, or implementing a row of trees between apartment buildings and freeways or arterial roads for noise pollution mitigation.

In this study, I compared the application of five tree segmentation algorithms and their performance in three locations in urban Los Angeles County. Overall, Dalponte2016 (50%) and AMS3D (42%) outperformed Silva2016 (20%), Watershed (40%), and Li2012 (32%) in correctly segmenting trees, and also had the least amount of under-segmentation (12% and 28%, respectively) and misclassification (0% and 4%, respectively) in the group. Silva2016 was the worst performing algorithm tested, and Watershed (10%) and Li2012 (0%) provided promising over-segmentation results, but underperformed elsewhere.

My comparison of canopy metrics derived from both Dalponte2016 and AMS3D produced expected results for most of our target species, but did have some unexpected results, especially concerning *Citrus* spp. (Fig. 11). One of the parameters set for all algorithms was specifying a minimum tree height of 4 m, or right around the maximum growth of mature *Citrus* spp. trees. Although open canopy individuals were correctly measured with tree heights between 4 and 9 m, individuals in closed canopy were instead under-segmented and assigned the height of the adjacent tree. This was especially true for *Citrus* spp. individuals less than 6 m in height, but taller individuals were successfully segmented in closed canopy (Fig. 11).

Figure 11. a) Google Street View of an example of *Citrus limon* under-segmentation in Altadena. With a minimum tree height parameter of 4 m, under-segmentation can occur for smaller (~4 m) understory trees. b) Google Street View of an example of *Citrus limon* correct segmentation in Altadena. A larger understory tree, in this case 9.66 m, may be correctly segmented since it deviates more from the minimum tree height of 4 m.

Comparing my results with the literature for each algorithm, I found agreement between their studies and mine. Ferraz et al. (2016) lacked ground measurements to accurately quantify their model's performance, but were able to compare the modeled crown widths and crown depths to show that the widths were slightly larger than the depths, which agreed with field measurements carried about by a separate study in 2006 (Ferraz et al., 2016). Dalponte and Coomes (2016) correctly detected 100% of trees with >80 cm DBH in their validation plots, but failed to detect small trees (Dalponte and Coomes, 2016). Both the Ferraz et al. (2016) and Dalponte and Coomes (2016) results align with my results. In our pilot sites, Dalponte2016 overestimated tree crowns by missing smaller trees found in close canopies with larger crowns, and Adaptive Mean Shift had low under and over segmentation (28% each) with higher accuracy (42%) indicating a preservation of individual tree shape.

4.1 Algorithm Performance in urban environments

While algorithmic approaches to individual tree segmentation offer valuable tools for urban forest management, it also comes with challenges that must be addressed to utilize their potential fully. Some of these challenges include:

- High spatial heterogeneity: Urban environments are characterized by a high degree of spatial heterogeneity, with a mix of built-up areas, vegetation, and open spaces. This complexity can make it difficult to accurately classify and analyze remote sensing data, as it can be challenging to distinguish between different land cover types and tree species.
- Spectral confusion: In urban environments, various features like buildings, roads, and different types of vegetation can have similar spectral signatures, making it difficult to distinguish between them in remote sensing imagery accurately. This can lead to errors in the classification and analysis of urban forests.
- Shadow and illumination effects: Urban areas often have tall buildings and structures that can cast shadows on the ground or vegetation, affecting the quality of remote sensing

data. Shadows can reduce classification accuracy and make detecting and assessing certain features in urban forests difficult.

- Data resolution: Remote sensing data can vary in spatial, spectral, and temporal resolution. High-resolution data is often required for detailed urban forest analysis, but acquiring and processing such data can be expensive and computationally intensive. Conversely, low-resolution data may not capture the fine-scale details necessary for accurate urban forest assessment.
- Data availability and accessibility: Although remote sensing data is becoming more widely available, accessing high-quality, up-to-date data can still be challenging for some users due to cost, data sharing policies, and technical expertise requirements.
- Integration with other data sources: To effectively manage urban forests, remote sensing data often needs to be combined with other data sources, such as land use data and ground-based surveys.
- Skill and expertise requirements: Analyzing and interpreting remote sensing data for urban forest management requires specialized knowledge and skills in remote sensing, GIS, and forestry. This can be a barrier for some organizations or individuals needing more expertise or resources to utilize remote sensing technology fully.

4.2 Recommendations for algorithm selection in urban environments

Despite these challenges, remote sensing remains a valuable tool for urban forest management. My research demonstrates the importance of comparing the performance of several algorithmic approaches to individual tree segmentation and offering the best approach based on project needs.

Should an application in urban forestry have substantial computational resources and avoid reproducible procedures across a diverse set of platforms and accessibility or only need to segment a minimal number of structurally homogeneous tree species, then I recommend deploying the Adaptive Mean Shift algorithm (Ferraz et al., 2016). However, should a project lack the financial or computational resources needed to perform a highly intensive algorithm like Adaptive Mean Shift, or if the target species span a heterogeneous mixture of structural variability, then I suggest deploying the Dalponte2016 algorithm using ground points as proxy tree tops with generalized parameters and implementing a several-meter buffer on the resulting tree crowns. For L. A. County, Dalponte2016 segmented 20 target species across an area greater than 45 km^2, producing a variety of crown metrics.

5 Conclusion

Accurate tree segmentation provides valuable information about the overall structure of the urban forest but also serves as a foundation for individual species classification, primarily through the application of segmentation results in species classification models. After deciding which of the better-performing algorithms best suits a project's needs, the resulting individual tree data can be integrated with additional information, such as multispectral or hyperspectral imagery and field survey data. This integrated dataset can comprehensively represent each tree, including its spectral, structural, and spatial characteristics. The integrated dataset can extract relevant features for each tree, such as spectral indices, canopy texture, crown shape, and height.

These features can serve as input variables for species classification models, enabling the differentiation of tree species based on their unique characteristics. Various classification algorithms can be applied to the extracted features to develop species classification models, including traditional machine learning methods (e.g., Random Forest, Support Vector Machines) and deep learning approaches (e.g., Convolutional Neural Networks). By training and validating these models on the integrated dataset, researchers can achieve high levels of classification accuracy, ultimately allowing for the identification of individual tree species within the urban forest.

In conclusion, accurate tree segmentation is essential for adequate urban forest inventory and management, providing critical information about individual trees and their spatial distribution. Additionally, segmentation results can be leveraged in species classification models to identify and manage tree species, further enhancing urban forest health and resilience. Species classification information can also inform species-specific management strategies, such as targeted pest control, customized pruning practices, and appropriate species selection for planting initiatives. This tailored approach to urban forest management can enhance overall forest health and resilience and direct the expertise of arborists and urban planning officials in a timely and efficient manner.

6 References

Avolio, M., Pataki, D. E., Gillespie, T., Jenerette, G. D., McCarthy, H. R., Pincetl, S., & Weller-Clarke, L. (2015). Tree diversity in southern California's urban forest: the interacting roles of social and environmental variables. *Frontiers in Ecology and Evolution*, 3, 73.

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and LiDAR data fusion. *Remote Sensing of Environment*, 148, 70-83.

Alonzo, M., McFadden, J. P., Nowak, D. J., & Roberts, D. A. (2016). Mapping urban forest structure and function using hyperspectral imagery and LiDAR data. *Urban Forestry & Urban Greening*, 17, 135-147.

Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J., ... & Vincent, G. (2019). A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests. *Remote Sensing*, 11(9), 1086.

Bodnaruk, E.W., Kroll, C.N., Yang, Y., Hirabayashi, S., Nowak, D.J., & Endreny, T.A. (2017). Where to plant urban trees? A spatially explicit methodology to explore ecosystem service tradeoffs. *Landscape and Urban Planning*, 157, 457-467.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and LiDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5), 1416-1427.

Dalponte, M., & Coomes, D. A. (2016). Tree‐centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in Ecology and Evolution*, 7(10), 1236-1245.

Degerickx, J., Roberts, D. A., McFadden, J. P., Hermy, M., & Somers, B. (2018). Urban tree health assessment using airborne hyperspectral and LiDAR imagery. *International Journal of Applied Earth Observation and Geoinformation*, 73, 26-38.

Dian, Y., Pang, Y., Dong, Y., & Li, Z. (2016). Urban tree species mapping using airborne LiDAR and hyperspectral data. *Journal of the Indian Society of Remote Sensing*, 44(4), 595-603.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., ... & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.

Ferraz, A., Saatchi, S., Mallet, C., & Meyer, V. (2016). Lidar detection of individual tree size in tropical forests. *Remote Sensing of Environment*, 183, 318-333.

Gillespie, T. W., de Goede, J., Aguilar, L., Jenerette, G. D., Fricker, G. A., Avolio, M. L., ... & Pataki, D. E. (2017). Predicting tree species richness in urban forests. *Urban Ecosystems*, 20(4), 839-849.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., … & Kommareddy, A. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850-853.

Hulley, M.E., 2012. The urban heat island effect: Causes and potential solutions. In *Metropolitan sustainability* (pp. 79-98). Woodhead Publishing.

Hulley, G., Shivers, S., Wetherley, E. and Cudd, R., 2019. New ECOSTRESS and MODIS land surface temperature data reveal fine-scale heat vulnerability in cities: A case study for Los Angeles County, California. *Remote Sensing*, 11(18), 2136.

Jones Jr, J. W. (2011). Using LiDAR Data and Geographical Information System (GIS) Technology to Assess Municipal Street Tree Inventories. Mississippi State University.

Kim, A. M., Kruse, F. A., & Olsen, R. C. (2016, May). Integrated analysis of light detection and ranging (LiDAR) and hyperspectral imagery (HSI) data. In *Laser Radar Technology and Applications* XXI (Vol. 9832, p. 98320W). International Society for Optics and Photonics.

Knapp N (2023). MeanShiftR: Tree delineation from lidar using mean shift clustering. R package version 0.1.1.

Krisanski, S., Taskhiri, M. S., Gonzalez Aracil, S., Herries, D., Muneri, A., Gurung, M. B., ... & Turner, P. (2021). Forest Structural Complexity Tool—An Open Source, Fully-Automated Tool for Measuring Forest Point Clouds. *Remote Sensing*, 13(22), 4677.

Li, W., Guo, Q., Jakubowski, M. K., & Kelly, M. (2012). A new method for segmenting individual trees from the lidar point cloud. *Photogrammetric Engineering and Remote Sensing*, 78(1), 75-84.

Liu, L., Coops, N. C., Aven, N. W., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. *Remote Sensing of Environment*, 200, 170-182.

Lin, Y., & Hyyppä, J. (2016). A comprehensive but efficient framework of proposing and validating feature parameters from airborne LiDAR data for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, 46, 45-55.

Locke, D. H., Romolini, M., Galvin, M., O'Neil-Dunne, J. P., & Strauss, E. G. (2017). Tree canopy change in coastal Los Angeles, 2009-2014. *Cities and the Environment* (CATE), 10(2), 3.

Los Angeles Region Imagery and Data Acquisition Consortium (LARIAC). (2016). LARIAC4 LiDAR Imagery (LiDAR). Los Angeles, CA: LARIAC. https://lariac-lacounty.hub.arcgis.com/pages/lariac4-documents-data.

McPherson, E.G., Simpson, J.R., Xiao, Q., & Wu, C. (2011). Million trees Los Angeles canopy cover and benefit assessment. *Landscape and Urban Planning*, 99(1), 40-50.

Mullaney, J., Lucke, T., & Trueman, S. J. (2015). A review of benefits and challenges in growing street trees in paved urban environments. *Landscape and urban planning*, 134, 157-166.

Murtha, P. A., & Fournier, R. A. (1992). Varying reflectance patterns influence photo interpretation of dead tree crowns. *Canadian Journal of Remote Sensing*, 18(3), 167-173.

Omasa, K., Hosoi, F., & Konishi, A. (2007). 3D lidar imaging for detecting and understanding plant responses and canopy structure. *Journal of Experimental Botany*, 58(4), 881-898.

Pincetl, S., Gillespie, T., Pataki, D. E., Saatchi, S., & Saphores, J. D. (2013). Urban tree planting programs, function or fashion? Los Angeles and urban tree planting campaigns. *GeoJournal*, 78(3), 475-493.

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Roman, L. A., Conway, T. M., Eisenman, T. S., Koeser, A. K., Ordóñez Barona, C., Locke, D. H., ... & Vogt, J. (2021). Beyond 'trees are good': Disservices, management costs, and tradeoffs in urban forestry. *Ambio*, 50, 615-630.

Roussel J, Auty D, Coops NC, Tompalski P, Goodbody TR, Meador AS, Bourdon J, de Boissieu F, Achim A (2020). "lidR: An R package for analysis of Airborne Laser Scanning (ALS) data." *Remote Sensing of Environment*, 251, 112061. ISSN 0034-4257,

doi:10.1016/j.rse.2020.112061,
https://www.sciencedirect.com/science/article/pii/S0034425720304314.

Roussel J, Auty D (2023). Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R package version 4.0.3, https://cran.r-project.org/package=lidR.

Silva, C. A., Hudak, A. T., Vierling, L. A., Loudermilk, E. L., O'Brien, J. J., Hiers, J. K., ... & Khosravipour, A. (2016). Imputation of individual longleaf pine (Pinus palustris Mill.) tree attributes from field and LiDAR data. *Canadian Journal of Remote Sensing*, 42(5), 554-573.

Tang, H., Armston, J., Hancock, S., Marselis, S., Goetz, S., & Dubayah, R. (2019). Characterizing global forest canopy cover distribution using spaceborne lidar. *Remote Sensing of Environment*, 231, 111262.

Yang, G., Zhao, Y., Li, B., Ma, Y., Li, R., Jing, J. and Dian, Y., 2019. Tree species classification by employing multiple features acquired from integrated sensors. *Journal of Sensors*, 2019.

Yang, M., Mou, Y., Liu, S., Meng, Y., Liu, Z., Li, P., ... & Peng, C. (2022). Detecting and mapping tree crowns based on convolutional neural network and Google Earth images. *International Journal of Applied Earth Observation and Geoinformation*, 108, 102764.

Yao, W., Krzystek, P., & Heurich, M. (2012). Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sensing of Environment*, 123, 368-380.

Zhang, C., & Qiu, F. (2012). Mapping individual tree species in an urban forest using airborne LiDAR data and hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing*, 78(10), 1079-1087.

Zhang, C., Zhou, Y., & Qiu, F. (2015). Individual tree segmentation from LiDAR point clouds for urban forest inventory. *Remote Sensing*, 7(6), 7892-7913.

Chapter 3.

What're the odds? Probabilistic tree inventories in urban Los Angeles County

Abstract

This research paper details the application of machine learning, specifically the Random Forest algorithm, in classifying tree species within urban settings. Motivated by the need for effective and efficient urban forest management, a data-driven model was developed to recognize and categorize tree species based on remote sensing data from the National Agriculture Imagery Program (NAIP) dataset (RGB, IR, Hue, Saturation, Value, vegetation indices) at a 60 cm spatial resolution and LiDAR data from the Los Angeles Region Imagery Acquisition Consortium (LARIAC) on tree height at a 1 m spatial resolution. The model demonstrated a high accuracy for common street tree species (N = 10,664), such as *Quercus agrifolia* and *Washingtonia robusta*, with an overall User's and Producer's Accuracy of between 0.859 to 0.919 across different neighborhoods. I also present the outputs from my model in two distinct formats: point data, pinpointing the precise location of each individual tree, and polygon data, outlining the canopy extent of the tree crowns. To enhance visualization and interpretability, a color-coded map was produced that categorizes trees based on the identified species, offering a vivid display of the species distribution across the studied area. This visualization approach not only provides a comprehensive overview of tree species distribution but also enables users to assess and improve the model's accuracy in real-world scenarios. Results were less robust in classifying trees on private property and less common species (0.620 accuracy for *Pinus pinea*, N = 42; 0.689 accuracy for *Citrus* spp., N = 196), indicating a bias towards species with larger training

samples. To mitigate this, the paper proposes strategies such as expanding the dataset to include more instances of private trees and using a binary classification approach. The successful application of the model will support initiatives such as public tree health monitoring, longitudinal studies, and urban forestry management, thereby enhancing our understanding of urban ecosystems and their roles in improving urban health. Future research will focus on refining the model with new data and investigating further adaptations for improved accuracy.

Keywords

Airborne Remote Sensing, Individual Tree Segmentation, Light Detection and Ranging, Urban Forestry

1 Introduction

Urban trees mitigate heat, clean air and water, and provide residents with an improvement in general well-being (Salmond et al., 2014; Gillner et al., 2015; Lee et al., 2016; Ozdemir, 2018; Los Angeles County Department of Public Health, 2020). In Los Angeles County, trees are a critical infrastructure as they help cool neighborhoods, clean air and water, and improve our communities' emotional and social health. Los Angeles County manages at least 800,000 public trees, with inventory costs using traditional methods exceeding $5 million (Los Angeles County Department of Public Health, 2020). These approaches are also limited to trees in the public right-of-way. The ability to inventory public and private trees using remote sensing will provide the county a faster, cheaper, and more efficient way to inventory trees throughout the region

(Fassnacht et al., 2014; Hutt-Taylor and Ziter, 2022; Liu et al., 2023). Redevelopment and land-use dynamics can disrupt urban tree species' natural life cycle. Regardless of their benefits to residents, urban trees are subject to how we choose to reconfigure urban spaces (Pincetl et al., 2013; Avolio et al., 2015; Bodnaruk et al., 2017). Moreover, invasive species, diseases, and climate change are also posing additional threats (Lausch et al., 2016; Schwantes et al., 2016; Pretzsch et al., 2017; Okin et al., 2018; Dong et al., 2023), and to address some of these challenges Los Angeles County has teamed up with the Department of Geography at UCLA to identify the best approach in tree species identification and health assessment to help ease the burden of inventorying local trees manually. Currently, no department within the county has a cost-effective or efficient way to know where and when to plant or manage trees. Those in charge of managing the urban forest are stretched thin and often have disparate tree inventory data, making hands-on management difficult to scale across the region (Hutt-Taylor and Ziter, 2022). This chapter will cover a remote sensing model designed to optimize and automate Los Angeles County's urban forest management to alleviate the burden of local departments that currently conduct visual assessments of local tree stock.

Over the last 20 years, there has been a dramatic increase in airborne and spaceborne multispectral sensors with fine spatial (1 m) and spectral (400 nm to 2300 nm) resolutions that can be used to identify tree individuals to species in urban settings (Xiao et al., 2004; Pu and Landry, 2012, Alonzo et al., 2013; Ferreira et al., 2016). Multispectral data consists of medium (10-30 m) to very high resolution (VHR) (5 cm – 5 m) imagery captured on both airborne and spaceborne platforms (Fassnacht et al., 2016; Immitzer et al., 2016; Fang et al., 2020). The sensors passively capture reflected energy from objects on the ground, including vegetation, and span the visible (red, blue, green) and infrared (longwave, shortwave, and near-infrared) spectra.

In addition, spectral datasets with high temporal resolutions, or consistent repeat observations, can provide a reliable measure of phenological information for target tree species (Bodnaruk et al., 2017; Gillespie et al., 2017; Pu and Landry, 2018). Another advancement in remote sensing technology is the introduction of Light Detection and Ranging (LiDAR) data. LiDAR uses the light from a laser to collect measurements of x,y, and z point clouds to create 3D models and maps of objects. Discrete return LiDAR with 2-66 points per meter squared are commonly used for urban tree identification, and airborne LiDAR alone is able to identify up to 61% of urban trees at species level (Liu et al., 2017). Several studies show that LiDAR data increases tree identification accuracy when combined with spectral imagery (Alonzo et al. 2014, Dian et al. 2016, Liu et al., 2017).

To map tree species across my study area, I use a supervised classification. Supervised classification methods, such as Random Forest, offer a range of advantages when applied to the classification of tree species using remote sensing data (Lim et al., 2019; Yang et al., 2019; Coleman et al., 2020; Zhang et al., 2020). In addition to handling complex classification tasks with high accuracy, Random Forest provides measures of variable importance, which can be used to identify the most informative spectral bands or features for the classification task (Gislason et al., 2006). This is particularly useful in remote sensing applications where high-dimensional data are common. Random Forest also does not make any assumptions about the underlying data distribution, making it suitable for remote sensing data that often exhibit nonlinear and complex relationships (Rodriguez-Galiano et al., 2012). Additionally, Random Forest is robust to overfitting, which is a common problem with decision trees (Belgiu and Dragut, 2016). This is because it takes the mode of the output of individual trees, thereby balancing out any individual tree's overfitting. The model can also handle large datasets with high dimensionality well, which is typical in remote sensing applications where you have multiple spectral bands and derived indices as predictor variables (Belgiu and Dragut, 2016; Yang et al., 2019).

Mapping tree species serves as a critical visualization that helps end-users in day-to-day operations, as well as other interested parties, in understanding forest biodiversity and aiding conservation efforts. By mapping results as polygon data for tree crown extents, I gain a detailed overview of tree distribution. Crucially, the color-coded map—for species classification—enables a dual assessment of both species presence and model reliability. Such an approach not only facilitates a comprehensive understanding of tree species dispersion but also provides insights into potential areas of model refinement (Pu and Landry, 2012; Dian et al., 2016). This leads me to my research question for this study. Can Random Forest, combined with spectral data imagery (NAIP) and LiDAR data (LARIAC), be used to assess the classification accuracy of urban trees in three unincorporated neighborhoods in urban Los Angeles County?

1.1 Research Objectives

Working in direct collaboration with the intended stakeholders, my research strives to understand daily end-users' needs to operationalize and optimize their urban forest management through remote sensing modeling. Several studies have explored the classification of tree species using remote sensing data in non-urban areas (Immitzer et al., 2012; Ferreira et al., 2016; Immitzer et al., 2016; Aubry-Kientz et al., 2019; Lim et al., 2019; Mosin et al., 2019; Zhang et al., 2020). For instance, Fassnacht et al. (2016) demonstrated that the integration of optical and structural remote sensing data significantly improves tree species classification in temperate forests.

However, the application of these methodologies in urban areas remains varied. Current approaches often lack the precision required for urban areas, where tree species are more diverse, and interspersed with anthropogenic structures. This research aims to address this gap by developing and applying a supervised classification model that combines high resolution optical and structural remote sensing data for tree species identification in urban Los Angeles County. This approach has the potential to refine our understanding of urban tree species composition and contribute to more effective urban biodiversity management strategies.

Specifically, my research focuses on determining the efficacy of machine learning algorithms in analyzing 4-band aerial remote sensing imagery, as well as assessing the potential improvements in accuracy and efficiency as compared to manual tree inventorying. Accurate identification of tree species plays a pivotal role in the maintenance of biodiversity, the provision of ecosystem services, and in the overall management of urban forests. Different tree species offer varied benefits and have distinct needs, emphasizing the necessity for accurate identification. Moreover, understanding the species composition of urban forests is critical for making informed decisions about future plantings, especially in the context of climate change and invasive species. This study, therefore, addresses critical factors in urban forestry, such as improving tree species identification (Fassnacht et al., 2016; Branson et al., 2018), optimizing resource allocation (Meddens et al., 2013; Caughlin et al., 2019), and enhancing our understanding of urban biodiversity (Isager and Niels, 2007; Lausch et al., 2016; Huang et al., 2019). By filling a substantial gap in the literature, this research will provide valuable insights that can guide future urban forestry management strategies. The proposed research will not only contribute to methodological advances in urban forestry but also enhance our understanding of urban biodiversity, facilitate efficient forest management, and foster resilience against emerging environmental challenges. This will provide data-driven insights to underpin urban forest management decisions at the species level. Second, by integrating results from my analyses with long-term monitoring of target species, data users/practitioners can gain a deeper understanding of the distribution and condition of urban forest species. My approach supports urban foresters' ability to optimize these benefits and mitigate environmental challenges such as climate change and habitat loss, which all starts with inventorying species to begin to track their health over time. The subsequent sections delve deeper into my study's methodology, the performance of a supervised classification algorithm, and the practical implications of this research for urban forest management in Los Angeles County.

2 Methodology

The methodology adopted in this research is a combination of remote sensing image processing and random forest supervised classification, performed through cloud-based computational software.

2.1 Pilot Sites

Public grants often opt for pilot sites to test the feasibility and efficacy of proposed research methodologies before full-scale implementation. This allows for identification and rectification of potential problems or challenges in a controlled setting, ensuring the method's reliability and accuracy. Pilot sites serve as a microcosm of the larger area of interest, providing valuable insights into how the method might perform when scaled up. They also offer a cost-effective way to fine-tune the approach, making it more robust and efficient before it is applied to a broader geographic context. Hence, the selection of appropriate pilot sites is a critical component of the research process, particularly in this study that involves technical approaches and large-scale data analysis.

Three pilot sites were chosen to serve as a representative sample of urban Los Angeles County. These sites include the unincorporated communities of Altadena, East Los Angeles, and Marina del Rey. They were chosen primarily for their differences in tree species, urban forest management, and demographics. Altadena is a community located approximately 20 km north of downtown Los Angeles. This neighborhood is nestled directly north of Pasadena at the base of the San Gabriel Mountains. With its diverse population, Altadena is known for its distinctive residential architecture, ranging from modest ranch-style homes to grand historic estates with a dense urban canopy. East Los Angeles, colloquially referred to as East LA, is one of the most densely populated urban areas in the United States. This predominantly Hispanic community boasts a vibrant culture, and its urban canopy is more open compared to Altadena. Marina del Rey, located on California's southern coastline, is an unincorporated seaside community in Los Angeles County. Known for its marina, the world's largest man-made, small-craft harbor, and is in close proximity to the Los Angeles International Airport. Marina del Rey's urban forest is managed by the County's Beaches and Harbors department, and it has one of the least diverse street tree populations in the group.

2.2 Ground Truth Data

Ground truth data (N = 16,947 total trees; 10,664 used in this study) were collected from several departments within Los Angeles County, including Public Works, Parks and Recreation, Beaches and Harbors, and Public Health. This data serves as an important reference point for evaluating the accuracy of tree species identification derived from remote sensing and machine learning techniques. It provides a "truth" against which the results of the machine learning algorithms can be compared and validated. The ground truth data is a critical component of the study, and can offer a reliable and objective means of assessing the efficiency and accuracy of the proposed tree species identification methodology (Mosin et al., 2019). It is important to quality assess and quality control ground truth data to ensure its reliability as species information and GPS location. In the case of the County ground truth data, contracted arborists identified tree individuals to species and used handheld GPS units to log individual tree locations at breast height (1.3 m), after which, the data was validated through a second independent contractor to ensure accuracy and reliability (Los Angeles County Department of Public Health, 2020). This research intends to enhance the User's Accuracy of species identification, thus facilitating more effective urban forest management in Los Angeles County.

This collection of 16,947 ground points were filtered to 10,664 tree locations by removing individual tree crown segments that overlapped with two or more species ground points. The resulting 10,664 street trees are all located in the public right-of-way that each department is responsible for maintaining within the three study areas. Table 1 provides an overview of the original 16,947 target tree species and Genera per community, and why they were chosen for this study. With over 500 species in the Los Angeles metropolitan area alone

(Pataki et al., 2013), the region has been found to be more diverse than any native forest in the United States (Pincetl et al., 2013; Gillespie et al., 2017; Love et al., 2022). Running a full classification model on every species available would be time consuming and computationally expensive for this pilot project. Instead, I opted to model a select number of species/genera that were most important for end-users. There are myriad ways of putting together such a technical approach, which can be a drawback if the proposed model does not consider the end-user. Also, no model is a perfect representation of our reality and acknowledging the presence of errors turns any drawbacks into opportunities to test and validate the model. My approach's success will need the end user to assess their expertise and feel confident evaluating the model output. Thus, stakeholder involvement becomes a crucial step in the model's conception and throughout its development. I surveyed regional stakeholders (Tree People, The Nature Conservancy), end users (day-to-day managers spanning eight County departments), and scientists (UCLA, UCSB, NASA JPL) as to which group of species would be best to test in L.A. Eleven target species/genera (four native to California) were chosen due to their importance when considering water stress, mortality, and hazards. The four native species were chosen for conservation and management purposes. I provide information on each species, the reason for its inclusion, and the number of training points included in the model in Table 1 in Chapter 2.

2.3 Remote Sensing Imagery

2.3.1 High-resolution Spectral Imagery

I used species-specific tree crowns from a Canopy Height Model derived from the Los Angeles Region Imagery Acquisition Consortium (LARIAC). The LARIAC dataset is a comprehensive, high-resolution geospatial data collection covering the entirety of Los Angeles County. This is a proprietary dataset provided courtesy of the LARIAC and EagleView (LARIAC, 2016). The richness and depth of the LARIAC dataset are particularly beneficial for this research as it aids in the accurate identification and classification of tree species across the diverse and expansive urban forests of Los Angeles County. The dataset includes a 4-band (Red, Green, Blue, and Near-Infrared) aerial imagery with a pixel resolution of 10.2 cm, enabling detailed and precise mapping of urban areas. At 10.2 cm spatial resolution, image processing becomes computationally expensive and timely. Instead, I used publicly available 4-band imagery provided by the National Agriculture Imagery Program (NAIP). NAIP is an initiative by the United States Department of Agriculture (USDA) to acquire aerial imagery during the agricultural growing seasons in the continental United States (NAIP, 2018). A key distinguishing feature of NAIP imagery is its high resolution, with a spatial resolution of 60 cm (NAIP, 2018), providing significant detail.

2.3.2 Light Detection and Ranging

The LARIAC dataset also includes Quality Level 2 discrete-return Light Detection and Ranging (LiDAR) point-clouds, which are instrumental for generating the canopy height model. The low point density, however, introduces a spatial mismatch between the very high resolution optical imagery provided by LARIAC and its LiDAR dataset. By using the 60 cm NAIP imagery, I can ensure that at least one or two returns in the point-cloud data align within single pixels in my spectral data. This alignment ensures height data is spatially located with spectral returns for a complete spectral and structural profile for every pixel within my tree crown segments.

2.4 Image Processing

2.4.1 Cloud Computing: Google Earth Engine

Google Earth Engine is a cloud-based platform for planetary-scale environmental data analysis (Gorelick, 2017). The colossal computing infrastructure of Google Earth Engine provides capabilities to run geospatial analysis at an unprecedented scale. It offers a vast array of public datasets that include satellite imagery, geospatial datasets, and client-side functions for manipulating and analyzing data. There are two primary reasons to use Google Earth Engine for image processing and supervised classification. First, Google Earth Engine excels in processing large-scale, high-resolution geospatial data, thanks to its cloud-based architecture. It performs on-the-fly computations, allowing users to visualize the results without first downloading and processing raw data locally. This feature is particularly beneficial for handling large-scale remote sensing datasets like those used in this project. Second, Google Earth Engine's supervised classification algorithms enable users to classify satellite imagery based on training datasets. This feature allows for the precise identification and mapping of different land cover types, in this case, tree species. By leveraging Google Earth Engine, I was able to expedite the process of image classification and ensure a high level of accuracy, thus facilitating efficient and effective urban forest management.

The process began with the acquisition of 2018 NAIP imagery from Earth Engine's extensive data catalog. I also uploaded two data assets: a Canopy Height Model (CHM) and a tree crown layer. Subsequently, the NAIP imagery underwent a series of processing steps, beginning with the generation of shadow masks to minimize classification errors due to varying light conditions. This was followed by a transformation of the RGB bands into the Hue, Saturation, and Value (HSV) color space, which is often more discriminative for vegetation analysis (Mostafa and Abdelhafiz, 2017; Han et al., 2020; Pu, 2021; Dyson et al., 2023). I also calculated two vegetation indices, the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and the Soil-Adjusted Vegetation Index (SAVI), to better capture the biophysical characteristics of the urban trees (Fassnacht et al., 2016). These layers, along with the processed NAIP bands, were then stacked to create a comprehensive set of spectral and spatial features.

Leveraging the tree crown layer, I sampled this stacked dataset to extract representative training data for supervised classification of individual tree species. To construct a robust classifier, I employed a random forest algorithm, partitioning the sampled data into an 80/20 split for training and validation purposes. The trained model was then applied across a larger geographic extent to perform species-level classification of public urban trees. The cloud-based computation capabilities of Google Earth Engine allowed for this entire process to be conducted without the prohibitive computational and storage overheads typically associated with local systems. Through this method, I achieved a rapid and precise classification of tree species, providing a solid foundation for informed decision-making in urban forest management.

2.4.2 Canopy Height Modeling

The CHM can be used as an alternative to point clouds for individual tree crown detection in discrete-return LiDAR data (Zhang et al., 2015; Lin and Hyyppa, 2016). Despite the faster processing speeds, using CHMs introduces complexities due to dependency on its construction parameters. For this study, I used the LARIAC LiDAR point-clouds to produce CHMs for my study areas at a 1 m spatial resolution with a points-to-raster, pitfree algorithm for image smoothing in the open-source lidR R package (Roussel et al., 2020; Roussel, 2023). This algorithm avoids "holes" in the CHM to avoid issues with individual tree detection and crown identification (Roussel et al., 2020; Roussel, 2023). The post-processing smoothing step includes applying a median filter. In addition to helping delineate individual tree crowns, the CHM also provides crucial height information to distinguish between species in our supervised classification.

2.4.3 Shadow Masking and Color Space Conversion

Shadow masking is an important step in remote sensing image processing as it helps to eliminate the effects of shadows in surface reflectance (Mostafa and Abdelhafiz, 2017; Han et al., 2020; Pu, 2021). Shadows can distort the accurate identification and interpretation of features within an image due to the reductions in illumination. Particularly in applications like supervised classification of tree species, shadows cast by taller objects can obscure or alter the perceived attributes of the underlying tree species in the imagery, leading to misclassification. By applying a shadow mask (Fig. 1), these areas are effectively 'masked out' of the analysis, thereby improving the accuracy and reliability of the classification results.

Figure 1. Visualization of vegetation shadow mask in Altadena, CA using NAIP data.

The HSV color space is an alternative way of representing color data that separates the chromatic information (Hue and Saturation) from the luminance information (Value). Unlike the Red, Green, Blue (RGB) color space, where each dimension corresponds to a primary color, the HSV color space is cylindrical, with hue represented as an angular dimension around the central vertical axis, saturation as the radial distance from the axis, and value as the vertical dimension (Mostafa and Abdelhafiz, 2017; Han et al., 2020).

The Hue in HSV corresponds to the dominant wavelength of light (e.g., red, green, blue, near-infrared), saturation corresponds to the intensity or purity of the hue, and value corresponds to the brightness of the color. By transforming the standard RGB color space into HSV, I gain the ability to work with colors in a way that is more aligned with how humans perceive color. The calculation of Hue in the HSV color space is a function of the Red, Green, and Blue (RGB) values of a pixel. The formula for computing Hue (H) differs depending upon which RGB channel has the maximum value.

If Red is the maximum, then Hue (H) is computed as: `(G-B)/(max-min) * 60°`

If Green is the maximum, then Hue (H) is computed as: `60° + (B-R)/(max-min) * 60°`

If Blue is the maximum, then Hue (H) is computed as: `120° + (R-G)/(max-min) * 60°`

In these formulae, `max` and `min` are the maximum and minimum values among the Red, Green, and Blue channels respectively. The resulting Hue value was then normalized to a range of 0° to 360°. Note that if the maximum value equals the minimum value, the Hue was defined as 0°, which represents a shade of gray and helps determine shadows in an image. Han et al. (2020) also provide the arithmetic used to compute both Saturation and Value using the RGB inputs. The formula for computing Saturation (S):

$$S = 1 - (3/(R+G+B))\min(R,G,B)$$

The formula for computing Value (V):

$$V = 1/3(R+G+B)$$

The Near-Infrared, Red, Green (NRG) color space can further enhance the differentiating potential of shadows within vegetation. Near-Infrared (NIR) light, invisible to the human eye, is strongly reflected by healthy vegetation. Thus, introducing NIR into the color space can significantly improve the ability to discern subtle spectral differences in the vegetation, leading to more accurate shadow detection and elimination. First, the NRG color space was calculated by reassigning the Near-Infrared, Red, and Green bands to the RGB channels, respectively. Then, similar to how Hue is calculated in the HSV color space, a new set of formulae is employed to calculate the Hue in the NRG color space. The computation can be expressed as:

If Near-Infrared is the maximum, then Hue (H) is computed as: `(R-G)/(max-min) * 60°`

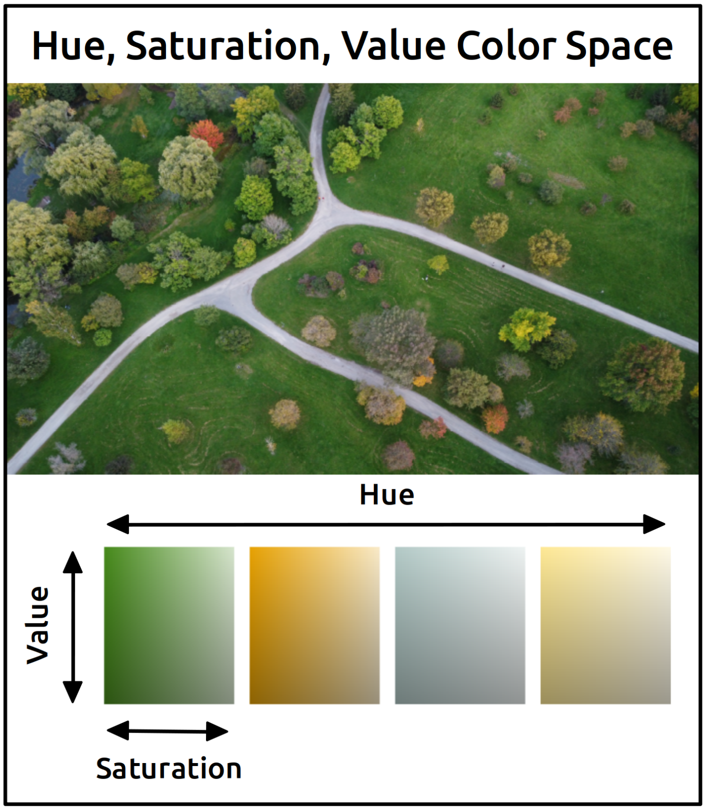
If Red is the maximum, then Hue (H) is computed as: `60° + (G-NIR)/(max-min) * 60°`

If Green is the maximum, then Hue (H) is computed as: `120° + (NIR-R)/(max-min) * 60°`

Again, `max` and `min` are the maximum and minimum values among the Near-Infrared, Red, and Green channels respectively. By normalizing the resulting Hue value to a range of 0° to 360°, it is then possible to better differentiate shadows in the vegetation, thus enhancing the accuracy of overall image analysis and feature classification.

In the context of shadow masking, the HSV color space is particularly useful because it isolates the luminance component (Value), which is significantly affected by shadows (Mostafa and Abdelhafiz, 2017; Han et al., 2020). By focusing on the value dimension of the HSV color space, it becomes easier to identify and mask out shadow-affected areas in an image. Specifically, areas of an image that are in shadow will have a significantly lower value, indicating less light. These can be identified and masked out, improving the quality of the image analysis and the accuracy of subsequent classification tasks.

Figure 2. HSV color space infographic.



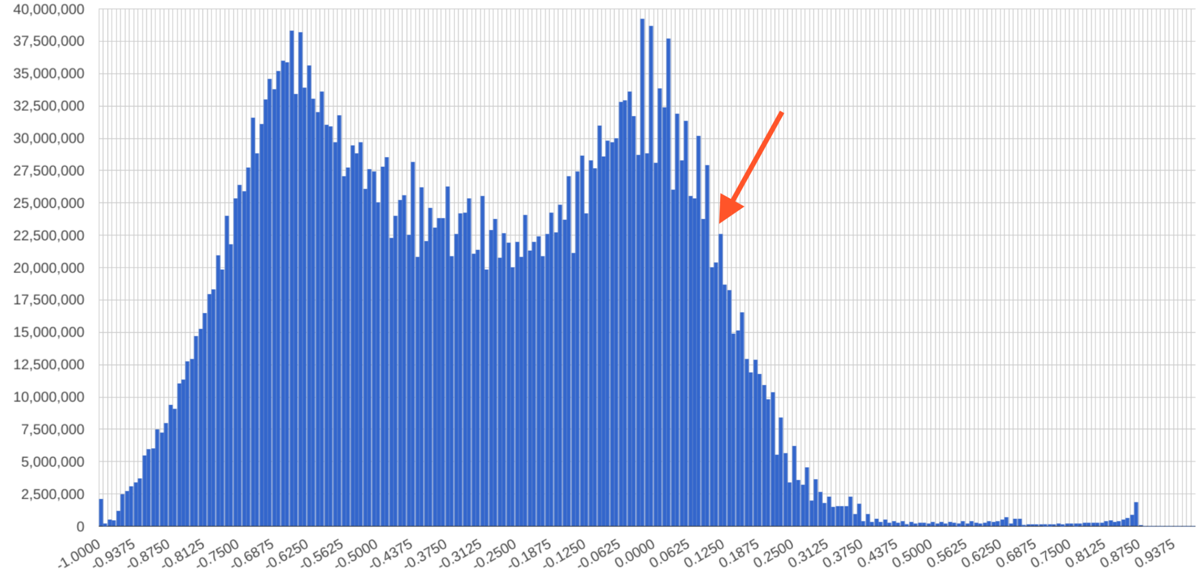
Normalized Saturation-Value Difference Index (NSVDI) is an index in remote sensing image analysis that is particularly effective in urban environments where vegetation is mixed with built-up areas (Mostafa and Abdelhafiz, 2017; Han et al., 2020). NSVDI is derived from the HSV color space, specifically using the Saturation (S) and Value (V) components. The formula for NSVDI is given as:

$$NSVDI = (S - V) / (S + V)$$

By normalizing the difference between saturation and value, NSVDI effectively accentuates the vegetation signal in an image, while suppressing non-vegetation elements. This is because in general, vegetation tends to have high saturation (due to the strong absorption of light in specific wavelengths by chlorophyll) and high value (due to the reflection of light in the near-infrared wavelength where vegetation strongly reflects) (Huang et al., 2020). The NSVDI therefore provides a useful tool for discriminating vegetation from non-vegetation cover types, improving the accuracy of urban vegetation mapping and other related analyses (Mostafa and Abdelhafiz, 2017; Han et al., 2020).

The inflection point in an NSVDI plot represents a transition in the spectral properties of the scene, often aligning with a change in land cover or the presence of a shadow (Mostafa and Abdelhafiz, 2017). Theoretically, an NSVDI value of 0.0 should serve as the threshold for distinguishing shadows in an image (Mostafa and Abdelhafiz, 2017). When plotted (Fig. 3), NSVDI of 0.0 is the most prominent feature, however, previous studies have found that using the final inflection point, or the value that experiences the greatest rate of change following the 0.0 prominence, is a better threshold to use for shadow masking (Mostafa and Abdelhafiz, 2017; Han et al., 2020). The inflection point may vary from study to study, but by identifying this value, one can determine an appropriate threshold for the NSVDI which can be used to create a shadow mask. Areas of the image with an NSVDI value below this threshold can be classified as shadow, as they represent areas with lower light saturation and value, typical characteristics of shadowed regions. This shadow mask can then be applied to the image, effectively 'masking out' these shadow-affected areas from subsequent analyses. In essence, the inflection point serves as a critical parameter for differentiating shadowed and non-shadowed areas in the imagery. By strategically applying this shadow mask, the accuracy of vegetation mapping and other related image analyses can be substantially improved, hence compensating for any spectral distortions caused by shadows. It is important to note that the threshold value determined from the inflection point may vary depending on the unique lighting conditions and land cover characteristics of each scene. Therefore, it may be necessary to adjust this value for different images or study areas to ensure optimal shadow detection and masking.

Figure 3. Identification of the NSVDI inflection point for shadow masking threshold.



2.4.4 Vegetation Indices

Vegetation indices (VIs) are combinations of spectral bands designed to highlight specific properties of vegetation in remote sensing imagery. They are computed mathematical calculations using the spectral reflectance of different bands, particularly those in the red and near-infrared (NIR) wavelengths, where the spectral response of vegetation is most distinctive (Huang et al., 2020).

VIs are crucial in remote sensing for several reasons. First, they enhance the visibility of vegetation in the imagery, making it easier to discriminate between vegetated areas and other land cover types. Second, they serve as proxies for vegetation properties such as biomass, leaf area index (LAI), and photosynthetic activity, which are otherwise difficult to measure directly at larger scales. Third, as VIs are sensitive to vegetation health and stress, they are applied extensively for monitoring vegetation dynamics, including changes in phenology, growth, and response to stressors such as drought, pests, or disease.

The use of VIs offers several benefits in the context of identifying individual trees to species. Similar to HSV color space, VIs can help normalize for different illumination conditions across an image, making the classification process more robust. Also, certain VIs are sensitive to specific physiological characteristics of vegetation, which can be used as discriminating features in the classification process. The reliability of VIs as key indicators of tree species is attributed to their correlation with various biophysical properties of vegetation, such as leaf area index, canopy cover, and photosynthetic activity (Rouse et al., 1973; Jackson, 1983; Huete, 1988; Purevdorj et al., 1998; Huete et al., 2002).

The most common VIs include the NDVI, EVI, and SAVI, each with its unique formula and use case (Fassnacht et al., 2016). For instance, NDVI is often used for broad assessments of vegetation cover and photosynthetic activity (Rouse et al., 1973; Jackson, 1983; Purevdorj et al., 1998), while EVI is more sensitive to canopy structural variations and better able to penetrate through atmospheric particles (Huete et al., 2002). SAVI, on the other hand, minimizes soil brightness influences, making it suitable for areas with sparse vegetation (Huete, 1988).

To calculate each of the mentioned vegetation indices, one requires specific band data from the imagery. The formulas for each index are as follows:

- **Normalized Difference Vegetation Index (NDVI)**: NDVI is computed using the Near Infrared (NIR) and red bands of the electromagnetic spectrum. The formula is given as:

  NDVI = (NIR - Red) / (NIR + Red)

- **Enhanced Vegetation Index (EVI)**: EVI takes into account atmospheric corrections, especially in areas with high aerosol content. It uses the blue band in addition to the NIR and red bands. The formula for EVI is:

  EVI = 2.5 * ((NIR - Red) / (NIR + 6 * Red - 7.5 * Blue + 1))

- **Soil Adjusted Vegetation Index (SAVI)**: SAVI minimizes the effects of soil brightness on the computed index, allowing for better discrimination of vegetated areas in sparse vegetation and bare soil areas. The formula is as follows:

  SAVI = (NIR - Red) / (NIR + Red + 0.5) * (1 + 0.5)

2.5 Supervised Classification: Random Forest

To deploy Random Forest for tree species classification, the model is 'trained' using labeled example objects, i.e., examples where the desired output (class labels) is known (Lim et al., 2019; Yang et al., 2019; Coleman et al., 2020; Zhang et al., 2020). The model then learns from these training examples and applies this learned knowledge to classify new objects. Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees during training and outputting the class that is the mode of the classes of individual trees (Breiman, 2001; Immitzer et al., 2012; Belgiu and Dragut 2016). In essence, it combines the predictions of several base estimators built with a given learning algorithm to improve robustness over a single estimator.

2.5.1 Model Inputs

I used a 14-band image stack as a comprehensive set of data for tree species classification. Besides the traditional RGBN (Red, Green, Blue, Near Infrared) bands, I also use HSV (Hue, Saturation, Value) derived from both RGB and NRG. These HSV layers can sometimes highlight subtle differences in color and brightness, crucial in distinguishing between different species. I also include Height (CHM) as an input to add a three-dimensional component to the classification. This captures the vertical structure of the forest, which can be a determinant in discriminating between species, especially those with markedly different growth forms or maturity stages. Moreover, I include all three vegetation indices to gain valuable information on vegetation health, productivity, and structure. Each of these indices captures different aspects of the vegetation spectral response and can be particularly insightful for differentiating between species based on their physiological properties or adaptation to specific environmental conditions.

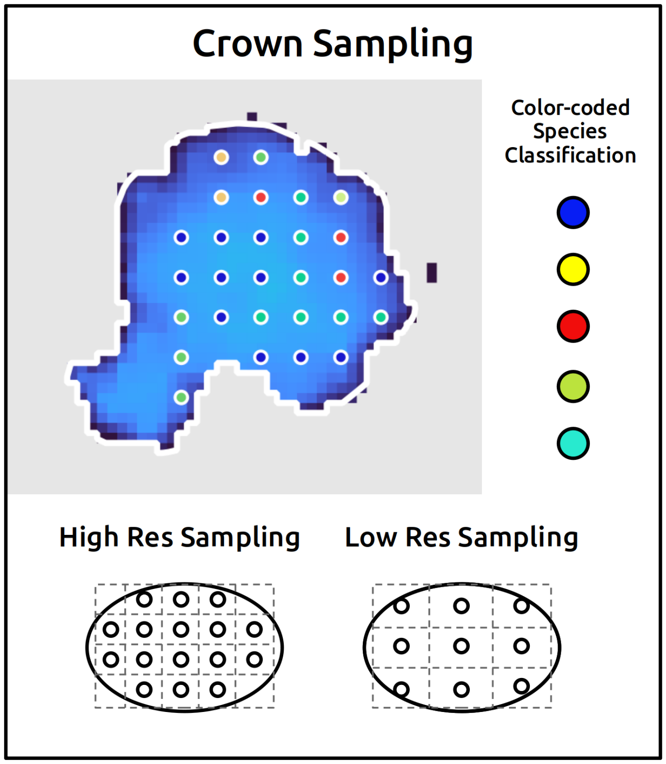
2.5.2 Parameter Fine-tuning

Model Parameters

The parameters selected for this Random Forest model are crucial for its performance and accuracy. The spatial resolution for point sampling plays a key role in determining the level of detail captured from each tree crown, which directly influences the classification outcome. The number of decision trees in the Random Forest is another critical parameter. A higher number increases the robustness of the model, as it reduces the likelihood of overfitting by averaging out anomalies, but at the cost of computational effort and time. The minimum leaf population, which determines the minimum number of samples required to be at a leaf node, is a parameter that governs the depth of the trees and can be adjusted to control overfitting.

For my study sites, I chose a spatial resolution of 2 m for point sampling. While fine-tuning this spatial scale, sampling band values at a spatial resolution of less than 1 m ran into computational limits in Google Earth Engine, even when accounting for the other parameters. Spatial scale less than 2 m but greater than 1 m ran for Marina del Rey and East LA, but not for Altadena. Scales greater than 3 m saw a drop in model performance across all sites, even when accounting for the number of decision trees and the minimum leaf population. For the number of decision trees, I initially chose 100 prior to applying a k-fold cross validation. After running the cross validation with 10 folds, I again ran into computational limits and reduced the number of trees to 10 and the minimum leaf population to 1. Despite reducing the number of decision trees, spatial resolution remained vital for determining model performance, and a reduction in decision trees ensured a higher resolution could be used while also performing a k-fold cross validation.

Figure 4. Random forest crown sampling and demonstrated species classification.



To resolve the multiple species classes returned by the model for each tree crown, a result of sampling multiple points per crown, I compute the mode - the most frequently occurring value - of the species classes predicted for the sampled points. This mode serves as the final classification label for that particular tree crown. This method capitalizes on the power of collective decision-making, ensuring that the most dominant classification label, as represented by the mode, is selected as the final output. This approach minimizes the impact of potential outliers or anomalies in the point sampling within individual tree crowns, thereby enhancing the overall accuracy of the tree species classification.

Model Validation and Data Split

For the training and testing data of individual tree species, an 80/20 split was used. This means that 80% of the dataset is used for training the model (N = 8,531 trees), and the remaining 20% is set aside for testing (N = 2,133). This ratio ensures a sufficient number of examples for the model to learn from, while also retaining a substantial subset for validation, thereby providing a robust measure of the model's predictive capability on unseen data (Gislason et al., 2006).

K-fold Cross Validation is another crucial step in evaluating the robustness and validity of a Random Forest model. This technique subdivides the original dataset into k number of equal-sized subsamples or 'folds'. Of these 'k' folds, a single fold is retained as the validation data for testing the model, while the remaining k-1 folds serve as training data. The cross-validation process is repeated k times, ensuring each fold is used exactly once as the validation data. The k results from the folds can then be averaged to produce a single estimation. This method is particularly effective in utilizing a limited dataset as it maximizes both the training and testing data. It provides a more comprehensive insight into how the model is performing across different subsets of data, rather than a single holdout sample. This is especially relevant in the context of Random Forest, where the model's performance can vary based on the specific training and testing split. By averaging across multiple folds, k-fold Cross Validation helps mitigate this variability and provides a more robust measure of model performance. For my model, I performed a 10-fold cross validation as this is a standard number of folds used in this approach (Kohavi, 1995; Tian, 2006).

2.5.3 Model Accuracy Assessment & Outputs

The output of this Random Forest classification model is a map and a comprehensive set of accuracy metrics. First, a confusion matrix is produced, which is a table layout that allows visualization of the machine learning algorithm performance. It yields the classification errors made by the classifier, distinguishing between the true and false positives and negatives.

Second, a classification report is generated. This report is a breakdown of User's Accuracy, Producer's Accuracy, F1-score, and support for each class. User's Accuracy quantifies the number of positive class predictions that actually belong to the positive class. Producer's Accuracy indicates the number of positive class predictions made out of all actual positive examples in the dataset. F1-score is a harmonic mean of User's and Producer's Accuracy, and support is the number of actual occurrences of the class in the specified dataset.

1. **User's Accuracy (Precision):** Also known as precision, this metric is the proportion of true positive predictions (i.e., instances where the model correctly identifies a class) out of all positive predictions made by the model. In other words, when the model predicts a certain species, the User's Accuracy represents how often it's correct. A higher User's Accuracy means fewer False Positives.
2. **Producer's Accuracy (Recall):** Also known as recall, this metric is the proportion of true positive predictions out of all instances that truly belong to that class. In simpler terms, it represents the model's ability to correctly identify all instances of a given class. A higher Producer's Accuracy means fewer False Negatives.

Additionally, the model was used to predict species locations and produce a map of results. The model, however, cannot predict species classes not included in the training data.

2.5.4 Mapping Output

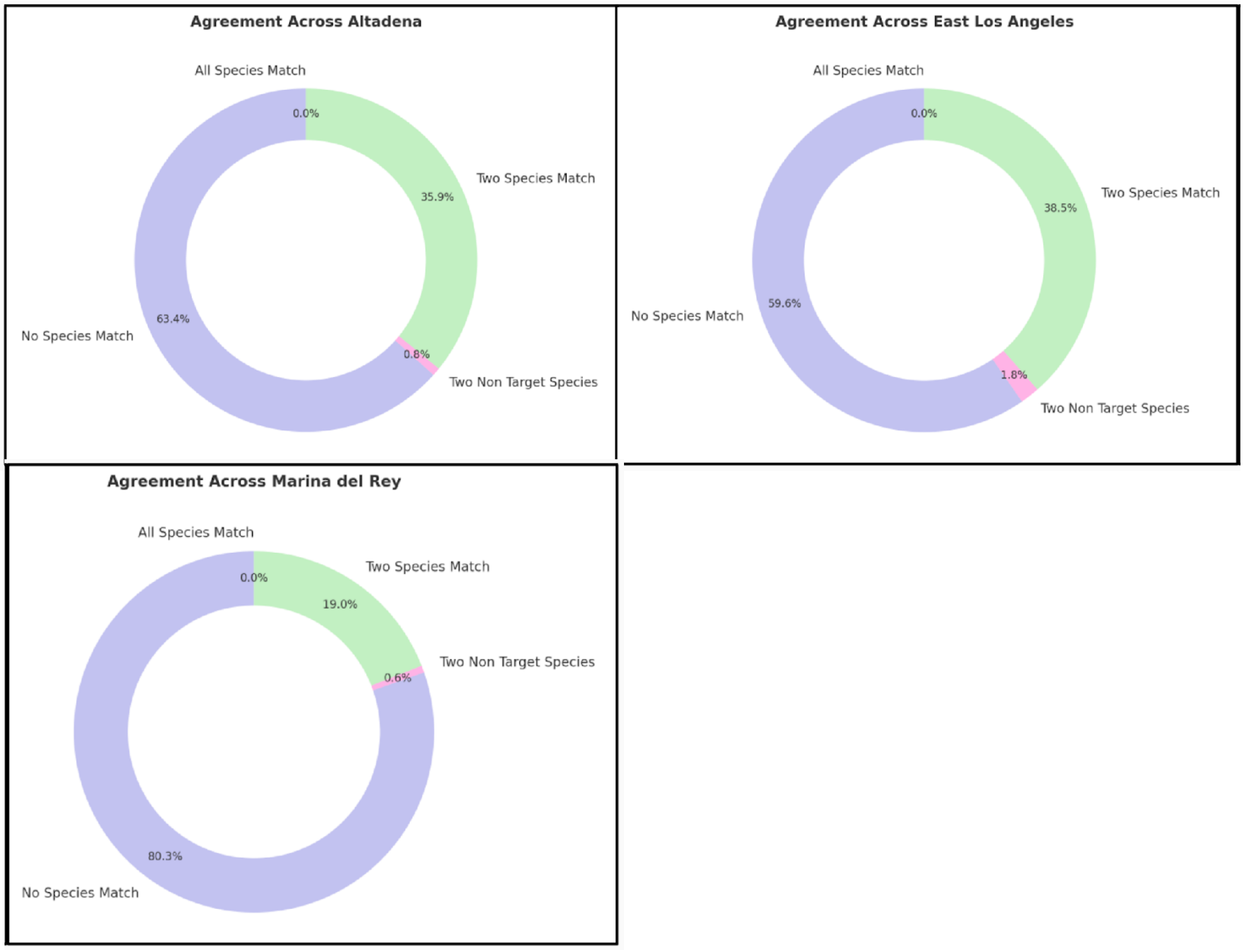
The mapping process I used starts with the classification outputs generated by Google Earth Engine, which were obtained in a table format for all three pilot sites. These tabulated classification outputs were spatially joined with the input tree crown polygons in the open-source, and free GIS software package, QGIS (QGIS Association, 2023). This data merger ensured that each tree crown polygon was affiliated with its respective classification result, enabling a seamless integration of tree species data with spatial features. After joining my results with the segmented tree crowns, I exported the joined crown polygons, representing individual tree locations. These were then mapped using a random color scheme, which was designed to differentiate tree species. This visualization method allowed for an immediate, spatially explicit understanding of classified tree species distribution across all study sites.

3 Results

3.1 Overall Accuracy

The individual performance of the model across the three study sites showed considerable promise, with high accuracies recorded for species identification (Fig. 6-8; Tables 1-3). However, these are individual results for three separately trained models using the ground points in the respective neighborhood. When attempting to scale into a single model across all three sites, I ran into user memory limits trying to sample over 10,000 tree crowns at a spatial resolution of 2 m. I had to resample at a lower, albeit still very high, spatial resolution (3 m) to avoid impossible computational needs on the Google Earth Engine server. This reduction in spatial resolution when sampling crowns was evident in the overall results and classification accuracy plummeted across all species, from an overall average of 89% to 64%. Additionally, when I used individual models to classify species in a different neighborhood, I saw very low agreement across all three models (Fig. 5). Across all three study sites, there was 0.00% agreement for any tree crown to have been classified as the same species in each model. There was an average of 31.13% agreement for two models to classify a tree crown as the same target species, and 1.07% agreement for non-target species. This indicated that one model could not be scaled to another neighborhood.

Figure 5. Agreement across all three models when each is used to classify species in all three neighborhoods.

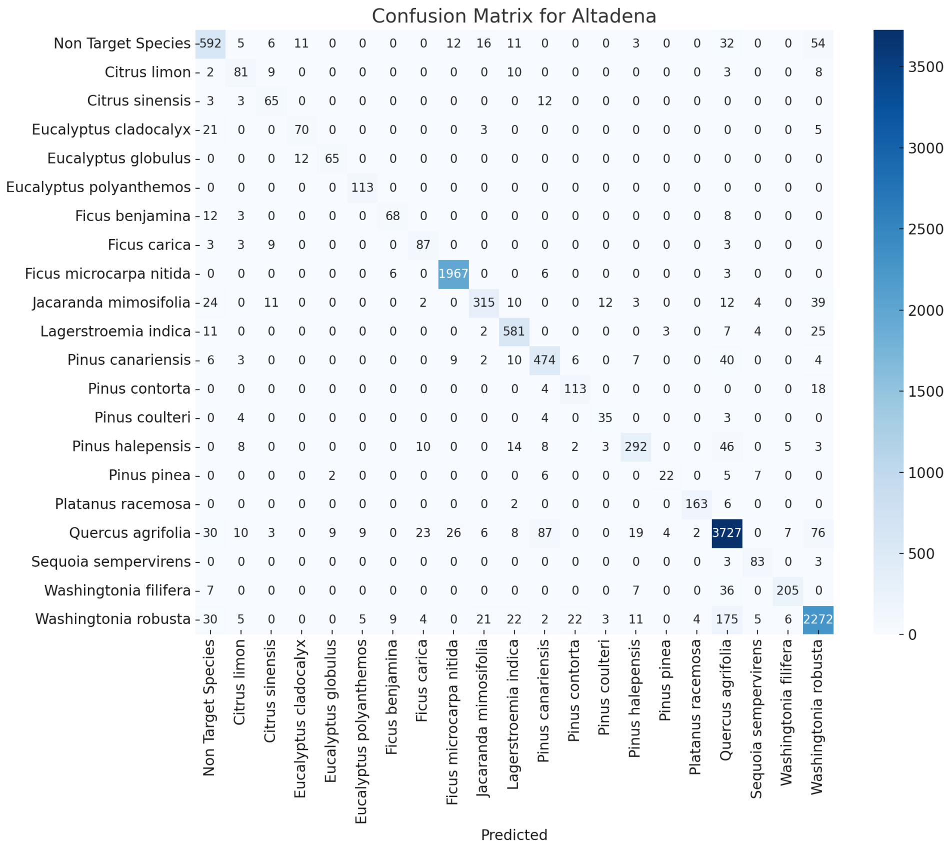


3.2 Accuracy Assessments: User's vs. Producer's Accuracy

Table 1. Accuracy Assessment for 20 tree species in Altadena.

Species	User's Accuracy	Producer's Accuracy	F1-Score	Support
Non Target Species	0.799	0.798	0.798	742
Citrus limon	0.648	0.717	0.681	113
Citrus sinensis	0.631	0.783	0.699	83
Eucalyptus cladocalyx	0.753	0.707	0.729	99
Eucalyptus globulus	0.855	0.844	0.850	77
Eucalyptus polyanthemos	0.890	1.000	0.942	113
Ficus benjamina	0.819	0.747	0.782	91
Ficus carica	0.690	0.829	0.753	105
Ficus microcarpa nitida	0.977	0.992	0.984	1,982
Jacaranda mimosifolia	0.863	0.729	0.790	432
Lagerstroemia indica	0.870	0.918	0.893	633
Pinus canariensis	0.786	0.845	0.814	561
Pinus contorta	0.790	0.837	0.813	135
Pinus coulteri	0.660	0.761	0.707	46
Pinus halepensis	0.854	0.747	0.797	391
Pinus pinea	0.759	0.524	0.620	42
Platanus racemosa	0.964	0.953	0.959	171
Quercus agrifolia	0.907	0.921	0.914	4,046
Sequoia sempervirens	0.806	0.933	0.865	89
Washingtonia filifera	0.919	0.804	0.858	255
Washingtonia robusta	0.906	0.875	0.890	2,596

Figure 6. Confusion Matrix for 20 tree species in Altadena.



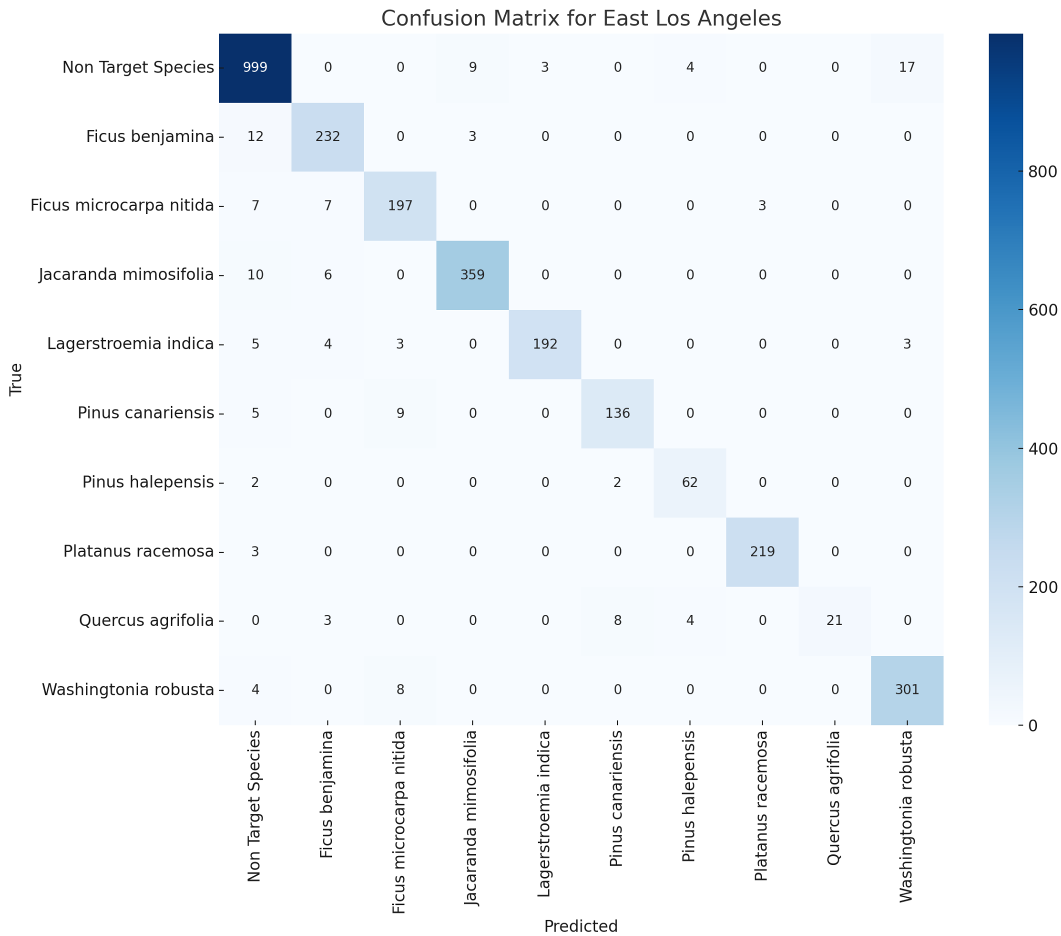
In Altadena, for *Eucalyptus polyanthemos*, the model had perfect Producer's Accuracy (1.000), which means it identified all instances of this species correctly. The User's Accuracy for this species was also high (0.890), implying that when the model predicted an instance to be *Eucalyptus polyanthemos*, it was correct about 89.0% of the time. For *Pinus pinea*, the model had the lowest Producer's Accuracy (0.524), indicating that it identified only about 52.4% of all actual *Pinus pinea* instances correctly. This implies a high False Negative rate for this species. For *Citrus limon*, the model had the lowest User's Accuracy (0.648), suggesting that when the model predicted an instance to be *Citrus limon*, it was correct only 64.8% of the time. This implies a high False Positive rate for this species.

The overall accuracy of the model in Altadena was approximately 0.889 for both User's Accuracy and Producer's Accuracy, which indicates a balanced performance in terms of False Positives and False Negatives. The species with the highest support (i.e., the most instances in the dataset) is *Quercus agrifolia*, with a count of 4,046 trees in the dataset. This species also had high User’s and Producer’s Accuracies, suggesting the model performed well on this species. The model appeared to perform well overall, with many species showing high User's Accuracy, Producer's Accuracy, and F1 scores. However, there were a few species (e.g., *Pinus pinea*) where the model's performance could be improved.

Table 2. Accuracy Assessment for nine tree species in East Los Angeles.

Species	User's Accuracy	Producer's Accuracy	F1-Score	Support
Non Target Species	0.954	0.968	0.961	1,032
Ficus benjamina	0.921	0.939	0.930	247
Ficus microcarpa nitida	0.908	0.921	0.914	214
Jacaranda mimosifolia	0.968	0.957	0.962	375
Lagerstroemia indica	0.985	0.928	0.955	207
Pinus canariensis	0.932	0.907	0.919	150
Pinus halepensis	0.886	0.939	0.912	66
Platanus racemosa	0.986	0.986	0.986	222
Quercus agrifolia	1.000	0.583	0.737	36
Washingtonia robusta	0.938	0.962	0.950	313

Figure 7. Confusion Matrix for nine tree species in East Los Angeles.



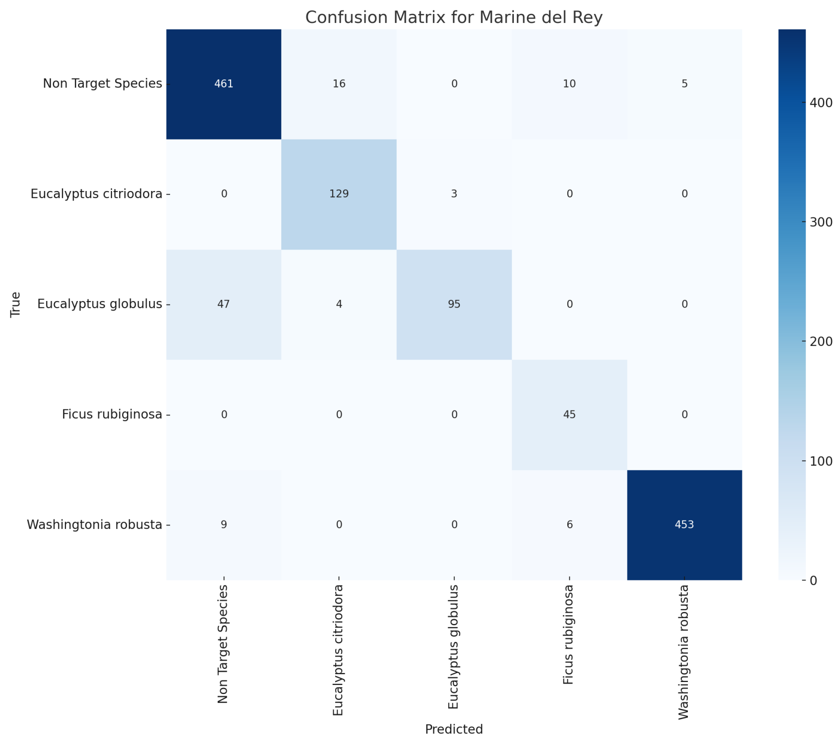
For East Los Angeles, the model performed well for most species, with both User's Accuracy and Producer's Accuracy often above 90%. However, for *Quercus agrifolia*, the User's Accuracy was perfect (1.000) but the Producer's Accuracy was quite low (0.583). This means that all instances predicted as *Quercus agrifolia* were indeed *Quercus agrifolia* (no false positives), but out of all actual instances of *Quercus agrifolia*, the model only correctly identified approximately 58% of them (high false negatives). Interestingly, there was confusion at this site between *Quercus agrifolia* and the two *Pinus* species (*P. canariensis* and *P. halepensis*).

Platanus racemosa showed excellent performance with both User's and Producer's Accuracy close to 1 (0.986). This suggests that the model was both precise and robust for this species. *Ficus benjamina*, *Ficus microcarpa nitida*, and *Jacaranda mimosifolia* showed good performance with both User's and Producer's Accuracy over 0.9. However, *Pinus halepensis* had relatively lower User's Accuracy (0.886), indicating a higher false-positive rate for this species. The group with the highest support (i.e., the most instances in the dataset) was "Non Target Species", with a count of 1,032.

Table 3. Accuracy Assessment for four tree species in Marina del Rey.

Species	User's Accuracy	Producer's Accuracy	F1-Score	Support
Non Target Species	0.892	0.937	0.914	492
Eucalyptus citriodora	0.866	0.977	0.918	132
Eucalyptus globulus	0.969	0.651	0.779	146
Ficus rubiginosa	0.738	1.000	0.849	45
Washingtonia robusta	0.989	0.968	0.978	468

Figure 8. Confusion Matrix for four tree species in Marina del Rey.



In Marina del Rey, for *Eucalyptus globulus*, User's Accuracy was relatively high at 0.969, indicating that when the model predicted this species, it was correct about 96.9% of the time. However, the Producer's Accuracy was only 0.651, meaning the model only correctly identified 65.1% of all actual *Eucalyptus globulus* instances. This implies a high rate of False Negatives for this species.

Ficus rubiginosa had perfect Producer's Accuracy of 1.000, indicating that the model identified all instances of this species correctly. However, the User's Accuracy was relatively low at 0.738, implying that when the model predicts an instance to be *Ficus rubiginosa*, it was correct only 73.8% of the time. This suggests a high rate of False Positives for this species. The model generally confused *Ficus rubiginosa* with non-target species and *Washingtonia robusta*.

Washingtonia robusta had both high User's Accuracy (0.989) and Producer's Accuracy (0.968), suggesting the model performed very well in identifying this species with both low False Positives and False Negatives.

3.3 Mapping Results

Figures 9 through 11 represent a species composition map, color-coded by tree species for all three pilot sites. All figures visualize model results of tree distribution through polygon data. The data is color-coded by species prediction and this methodology ensures an opportunity to visualize tree species prediction across all study sites (Fig. 9-11).

Figure 9. Classified tree species map for public street trees in Altadena, CA. A total of 19 species were classified using NAIP imagery and a random forest supervised classification in Google Earth Engine.

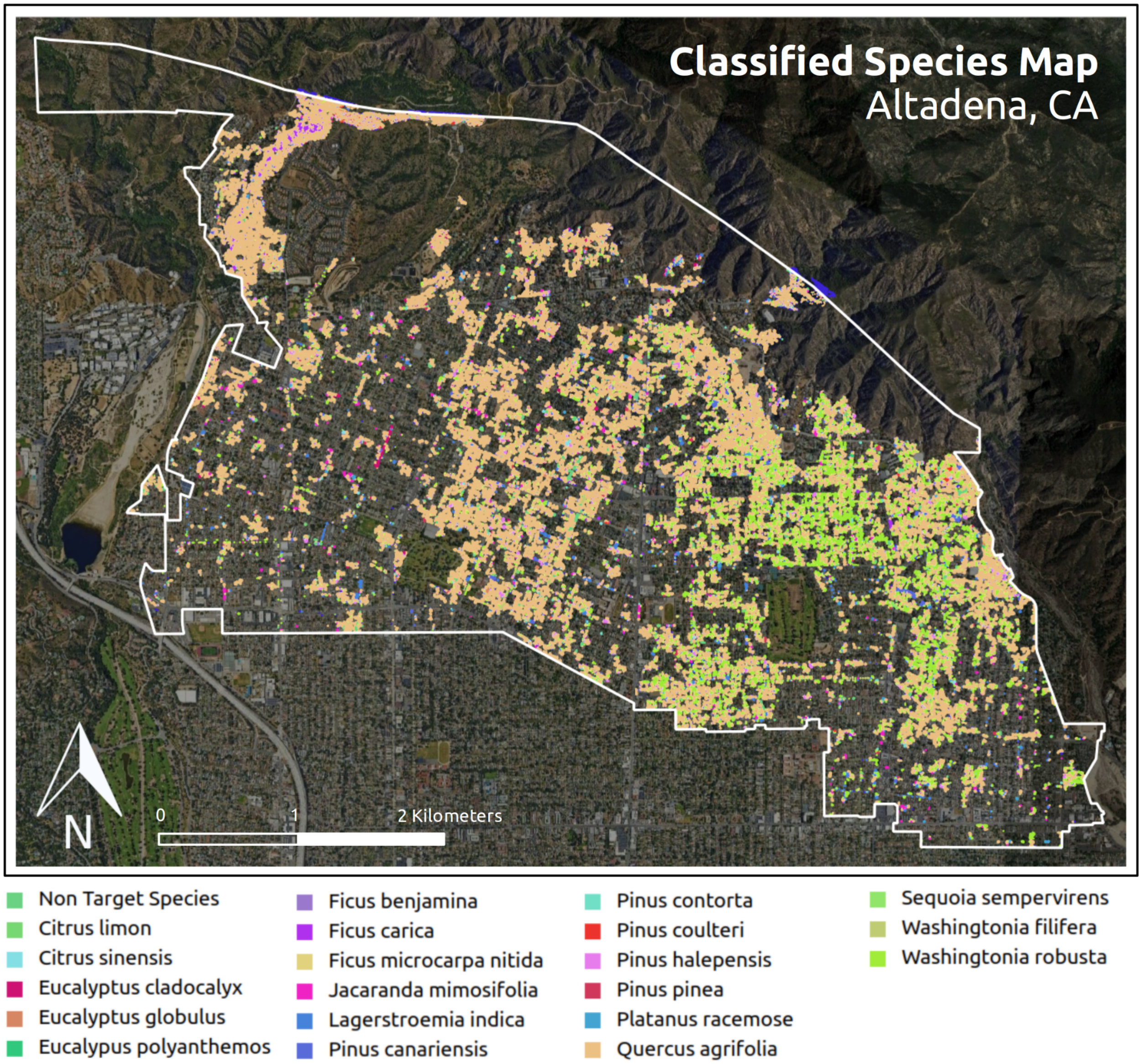


Figure 10. Classified tree species map for public street trees in East Los Angeles, CA. A total of nine species were classified using NAIP imagery and a random forest supervised classification in Google Earth Engine.

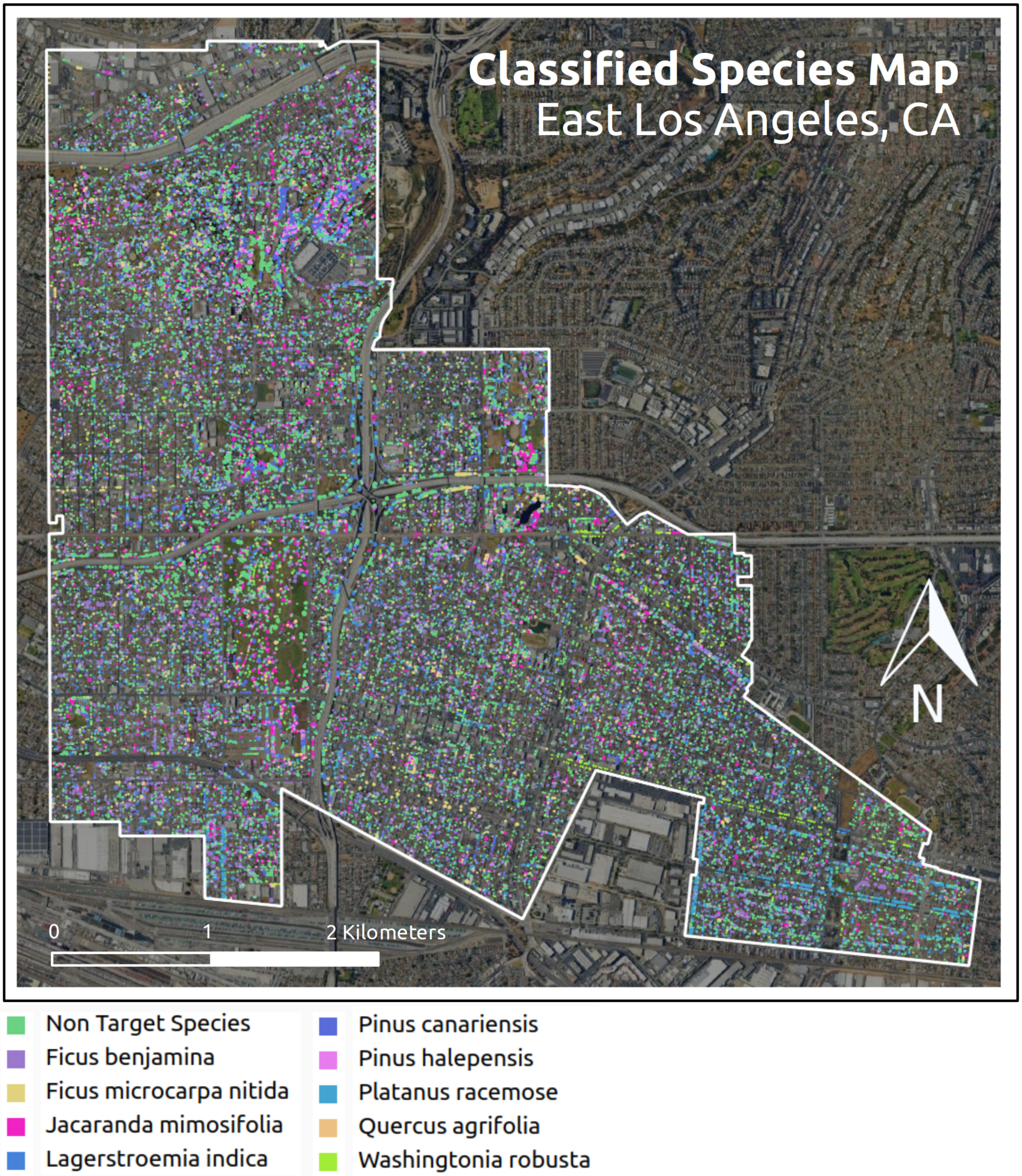
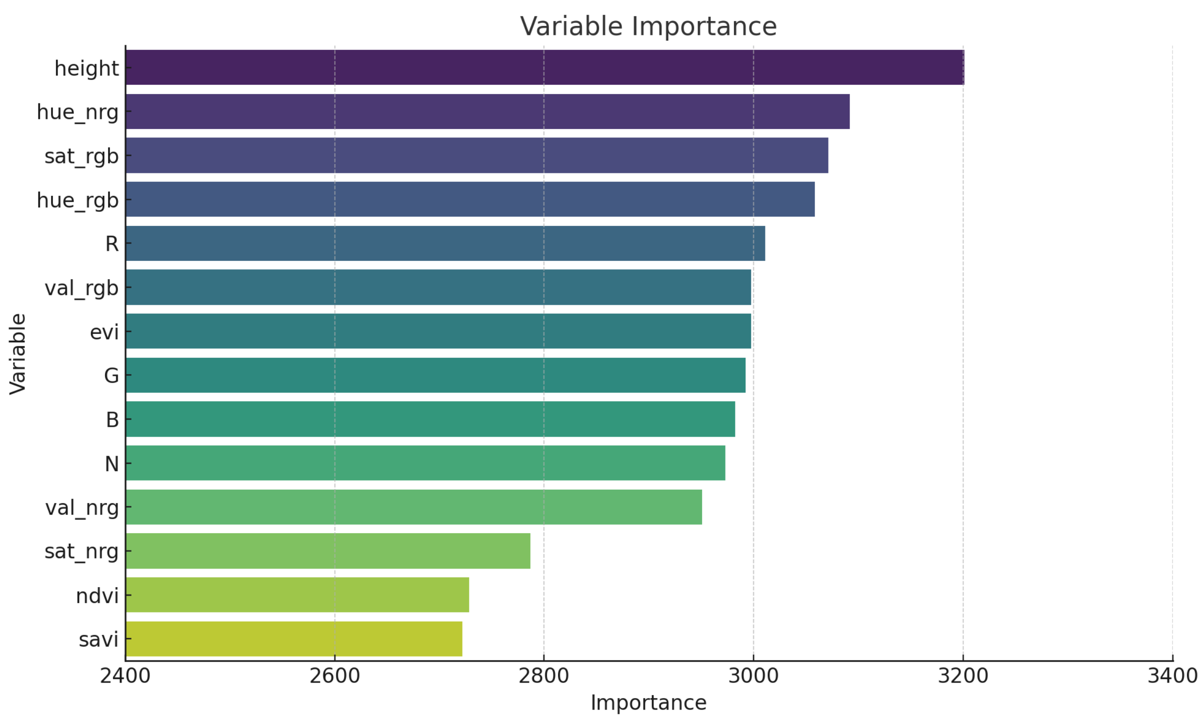


Figure 11. Classified tree species map for public street trees in Marina del Rey, CA. A total of four species were classified using NAIP imagery and a random forest supervised classification in Google Earth Engine.

In addition to the classification results, I computed feature importance to better understand which input variables were most influential in the species classification process (Fig. 12). Feature importance provides insights into the relative importance of each feature in making accurate predictions. In the context of this study, the height variables and HSV color space variables emerged as the most important contributors to the differentiation of species. Conversely, the NDVI and SAVI, which are primarily used to assess vegetation productivity, proved to be the least important in species classification.

Figure 12. Feature importance using Random Forest to identify 22 of tree species in Altadena, East Los Angeles, and Marina del Rey.



4

Discussion

4.1 Feature Importance

Understanding which features from my remote sensing imagery were most important for distinguishing species using a Random Forest model in Google Earth Engine is critical for fine tuning model hyper parameters and exploratory future work. The most important feature for identifying individual public street trees was tree height derived from my LiDAR data. This structural information is critical when separately species classes and corroborates findings from previous studies demonstrating an ability to increase classification results when including LiDAR with spectral information (Alonzo et al., 2014; Dian et al., 2016; Liu et al., 2017). HSV bands were unexpectedly of greater importance for classification than the original RGB spectral information or the vegetation indices that are commonly used in remote sensing studies of vegetation.

4.2 Accuracy Assessment

The accuracy assessments, particularly User's and Producer's accuracy, are critical in assessing the performance of a classification model. They allow us to understand not only how often the classifier is correct, but also what kinds of mistakes it was making (e.g., false positives versus false negatives). This can help guide efforts to improve the system and assess its suitability for future applications, or for its use in targeted studies for specific species. In general, the accuracy of the model was very good: for Altadena, the overall User's Accuracy and Producer's Accuracy are both 0.889; for East Los Angeles, they are both 0.919; and for Marina del Rey, they are 0.859

and 0.869 respectively. This suggests that when using this model to classify species, it gives accurate and reliable results.

In comparison with other studies that also used LiDAR and multispectral sensors, the best performing model used WorldView 2 and 3 imagery to classify eight species with an overall accuracy of 0.820 in St. Louis, MO, USA (Hartling et al., 2019). My model not only outperformed this study across all three pilot sites, but it is also the only study to use NAIP imagery and HSV color space as input variables in urban tree species classification. The higher spatial resolution of NAIP in addition to the variable importance of HSV demonstrated in this study likely explain the higher overall accuracies and lend important considerations for future studies.

Table 4. Articles that use remote sensing to identify urban trees to species in chronological order and total number of species examined, mean number of tree individuals per species and total number of individuals included in study, overall accuracy, sensors, and location.

Source	Species (#)	Mean Indiv. per Species	Total Indiv.	Overall Accuracy	Sensors	Location
This study	20	603	12,060	89%	LARIAC/ NAIP	Altadena, CA
	9	203	1,830	92%		East L.A., CA
	4	198	791	86%		Marina del Rey, CA
Pu & Landry 2012	7	81	573	63%	IKONOS/ WV2	Tampa, FL
Zhou et al., 2016	11	63	697	73%	Digital camera	Shanghai, China
Pu et al., 2018	7	111	777	61%	Pléiades	Tampa, FL
Hartling et al., 2019	8	194	1,552	82%	WorldView 2/3	St Louis, MO

Fang et al., 2020	19	868	16,486	61%	WorldView 3	Washington DC
Mesquita et al., 2020	16	50	1,329	70%	WorldView 2	Teresina, Brazil

WV = World View

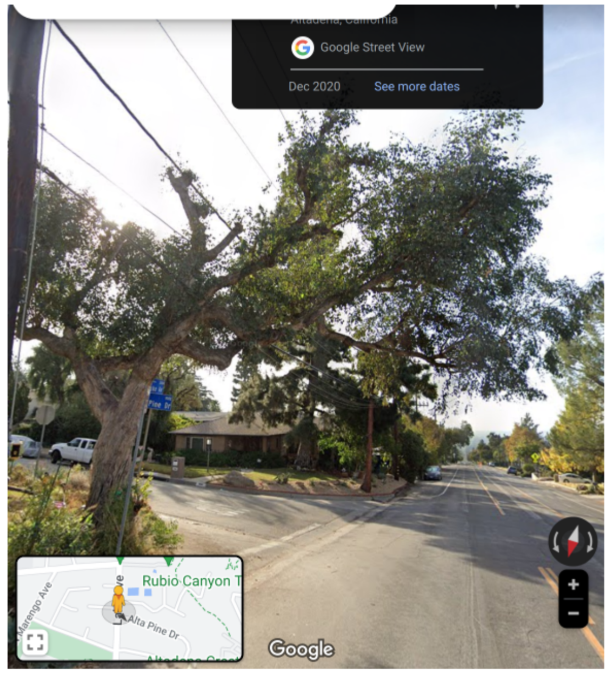
However, there is still room for improvement in certain areas. For instance, the model has trouble recognizing *Pinus pinea* in Altadena (Producer's Accuracy of 0.524) as well as *Ficus rubiginosa* in Marina del Rey (User's Accuracy of 0.738). This indicates that either more data or a different model may be needed to improve the accuracy of the classifier. Furthermore, while it is possible to scale this model across different neighborhoods, the results are not always consistent. This suggests that one model cannot be applied universally, and must incorporate local characteristics, like species present or tree maturity, to perform adequately.

4.3 Mapping Output

In this study, I have demonstrated that Random Forest in Google Earth Engine can classify public street trees with relative success with an average overall accuracy of 89% across my three pilot sites. However, the urban forest in Los Angeles is incredibly diverse, and although my training data set provided a well-documented and representative sample of trees planted in the public right-of-way, it does not capture the species variety on private property. Without geolocated and validated ground data for private lawns and commercial properties, my model lacks the ability to adequately classify and map privately managed trees. Mapping my results shows the discrepancy between model performance with public street trees and private trees (Fig. 13-15).

Figure 13. Mapping correctly classified public street tree crowns with Google Street View validation.

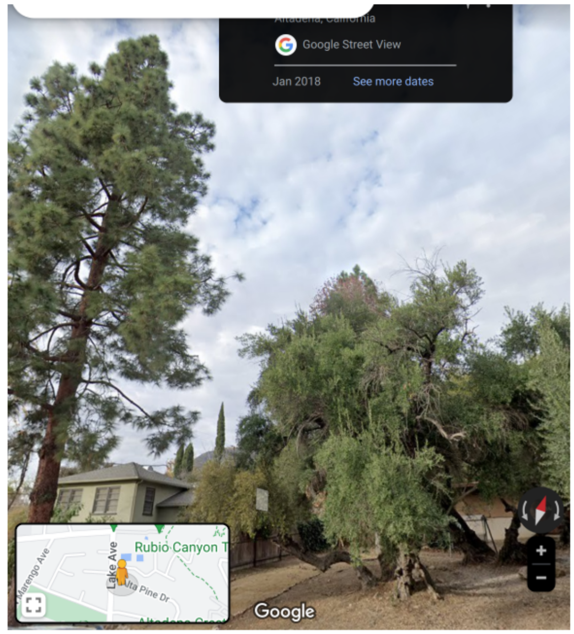
Public Street Tree Classification:
Eucalyptus gongylocarpa (non-target species)



Google Street View

Figure 14. Mapping incorrectly classified private property tree crowns with Google Street View validation.

Private Property Tree Classification: *Quercus agrifolia*



Google Street View

Figure 15. Visualization of poor model performance across trees on private property.

Predicted Species Map: Over classification of the two most abundant training species. Magenta symbolizes *Quercus agrifolia* and lavender is *Washingtonia robusta*

4.4 Future Work

The performance of the model appears to be biased towards the largest training samples by height, delivering good performance for street trees, which are the most representative sample in our dataset. However, the performance falters significantly for private trees, which are non-representative in our dataset (Hutt-Taylor and Ziter, 2022). To combat this, I propose a few adaptations:

1. Train a different model on new data that includes more instances of private trees.

2. Build profiles for unknown crowns and compare them to known crowns. This will yield insights to whether it is possible to distinguish between species with high confidence without needing to acquire accurate data for private trees.
3. Switch to a binary classification approach, where each tree species is distinguished from all others (i.e., oak vs everything else, palm vs everything else, etc.).

The applications of these adaptations are broad, including monitoring public tree health and longitudinal studies (Hart and Veblen, 2015; Byer and Jin, 2017; Furniss et al., 2020; Furniss et al., 2020). With further development, this method could be used to support smart city initiatives such as tree mapping and urban forestry management (Xiao and McPherson, 2005; Timilsina et al., 2020). This approach has the potential to improve our understanding of the relationship between landscape elements and their ecological role in improving urban health, as well as helping cities monitor and protect their unique ecosystems (Pretzsch et al., 2017; Zhu et al., 2019; Rodman et al., 2021; Cavender-Bares et al., 2022). Furthermore, the impending availability of new LARIAC data in 2024 is expected to further refine and upgrade the model. This progress, however, brings to light certain limitations. The classification of private trees, for instance, may necessitate additional methodologies, entailing a greater demand for resources and efforts to boost model accuracy in such contexts.

Nevertheless, the application of these methods must contend with the inherent disparities in data access and quality across different urban areas. Cities like Los Angeles County, equipped with ample resources and advanced technologies, are at the forefront of sophisticated data collection and analysis. Conversely, smaller municipalities often grapple with limitations in resources, resulting in data that may be less detailed or accurate. This divergence in capabilities highlights a crucial challenge: the potential for uneven application of urban forestry management and policy implementation across various urban settings. While the advancements in urban forestry management methodologies are promising, their implementation is nuanced and requires consideration of the disparities in data access and technological resources among different municipalities. This understanding is critical for ensuring that urban forestry management and smart city initiatives are effectively and equitably applied across diverse urban landscapes.

Contemporary urban forest planning is also shaped by the history of land use and land cover in various ways. The lasting impacts of past land uses, known as legacy effects, directly influence the current infrastructure needed for tree planting, such as soil quality and the presence of brownfield sites. These factors necessitate a careful selection and management of tree species suitable for the specific conditions of each site. Historical biodiversity is also a critical factor in current decisions about species selection and biodiversity conservation. Urban planners often focus on native species that are well-suited to the local environment to maintain ecological balance. For instance, in Los Angeles, the diverse urban forest seen today is a relatively recent development. Aerial photographs from the 1930s and 1950s compared to present-day images, as shown in Appendix A, illustrate the evolution of the urban forest canopy in the city. Moreover, the development history of an urban area determines its present structure, affecting how urban forest planning can be approached. In older, densely populated neighborhoods in Los Angeles, limited space requires creative greening strategies, whereas newer or less developed areas might provide more space for extensive tree planting.

5 Conclusion

My research has focused on the application of machine learning algorithms, specifically the random forest algorithm, for categorizing tree species and assessing individual health status. My model showed strong performance for the most common species in the dataset, such as *Quercus agrifolia* and *Washingtonia robusta*. Conversely, instances of uncommon or private tree species were less accurately predicted, indicating the model's bias towards species with larger training samples. To enhance model performance, I suggest several strategies, including expanding the dataset to include more instances of private trees, and shifting to a binary classification approach. The model's benefits extend to monitoring public tree health, conducting longitudinal studies, and predicting vulnerabilities to climate change, drought, and natural disasters. Ultimately, this research equips relevant stakeholders with the necessary tools for effective and efficient urban forest management.

6 References

Alonzo, M., Roth, K., & Roberts, D. (2013). Identifying Santa Barbara's urban tree species from AVIRIS imagery using canonical discriminant analysis. *Remote Sensing Letters*, 4(5), 513-521.

Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J., ... & Vincent, G. (2019). A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests. *Remote Sensing*, 11(9), 1086.

Aval, J., Demuynck, J., Zenou, E., Fabre, S., Sheeren, D., Fauvel, M., ... & Briottet, X. (2018). Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 197-210.

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

Bodnaruk, E.W., Kroll, C.N., Yang, Y., Hirabayashi, S., Nowak, D.J., & Endreny, T.A. (2017). Where to plant urban trees? A spatially explicit methodology to explore ecosystem service tradeoffs. *Landscape and Urban Planning*, 157, 457-467.

Branson, S., Wegner, J. D., Hall, D., Lang, N., Schindler, K., & Perona, P. (2018). From Google Maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135, 13-30.

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

Byer, S., & Jin, Y. (2017). Detecting drought-induced tree mortality in Sierra Nevada forests with time series of satellite data. *Remote Sensing*, 9(9), 929.

Caughlin, T. T., et al., (2019). High-resolution remote sensing data as a boundary object to facilitate interdisciplinary collaboration. Collaboration Across Boundaries for Social-Ecological Systems Science. *Palgrave Macmillan*, 295-326.

Cavender-Bares, J., et al., (2022). Integrating Remote Sensing with Ecology and Evolution to Advance Biodiversity Conservation." *Nature Ecology & Evolution*, 6, 506–519.

Coleman, R. W., Stavros, N., Yadav, V., & Parazoo, N. (2020). A Simplified Framework for High-Resolution Urban Vegetation Classification with Optical Imagery in the Los Angeles Megacity. *Remote Sensing*, 12(15), 2399.

Dian, Y., Pang, Y., Dong, Y., & Li, Z. (2016). Urban tree species mapping using airborne LiDAR and hyperspectral data. Journal of the Indian Society of Remote Sensing, 44, 595-603.

Dong, C., et al., (2023). Drought-Vulnerable Vegetation Increases Exposure of Disadvantaged Populations to Heatwaves under Global Warming: A Case Study from Los Angeles. Sustainable *Cities and Society*, 93, 104488.

Dyson, K., Nicolau, A. P., Clinton, N., & Saah, D. (2023). Advanced Pixel-Based Image Transformations. In *Cloud-Based Remote Sensing with Google Earth Engine: Fundamentals and Applications* (pp. 169-194). Cham: Springer International Publishing.

Fang, F., et al., (2020). Street Tree Health from Space? An Evaluation Using Worldview-3 Data and the Washington D.C. Street Tree Spatial Database. *Urban Forestry & Urban Greening*, 49, 126634.

Fassnacht, F. E., Hartig, F., Latifi, H., Berger, C., Hernández, J., Corvalán, P., & Koch, B. (2014). Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sensing of Environment*, 154, 102-114.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., ... & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.

Ferreira, M. P., Zortea, M., Zanotta, D. C., Shimabukuro, Y. E., & de Souza Filho, C. R. (2016). Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. *Remote Sensing of Environment*, 179, 66-78.

Furniss, T. J., Kane, V. R., Larson, A. J., & Lutz, J. A. (2020). Detecting tree mortality with Landsat-derived spectral indices: Improving ecological accuracy by examining uncertainty. *Remote Sensing of Environment*, 237, 111497.

Furniss, T. J., et al., (2020). Detecting Tree Mortality with Landsat-Derived Spectral Indices: Improving Ecological Accuracy by Examining Uncertainty. *Remote Sensing of Environment*, 237, 111497.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern recognition letters*, 27(4), 294-300.

Gillner, S., Vogt, J., Tharang, A., Dettmann, S., & Roloff, A. (2015). Role of street trees in mitigating effects of heat and drought at highly sealed urban sites. *Landscape and Urban Planning*, 143, 33-42.

Gillespie, T. W., de Goede, J., Aguilar, L., Jenerette, G. D., Fricker, G. A., Avolio, M. L., ... & Pataki, D. E. (2017). Predicting tree species richness in urban forests. *Urban Ecosystems*, 20(4), 839-849.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18-27.

Guo, Q., Kelly, M., Gong, P., & Liu, D. (2007). An object-based classification approach in mapping tree mortality using high spatial resolution imagery. *GIScience & Remote Sensing*, 44(1), 24-47.

Han, H., Han, C., Lan, T., Huang, L., Hu, C., & Xue, X. (2020). Automatic shadow detection for multispectral satellite remote sensing images in invariant color spaces. *Applied Sciences*, 10(18), 6467.

Hart, S. J., & Veblen, T. T. (2015). Detection of spruce beetle-induced tree mortality using high-and medium-resolution remotely sensed imagery. *Remote Sensing of Environment*, 168, 134-145.

Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M., & Carron, J. (2019). Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning. *Sensors*, 19(6), 1284.

Ozdemir, H. (2019). Mitigation impact of roadside trees on fine particle pollution. *Science of the Total Environment*, 659, 1176-1185.

Huang, C., et al., (2019). Remote Sensing of Forest Die-off in the Anthropocene: From Plant Ecophysiology to Canopy Structure. *Remote Sensing of Environment*, 231,111233.

Huang, S., et al., (2020). A Commentary Review on the Use of Normalized Difference Vegetation Index (NDVI) in the Era of Popular Remote Sensing. *Journal of Forestry Research*, 32, 1–6.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 25(3), 295-309.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment*, 83(1-2), 195-213.

Hutt-Taylor, K., & Ziter, C. D. (2022). Private trees contribute uniquely to urban forest diversity, structure and service-based traits. *Urban Forestry & Urban Greening*, 78, 127760.

Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sensing*, 4(9), 2661-2693.

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*, 8(3), 166.

Isager, L., and Niels, H. B. (2007). Combining remote sensing and anthropology to trace historical land-use changes and facilitate better landscape management in a sub-watershed in North Thailand. *Landscape Research*, 32.2, 147-169.

Jackson, R. D., Slater, P. N., & Pinter Jr, P. J. (1983). Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres. *Remote sensing of environment*, 13(3), 187-208.

Jones Jr, J. W. (2011). Using LIDAR Data and Geographical Information System (GIS) Technology to Assess Municipal Street Tree Inventories. Mississippi State University.

Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, No. 2, pp. 1137-1145).

Lausch, A., et al., (2016). Understanding Forest Health with Remote Sensing -Part I—a Review of Spectral Traits, Processes and Remote-Sensing Characteristics. *Remote Sensing*, 8, 1029.

Lee, H., Mayer, H., & Chen, L. (2016). Contribution of trees and grasslands to the mitigation of human heat stress in a residential district of Freiburg, Southwest Germany. *Landscape and Urban Planning*, 148, 37-50.

Lim, J., Kim, K. M., & Jin, R. (2019). Tree species classification using hyperion and sentinel-2 data with machine learning in South Korea and China. *ISPRS International Journal of Geo-Information*, 8(3), 150.

Lin, Y., & Hyyppä, J. (2016). A comprehensive but efficient framework of proposing and validating feature parameters from airborne LiDAR data for tree species classification. International *Journal of Applied Earth Observation and Geoinformation*, 46, 45-55.

Liu, D., Jiang, Y., Wang, R., & Lu, Y. (2023). Establishing a citywide street tree inventory with street view images and computer vision techniques. *Computers, Environment and Urban Systems*, 100, 101924.

Los Angeles Region Imagery and Data Acquisition Consortium (LARIAC). (2016). LARIAC4 LiDAR Imagery (LiDAR). Los Angeles, CA: LARIAC. https://lariac-lacounty.hub.arcgis.com/pages/lariac4-documents-data.

Love, N.L., Nguyen, V., Pawlak, C., Pineda, A., Reimer, J.L., Yost, J.M., Fricker, G.A., Ventura, J.D., Doremus, J.M., Crow, T. and Ritter, M.K., 2022. Diversity and structure in California's urban forest: What over six million data points tell us about one of the world's largest urban forests. *Urban Forestry & Urban Greening*, 74, p.127679.

Meddens, A. J., Hicke, J. A., Vierling, L. A., & Hudak, A. T. (2013). Evaluating methods to detect bark beetle-caused tree mortality using single-date and multi-date Landsat imagery. *Remote Sensing of Environment*, 132, 49-58.

Mesquita, M. R., Agarwal, S., de Morais Lima, L. H. G., Soares, M. R. A., Barbosa, D. B. E. S., Silva, V. C., ... & Costa, C. H. N. (2022). The use of geotechnologies for the identification of the urban flora in the city of Teresina, Brazil. *Urban Ecosystems*, 1-12.

Mosin, V., Aguilar, R., Platonov, A., Vasiliev, A., Kedrov, A., & Ivanov, A. (2019). Remote sensing and machine learning for tree detection and classification in forestry applications. *In Image and Signal Processing for Remote Sensing*, 11155, 111550F).

Mostafa, Y., & Abdelhafiz, A. (2017). Accurate shadow detection from high-resolution satellite images. *IEEE Geoscience and Remote Sensing Letters*, 14(4), 494-498.

Okin, G. S., et al., (2018). The Impact of Drought on Native Southern California Vegetation: Remote Sensing Analysis Using Modis-Derived Time Series. Journal of Geophysical Research: *Biogeosciences*, 123, 1927–1939.

Pataki, D. E., McCarthy, H. R., Gillespie, T., Jenerette, G. D., & Pincetl, S. (2013). A trait‐based ecology of the Los Angeles urban forest. *Ecosphere*, 4(6), 1-20.

Pincetl, S., Gillespie, T., Pataki, D. E., Saatchi, S., & Saphores, J. D. (2013). Urban tree planting programs, function or fashion? Los Angeles and urban tree planting campaigns. *GeoJournal*, 78(3), 475-493.

Pretzsch, H., Biber, P., Uhl, E., Dahlhausen, J., Schütze, G., Perkins, D., … & Chavanne, A. (2017). Climate change accelerates growth of urban trees in metropolises worldwide. *Scientific Reports*, 7(1), 1-10.

Pu, R., & Landry, S. (2012). A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment*, 124, 516-533.

Pu, R., Landry, S., & Yu, Q. (2018). Assessing the potential of multi-seasonal high resolution Pléiades satellite imagery for mapping urban tree species. *International Journal of Applied Earth Observation and Geoinformation*, 71, 144-158.

Pu, R., 2021. Mapping tree species using advanced remote sensing technologies: a state-of-the-art review and perspective. *Journal of Remote Sensing*, 2021, p.26.

Purevdorj, T. S., Tateishi, R., Ishiyama, T., & Honda, Y. (1998). Relationships between percent vegetation cover and vegetation indices. *International journal of remote sensing*, 19(18), 3519-3535.

QGIS.org, 2023. QGIS Geographic Information System. QGIS Association. http://www.qgis.org

Rodman, K. C., Andrus, R. A., Veblen, T. T., & Hart, S. J. (2021). Disturbance detection in landsat time series is influenced by tree mortality agent and severity, not by prior disturbance. *Remote Sensing of Environment*, 254, 112244.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS journal of photogrammetry and remote sensing*, 67, 93-104.

Rouse Jr, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation* (No. NASA-CR-132982).

Roussel J, Auty D, Coops NC, Tompalski P, Goodbody TR, Meador AS, Bourdon J, de Boissieu F, Achim A (2020). "lidR: An R package for analysis of Airborne Laser Scanning (ALS) data." *Remote Sensing of Environment*, 251, 112061. ISSN 0034-4257, doi:10.1016/j.rse.2020.112061, https://www.sciencedirect.com/science/article/pii/S0034425720304314.

Roussel J, Auty D (2023). Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R package version 4.0.3.

Salmond, J. A., Tadaki, M., Vardoulakis, S., Arbuthnott, K., Coutts, A., Demuzere, M., ... & Wheeler, B. W. (2016). Health and climate related ecosystem services provided by street trees in the urban environment. *Environmental Health*, 15(1), 95-111.

Schwantes, A. M., Swenson, J. J., & Jackson, R. B. (2016). Quantifying drought-induced tree mortality in the open canopy woodlands of central Texas. *Remote Sensing of Environment*, 181, 54-64.

Tian, H. (2006). *Variance estimation of the cross validation estimator of the generalization error*. Columbia University.

Timilsina, S., et al., (2020). Mapping Urban Tree Cover Changes Using Object-Based Convolution Neural Network (OB-CNN). *Remote Sensing*, 12, 3017.

United States Department of Agriculture (USDA), National Agricultural Imagery Program (NAIP). (2016). Imagery collected during 2016 agricultural season. Washington, D.C.: USDA's Farm Service Agency. https://developers.google.com/earth-engine/datasets/catalog/USDA_NAIP_DOQQ#citations.

Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. *New Phytologist*, 186(4), 795-816.

Xiao, Q., Ustin, S. L., & McPherson, E. G. (2004). Using AVIRIS data and multiple-masking techniques to map urban forest tree species. *International Journal of Remote Sensing*, 25(24), 5637-5654.

Xiao, Q., and McPherson, E. G. (2005). Tree Health Mapping with Multispectral Remote Sensing Data at UC Davis, California. *Urban Ecosystems*, 8, 349–361.

Yang, G., Zhao, Y., Li, B., Ma, Y., Li, R., Jing, J., & Dian, Y. (2019). Tree Species Classification by Employing Multiple Features Acquired from Integrated Sensors. *Journal of Sensors*, 2019.

Zhang, C., Zhou, Y., & Qiu, F. (2015). Individual tree segmentation from LiDAR point clouds for urban forest inventory. *Remote Sensing*, 7(6), 7892-7913.

Zhang, R., Li, Q., Duan, K. F., You, S. C., Zhang, T., Liu, K., & Gan, Y. H. (2020). Precise Classification of Forest Species Based on Multi-source Remote-sensing Images. *Applied Ecology and Environmental Research*, 18(2), 3659-3681.

Zhou, T., Popescu, S. C., Lawing, A. M., Eriksson, M., Strimbu, B. M., & Bürkner, P. C. (2018). Bayesian and classical machine learning methods: a comparison for tree species classification with LiDAR waveform signatures. *Remote Sensing*, 10(1), 39.

Zhu, Zhe, et al., (2019). Understanding an urbanizing planet: Strategic directions for remote sensing. *Remote sensing of environment*, 228, 164-182.

Chapter 4.

"Eyes" on the street: Perspectives, promises, and the practice of geospatial technology in science applications and public engagement

Abstract

This chapter critically examines the intersection of geospatial technologies, public engagement, and science applications within the context of socio-environmental research and policy making. It dissects the inherent complexities of the transformation steered by the synthesis of technology and data science, focusing specifically on the curation and utilization of remote sensing and GIS data. The chapter underscores the roles and responsibilities of various stakeholders in shaping the outcomes of this data-driven process. These range from Data Producers who collect raw data and Data Scientists who interpret and model this information, to Data Users who apply the findings to new management practices and Impacted Persons who experience the real-world consequences of these interactions. I emphasize the often overlooked or marginalized voices of Impacted Persons. Recognizing the 'knowledge gap' that arises from the technical intricacy of geospatial data, this chapter discusses the necessity to democratize geospatial data knowledge. This chapter also emphasizes how boundary objects, such as maps, aid in enhancing communication and shared understanding between stakeholders from different disciplines. I conducted interviews and surveys along with past participatory science applications to provide recommendations for more inclusive engagement and cooperative processes in policy initiatives.

The overarching aim is to shape policy-making processes that are not only informed by data but are also inclusive, equitable, and responsive to the needs and perspectives of all stakeholders.

Keywords

Remote Sensing, GIS, Political Ecology, Socio-environmental Systems, Boundary Objects

1 Introduction

In recent years, remote sensing and Geographic Information Systems (GIS) technology advancements, from enhanced satellite imagery resolution to cutting-edge machine learning algorithms, have significantly transformed data analysis and visualization (Acevedo et al., 2008; Pettorelli et al., 2014). This evolution affects many stakeholders, including government agencies, businesses, researchers, and civil society organizations. Remote sensing data is pivotal in tracking urban, rural, and agricultural changes, offering a dynamic understanding of evolving scenarios (Galvin et al., 2001; Foster and Dunham, 2015; Peek et al., 2020). This data facilitates longitudinal studies that gauge real-time changes. Understanding the interplay between human decisions and environmental impacts necessitates data that spans appropriate spatial, temporal, and socioeconomic scales (Cumming et al., 2006). For nearly four decades, public access to moderate (10 – 60 m) to coarse (> 60 m) spatial and spectral resolution imagery has enriched insights into phenomena such as effects of logging on fire patterns (Dennis et al., 2005) or factors behind urban heat islands (Hulley et al., 2020). High-resolution data (< 10 m) further sharpens these insights by connecting broad issues, like drought, to hyperlocal, leaf-level observations, such as canopy water content (Ordway et al. 2021). The majority of existing studies primarily focus on empirical evidence, often overlooking the importance of diversity in stakeholder perspectives. In this study, a comparative analysis is conducted between the modern applications, promises, and practices in applied remote sensing and those discussed 25 years ago in *People and Pixels* (Rindfuss et al., 1998). The aim is to investigate if the recent intentional trends towards diversity, equity, and inclusion have succeeded in establishing more accessible standards and inclusive environments in the production of geospatial knowledge.

Emerging from the global COVID-19 pandemic and Black Lives Matter movements, agencies and institutions have experienced a shift in diversity, equity, and inclusion politics, aimed at overcoming the inequities and unfair power dynamics previously criticized by scholars, community members, and those underrepresented within the sciences. Analyzing the extent to which this transformation has been successful can yield valuable insights into the opportunities and challenges faced by remote sensing in achieving its full potential as an actionable, co-produced science. The primary focus lies beyond merely "socializing the pixel" which integrates spatially explicit social data within remote sensing applications. Instead, the focus will be on how socio-ecological insights have been incorporated into remote sensing applications in the past 25 years, celebrating the anniversary of this seminal concept in Land Use and Land Cover Change (LUCC) science (Geohegan et al., 1998). I also explore a possible redefinition of this concept, shifting from just spatially explicit social data to a broader examination of the sociopolitical components that form the pixel pipeline.

1.1 Applications of Geospatial Technology

The growing reliance on remote sensing data has led to an expanding array of applications and user groups. From ecological monitoring (Kerr and Ostrovsky, 2003; Svoray et al., 2013; Kugler et al., 2019) and disaster response (Peek et al., 2020) to urban planning and political decision-making (Hartmann et al., 2015; Heider et al., 2018; Niemiec et al., 2018), remote sensing provides unparalleled insights for understanding diverse processes and tackling various challenges. Contemporary remote sensing and GIS technologies have advanced beyond mere mapping and monitoring capabilities to offer sophisticated data analysis, integration, and visualization tools. These advancements enable interested parties to access, manipulate, and interpret geospatial data in novel ways. For example, recent developments in satellite imagery, drone technology, and machine learning have improved the accuracy and accessibility of remote sensing data, broadening its appeal to a broader range of users (Ustin and Gamon, 2010; Jones, 2011; Aval et al., 2018; Zhou et al., 2018; Aubry-Kientz et al., 2019).

Remote sensing technologies inherently have various ethical, economic, and environmental implications (Robbins, 2001; Robbins, 2003; Dennis et al., 2005). On the one hand, these technologies have the potential to vastly improve decision-making by providing accurate, timely, and comprehensive data (Walker and Peters, 2007). This can lead to more efficient resource allocation (Isager et al., 2007; Oestreich et al., 2020), better environmental management (Isager et al., 2007; Berger et al., 2019; Germestani et al., 2020), and increased economic opportunities (Walters et al., 2008). On the other hand, issues such as privacy, surveillance, and data ownership have emerged alongside the proliferation of remote sensing technologies (Rissman et al., 2017). In addition, unequal access to geospatial data and resources can exacerbate existing socioeconomic inequalities and power imbalances between stakeholders (Rindfuss et al., 1998; Chrisman, 1999; Lave et al., 2014; Adler et al. 2018; Arnaiz-Schmitz et al., 2018; Allan et al., 2022). Different stakeholder groups, including businesses, governments, and local communities, perceive and interact with remote sensing data differently. This, in sum, has significant implications for policy and decision-making processes. For example, remote sensing data have been co-opted by powerful entities to pursue their agendas, leading to the marginalization of local communities with limited access to or understanding of the data (Escobar, 1996; Rindfuss et al., 1998; Liverman and Cuesta, 2008; McGinnis and Ostrom, 2014; Envvist et al., 2018; Dinko and Nyantakyi-Frimpong, 2023).

1.2 Developing a Common Language: Boundary-spanning Terminology

In remote sensing and GIS data application, stakeholders—from Data Producers to Impacted Persons—each assume distinct, pivotal roles. Each group has unique perceptions, roles, and responsibilities. Understanding the dynamics between these groups and the potential issues that arise from their interactions is crucial to addressing broader concerns related to equity and power dynamics:

- Data Producers: These actors generate the foundational raw data that propels remote sensing and GIS technologies—the data's accuracy, quality, and accessibility hinge on their priorities and objectives.
- Data Scientists: As the intermediaries, my fellow data scientists and I, interpret and translate this raw data, utilizing diverse modeling techniques to render the data into a format that can be understood and applied by a broader audience. We ensure data integrity, select the most appropriate analytical methods, and continually recognize and communicate our work's intrinsic limitations.
- Data Users: This group harnesses the refined data to dictate policies, formulates decisions, and guides actions across sectors, from urban planning to public policy. Their task demands critically examining data and discerning potential biases, uncertainties, and ethical implications. However, late inclusion or inadequate engagement with these users can compromise data processing and outcomes. Such arrangements often demand added time for alignment and potentially extra resources for training.
- Impacted Persons: Often an expansive group, these individuals and ecosystems bear the direct or indirect consequences of the choices made by Data Users. Yet, many among them, especially those from marginalized communities, might find their voices silenced or overlooked entirely.
- Boundary Object: A tool or product, such as maps, facilitating effective communication often between Data Producers, Scientists, and Users. While scientists might derive measurements from these, stakeholders might utilize them for conceptualizing and marking significant activities.

Impacted Persons, who bear the brunt of changes and impacts due to such science, are often missing from the table. Too often, collaborative projects predominantly feature Data Producers, Scientists, and Users—representatives from government, academia, industry, and occasionally non-governmental organizations. Such discrepancies in access to and control over geospatial data can amplify prevailing power disparities, leading to skewed benefits and further marginalization.

The intricacy of geospatial data presents a formidable barrier to its understanding and utilization, particularly to those not well-versed in its technical aspects. Geospatial data, which includes remote sensing and GIS data, involves complex concepts and methodologies that require a specialized education to comprehend fully. This data is often multi-dimensional, time-variant, and spatially referenced, necessitating an understanding of advanced mathematical and statistical concepts, as well as proficiency in specialized software tools. This technical complexity can inadvertently lead to a 'knowledge gap.' Stakeholders, particularly the Data Users and Impacted Persons, may struggle to grasp the implications of the data, interpret the results accurately, or voice their concerns effectively. The education needed to understand geospatial data, therefore, becomes a hurdle in achieving meaningful and inclusive collaboration.

Furthermore, this gap can exacerbate existing power dynamics, with those possessing the technical knowledge potentially dominating the discourse. It underscores the necessity for efforts towards democratizing geospatial data knowledge—whether through accessible education initiatives, transparent communication, or intuitive visualization tools—that make the data understandable and usable to all stakeholders, irrespective of their technical acumen.

1.3 Situating My Work: Political Ecologies of GIScience and Technology

The aspiration to "socialize the pixel" continues to be an interdisciplinary research mainstay in applied remotes sensing (Geoghegan et al., 1998; Liverman and Cuesta, 2008; Tellman, 2018; Dinko and Nyantakyi-Frimpong, 2023). The processes underlying the technology and programming to model intricate, multi-temporal, and multiscalar socio-ecological changes wield significant power and responsibility. Given the profound effects of recent advancements on policy, planning, and ecosystem science, recognizing the agents co-producing SES research and comprehending the dynamics in decision-making become paramount (Keshkamat et al., 2012; Kramer et al., 2017; Perz 2020). As landscape classification methods gain institutional traction, their embedded complexities often go unchallenged. This acceptance necessitates ground truth observations and a perception that maps, as Zubrow (2003) noted, are authoritative representations. Thus, rigorous scrutiny of our methodologies, data collection techniques, and models, is essential to discern their broader societal ramifications.

The advent of remote sensing and GIS has transformed spatial data acquisition, analysis, and visualization. Despite the undeniable progress they've spurred, critical scholars around the turn of the 21st Century highlighted the potential for imbalanced power dynamics in applied remote sensing and GIS, which sometimes culminated in detrimental consequences (Escobar, 1996; Rindfuss et al., 1998; Liverman and Cuesta, 2008; McGinnis and Ostrom, 2014; Envvist et al., 2018; Dinko and Nyantakyi-Frimpong, 2023). With the recent and deliberate move towards diversity, equity, and inclusion (DEI) in scientific applications (Swartz et al., 2019), this chapter probes the endurance of these inequities within geospatial sciences and the adoption rate of more equitable standards in the field. As Escobar (1996) suggests, understanding the synergy between self-awareness and the world and between knowledge and its societal underpinnings might usher in a renewed biology and ecology. Introducing discursive analysis to materialistic understandings can blur the traditional boundaries separating nature, culture, and science (Escobar, 1996; Blue and Brierley, 2016).

1.3.1 Theoretical Frameworks

As geospatial technologies become more accessible and embedded in different aspects of our lives, it is essential to critically examine their broader sociopolitical implications. My trajectory as a remote sensing researcher underscores the urgency to harmonize dialogues among geospatial technology developers, users, and the communities they impact. The dominant structures in contemporary science and technology, laden with historical biases, necessitate a fresh perspective, challenging prevailing norms. Historically, remote sensing propagated Western ideologies, often packaged within the Global North's development paradigms (Olbrich, 2019; Vurdubakis and Rajāo, 2022). The persistent colonial undertones in Western science have metamorphosed from overt racial biases to subtle intellectual dominance over emerging economies (Roy, 2018). I take this one step further and suggest that scientific exploitation also extends into citizen science and co-production among local stakeholders. Within applied sciences, the unchecked authority of scientists often overshadows the local knowledge or indigenous expertise of those intimately involved in data collection and processing (Carroll et al. 2019; Carroll et al., 2020; Rammage et al., 2020: Rattling Leaf Sr. et al., 2020). Moreover, the technology's history is tainted by its misuse for diverse ulterior motives especially in resource-rich areas like the Amazon (Vurdubakis and Rajão, 2022), including using remote sensing science as a mechanism in the expulsion of local, indigenous, and traditional knowledge (Rajão, 2013). Convergent research naturally introduces a mosaic of stakeholders and various disciplinary traditions. A comprehensive evaluation demands recognizing all actors in knowledge production, understanding the dynamics of agency, and scrutinizing narratives driving decisions.

Sociology of Geospatial Technology: Delving into the sociology of remote sensing unveils how societal undercurrents mold the creation, interpretation, and deployment of GIS data (Geoghagen et al., 1998; Olbrich, 2019; Vurdubakis and Rajão, 2022). Societal norms and production contexts influence data interpretation, with consequent societal effects. An integrative sociological approach transcends the technical grasp of remote sensing and GIS, offering a holistic view of the nexus between technology and society and facilitating informed and inclusive policy decisions (Tellman, 2018).

In their seminal work, Geoghegan et al. (1998) posited the notion of "Socializing the Pixel and Pixelizing the Social" to articulate the intersection of geospatial technology and sociological paradigms. The central tenet of this theory underscores the reciprocal influence between the digital representation of geographical space (pixels) and the social dynamics inscribed within these spaces. The crux of the discourse pivots around two thematic axes:

"Socializing the Pixel" and "Pixelizing the Social," each diving into distinct yet interlinked facets of this interdisciplinary dialogue.

"Socializing the Pixel" underscores the imperative of embedding social paradigms within the digital abstraction of geographical space. It brings a departure from a quantitative appraisal of land-use and land-cover change, instead urging a more nuanced understanding fostered through the integration of social science insights. This strand of thought accentuates the contextual social processes that shape and are shaped by land-use dynamics, thereby enriching the interpretative lens through which geospatial data is analyzed and understood. Conversely, "Pixelizing the Social" entails the digital representation of social phenomena within the geographical space, bringing a nuanced understanding of social dynamics through a geospatial lens. This facet underscores the potential of geospatial technology in rendering visible the spatial manifestations of social phenomena, thereby fostering a more grounded understanding of social processes and their spatial correlates.

Fast forward nearly three decades, the resonance of this concept in applied remote sensing still rings true. The advances in geospatial technology have significantly enhanced the capacity to articulate and analyze the spatial dimension of social phenomena. Concurrently, the growing discourse in social science has continually informed the methodologies and interpretative frameworks employed in geospatial analysis. The synthesis of these domains has brought a more holistic understanding of land-use and land-cover change, embodying the ethos of "Socializing the Pixel and Pixelizing the Social." The discourse has evolved in contemporary research, spanning advancements in remote sensing technology and computational methodologies. The growing body of research underscores the ongoing relevance of this conceptual framework in understanding complex socio-ecological systems.

Socio-ecological Systems: SES, rooted in the human-nature dichotomy, probes intricate systems across spatial and temporal scales (Liu et al. 2007; Turner et al. 2016). It embraces interdisciplinary networks and diverse knowledge sources, emphasizing equitable solutions to global challenges while prioritizing environmental integrity. As represented by LENS, SES's collaborative ethos fosters dialogue among diverse stakeholders, raising critical questions about existing power dynamics and research objectives.

Applied remote sensing, with its ability to provide valuable data on ecosystem properties and socioeconomic metrics, finds a significant application within the SES framework. This technological intervention enables the capture of nuanced environmental dynamics, thus facilitating a more robust understanding of socio-ecological interactions. Particularly, remote sensing data, with its varying levels of spatiotemporal resolution, unveils novel attributes and interactions within these systems.

Within the SES framework, LENS represents a collaborative ethos that promotes dialogue among a diverse array of stakeholders. This dialogue is imperative for addressing power dynamics, ethical concerns, and aligning research objectives with broader societal and environmental goals. For example, the critique extends to data curation and its role as boundary objects in this discourse. Data, in its curated form, acts as a bridge, facilitating communication, understanding, and collaborative action among varied stakeholders. However, the process of data curation itself is imbued with power dynamics and ethical considerations. Who curates the data, how it is curated, and whose perspectives are represented or marginalized in this process, are critical questions that echo the broader concerns of justice, equity, and transparency within the SES framework. Thus, the nexus between applied remote sensing and SES is a rich, multidimensional interface. It embodies the potential for harnessing technological advancements in remote sensing to foster a deeper, more nuanced understanding of complex socio-ecological systems, while simultaneously challenging and re-evaluating existing power structures and ethical paradigms. Through this lens, remote sensing transcends its technical utility, morphing into a potent tool for social and environmental justice, as well as a catalyst for meaningful, collaborative engagement in addressing global challenges.

(Geo)Science and Technology Studies: Science and technology studies present robust frameworks delineating the nexus between technology, society, and culture. This research amplifies this dialogue by assessing knowledge and policy co-creation within remote sensing (Liverman et al., 1998; Dennis et al., 2005; Liverman and Cuesta, 2008; Haraway, 2016; Kugler et al., 2019). In synthesizing the Geo STS narrative, the sociological and policy co-creation dimensions within remote sensing are shown. They echo the STS ethos of dissecting the socio-technological dialogues and the co-evolution of society and technology. The intertwined pathways of knowledge creation, policy discourse, and technological innovation within remote sensing show the enduring relevance of the STS frameworks in applied geoscience. Through the

lens of STS, the socio-technological landscapes of remote sensing offer a fertile ground for academic discourse, policy deliberation, and societal engagement in the further development of geospatial technology and curation of spatial data.

Critical GIS & Remote Sensing: Robbins (2001) emphasizes that while remotely sensed data can shed light on complex systems, it cannot address foundational disagreements regarding the nature of the environment. Critical GIS literature underscores the need to confront power imbalances, ethical dilemmas, and justice concerns. This study extends this critique, examining how data curation impacts its role as boundary objects. Robbins (2001) also underscores a critical limitation of remotely sensed data, emphasizing that while it can help uncover complex environmental systems, it falters in addressing foundational disagreements regarding the nature of the environment. I extend this critique to data curation and its role as boundary objects, thus fostering a nuanced understanding of applied remote sensing within these socio-technological and ethical frameworks.

Local Ecological Knowledge: While indigenous populations manage over half of the Earth's terrestrial expanse, only 10% receive formal recognition by nation-states (Ramage et al., 2020; Rattling Leaf Sr., et al. 2020), emphasizing the historical discord between indigenous landscapes and colonialism. Carroll et al. (2019) query the prospects of data-driven futures for communities entrenched in data inequities. Although geographical traditions present certain overlaps, indigenous data sovereignty is a stark challenge in SES integration. Carroll et al. (2020) offer a potential solution through the CARE principles for Indigenous Data Governance, fostering responsible knowledge transfer while preserving indigenous data rights. Furthermore, aligning remote sensing with Local Ecological Knowledge (LEK) can amplify its potential. The symbiotic relationship between remotely sensed imagery and LEK provides a novel method already gaining traction among global indigenous communities (Rattling Leaf Sr., 2022). Ground-referenced data integration can elevate SES, combining technological tools like cloud computing, citizen science, and "humans-as-sensors" for comprehensive socio-ecological evaluations (Pricope et al. 2019).

The integration of LEK with remote sensing brings a distinctive methodological pursuit that targets equitable and contextually nuanced rangeland management schemes. The convergence of LEK and remote sensing grows across diverse realms of natural resource management, including fisheries, forests, and rangelands, with local expert opinion highly valued for map validation, comparison, and evaluation. This symbiotic relation unveils a landscape-scale synergy, where LEK complements remote sensing in monitoring species, conservation endeavors, and capturing ground-level data pivotal for discerning threats visible through remote sensing, such as overhunting and overfishing.

The ethos of SES accentuates the imperative of understanding and integrating LEK, fostering enhanced sustainability and resilience of social-ecological systems. Most protected areas rely on scientific ecological knowledge alone, albeit the infusion of LEK can augment the understanding of ecosystem service provision and landscape vulnerability, thus enriching the sustainability science paradigm. The CARE principles for Indigenous Data Governance, as posited by Carroll et al. (2019, 2020), emerge as a potential avenue for nurturing responsible knowledge transfer while safeguarding indigenous data rights and sovereignty. This discourse resonates with the novel approaches touched on by Rattling Leaf Sr. (2022), where the synergistic relation between remotely sensed imagery and LEK brings about a growing methodological paradigm among engaged indigenous communities.

Furthermore, the integration of ground-referenced data ultimately augments SES research beyond indigenous communities, amalgamating technological tools like cloud computing, citizen science, and "humans-as-sensors" for exhaustive socio-ecological evaluations (Pricope et al. 2019). This narrative accentuates the indispensable role of LEK in both an (non)indigenous context, and in navigating the complex socio-ecological landscapes through remote sensing applications, whilst fostering a dialogue aware of the intentional inclusive politics growing within the scientific community.

1.4 Research Objectives

In our data-centric era, the weight of remote sensing and GIS data in steering policy and decision-making is undeniable. The ripple effects from generating and applying such data influence the decisions that sculpt our local, regional, and global landscapes. Yet, each stakeholder brings a unique perspective to the table within these conversations. Those who engage with data largely through research, either in the procurement of new information, or through its analysis, hold the technical acumen and nuanced understanding of data methodologies that lay the foundation for all subsequent science applications. Meanwhile, interested parties that wish to apply spatial analysis to real-world scenarios directly shape the policies and decisions enacted. At the receiving end of these decisions stand private citizens, sovereign nations, communities of varying sizes, and more-than-human individuals, experiencing tangible repercussions on their lives and surroundings.

This convergence of varying perspectives underscores the importance of inclusivity and grounded realism in policy-making. It necessitates the creation of policies that, while being data-driven, are also sensitive to the varied experiences and narratives of those they impact. This is where the concept of GIS data as 'boundary objects' comes into play, bridging the gap between different stakeholders and fostering a shared understanding that transcends individual biases and viewpoints (Caughlin et al., 2019). Boundary objects, such as remote sensing data, can be a common language for interdisciplinary engagement to communicate in mutual understanding. However, these objects can also become sites of contestation, as different Data Users may interpret and utilize them for diverging purposes. Within this dense web of roles, GIS data's role as 'boundary objects' offers a potential convergence point. Maps and visualized GIS data can bridge disparate viewpoints, serving as a shared visual language. Emphasizing inclusive engagement in applied science becomes an ethical mandate. By prioritizing public engagement, we can shape co-production processes that benefit all involved. This chapter intends to critically explore the inclusive dynamics of spatial data's lifecycle, spotlighting stakeholder engagement techniques, co-production processes, and the art of cross-disciplinary communication.

By viewing remote sensing and GIS data through this multi-faceted lens, we can begin to envision a model of policy-making that is not only driven by hard data but is also deeply rooted in the complex realities of (non)human experiences. A model that acknowledges the power of maps as boundary objects and leverages this power to create more inclusive and grounded policies. As I continue this exploration, I dive deeper into these concepts, shedding light on how a more holistic approach to data utilization can guide the development of policies that truly reflect the complexities and nuances of our world. I advocate for a policymaking model anchored in empirical data yet woven with the intricacies of (non)human narratives. This model acknowledges maps as instrumental boundary objects, capitalizing on their potential to architect inclusive and nuanced policies.

My central questions for this chapter are:

1. How do diverse stakeholders, including Data Producers, Data Scientists, Data Users, and Impacted Persons, engage with and interpret remote sensing and GIS data, and how does this interaction influence the development of inclusive policies in geospatial sciences?
2. Which stakeholders remain underrepresented or absent within the discourse around remote sensing and GIS data, and what are the implications of this exclusion on policy-making and the resulting tangible changes felt by Impacted Persons?

Question 1 integrates aspects of curation, application, and interpretation of remote sensing data, as well as the roles and perceptions of different stakeholder groups. By focusing on engagement and interpretation, I conducted interviews, surveys, and participant observation to focus on the various engagements with GIS and remote sensing data and interrogate participants' interpretation of such data. Question 2 emphasizes the gaps in the discourse, which can be particularly relevant when evaluating the comprehensiveness and inclusivity of policies and practices in the geospatial domain.

2 Methodology

By examining the experiences and perspectives of various stakeholders across a broad spectrum in the remote sensing domain, I am able to provide a unique look into the tensions and synergies between public engagement with governmental agencies, public interest groups, or for the implementation of public policy, and private collaboration through Non-Governmental Organizations (NGO) and private corporations in the field of applied remote sensing. By using a combination of participant observation, anonymized interviews, and surveys, this research aims to weave together a narrative that reflects the diverse voices and experiences of those who interact with remote sensing data in different capacities.

2.1 Multi-faceted Analysis of Engaged Perspectives

Incorporating stakeholder perspectives is critical to addressing the main research question of this study. Stakeholder perspectives not only enrich our understanding of practical applications of GIS data, but also reveal the power dynamics, biases, and social implications inherent in the data production and usage. As a result, integrating these perspectives can lead to more democratic, inclusive, and grounded decision-making processes, bridging the gap between technical knowledge and societal realities. Ultimately, by viewing stakeholder perspectives as an integral part of the data life cycle, we can ensure that the potential of remote sensing and GIS data is realized in a manner that is both technically robust and socially relevant.

In order to gain a comprehensive understanding of the interaction between various data users and geospatial technologies, such as remote sensing and GIS, a systematic examination is essential. This examination entails identifying Data Producers, Scientists, Users, and Impacted Persons, and understanding their unique roles, responsibilities, and interests. My methodological approach is rooted in an examination of the distinct perspectives and expertise of the Data Producers, Scientists, and Users. This adopts a focus on how these technologies are developed, applied, and the subsequent effects on Impacted Persons. By scrutinizing the multifaceted interactions these data users have with boundary objects, such as remote sensing data, I can build a more intricate picture of the obstacles and possibilities within the pixel pipeline.

2.1.1 Anonymous, Semi-Structured Interviews

The decision to employ semi-structured interviews as a primary data collection tool is interwoven with my research objectives. These interviews provide a strategic medium to delve into the complexities of diverse stakeholder engagement and interpretation of remote sensing and GIS data. They offer the quantitative robustness necessary to address the who and what of underrepresentation, while simultaneously retaining the qualitative richness that permits a nuanced exploration of the why and how. They serve as a conduit to capture the subtleties of

Data User experiences and attitudes, highlighting areas of policy-making that may have been influenced or overlooked due to the exclusion of Impacted Persons. Thus, the conversational, yet structured nature of these interviews aligns seamlessly with the goal of investigating the interdisciplinary landscape of geospatial sciences and the socio-political implications of socializing the pixel.

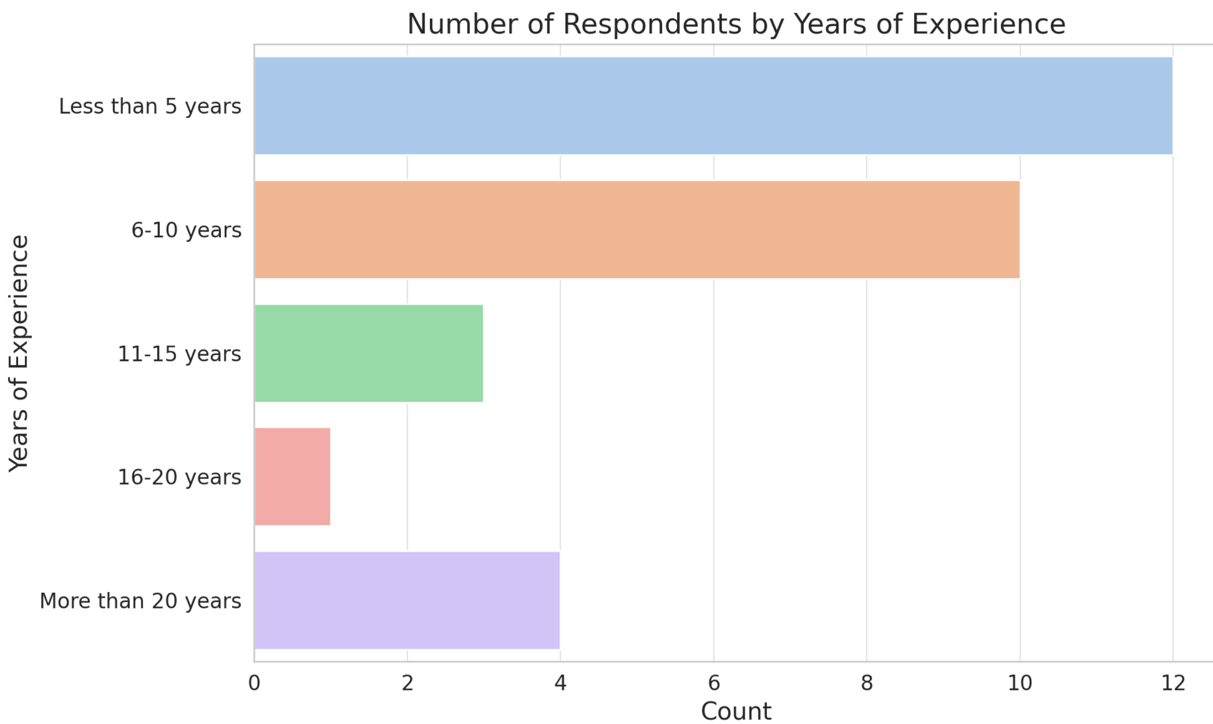
These interviews are conducted with predetermined questions (Appendix B) that allow room for dialogue and exploration of ideas among participants. The interviewer has the flexibility to ask follow-up questions or prompt participants to explore certain topics further, while the participants are encouraged to share their perspectives on a variety of topics. This approach requires thoughtful preparation before each interview session, including the creation of a detailed agenda and list of questions. It is essential that the interviewer has a clear understanding of their role and the purpose of the interview before beginning each session. Throughout, I adhered to ethical guidelines, including obtaining informed consent from participants and ensuring confidentiality and anonymity throughout the research process.

In the preparation of this chapter, I interviewed two professionals specializing in remote sensing. Both experts possess advanced degrees in forestry applications of geospatial technology and currently contribute their expertise to the private sector. They provide consultancy services for global carbon offset projects, applying their specialized knowledge and skills to address pressing environmental challenges. Their insights form a significant part of the discussion in this chapter, enhancing its academic rigor and real-world relevance.

2.1.2 Surveys

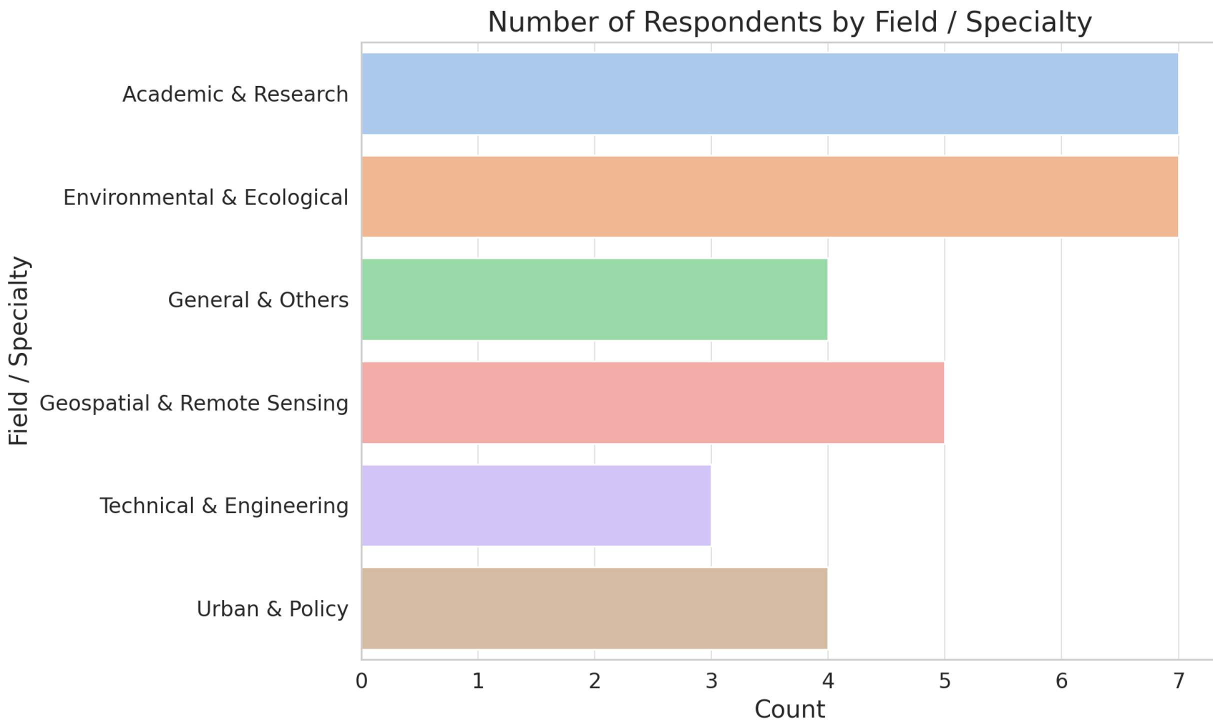
In a bid to broaden the scope of my research and gain a more comprehensive understanding beyond my direct interactions with Data Producers, Scientists, and Users, I collected additional data through a targeted survey. The survey consisted of 10 questions aimed at professionals across academia, public institutions, and private industries. The survey's objective was to assess the respondents' familiarity with geospatial technology, including their prior experiences with its application. Additionally, it sought to gauge their perception of the role and significance of remote sensing in scientific applications. The survey was completed by 30 individuals, representing a diverse mix of professionals from all three sectors: academic, public, and private. The responses gathered offer valuable insights into the practical usage and perceived importance of geospatial technology and remote sensing across different fields and sectors. The distribution of respondents based on years of experience (Fig. 1) reveals the spectrum of expertise and knowledge encompassed within the sample. The range of experience, from novices to seasoned professionals, augments the depth and breadth of perspectives garnered through the survey.

Figure 1. Distribution of survey respondents' years of experience.



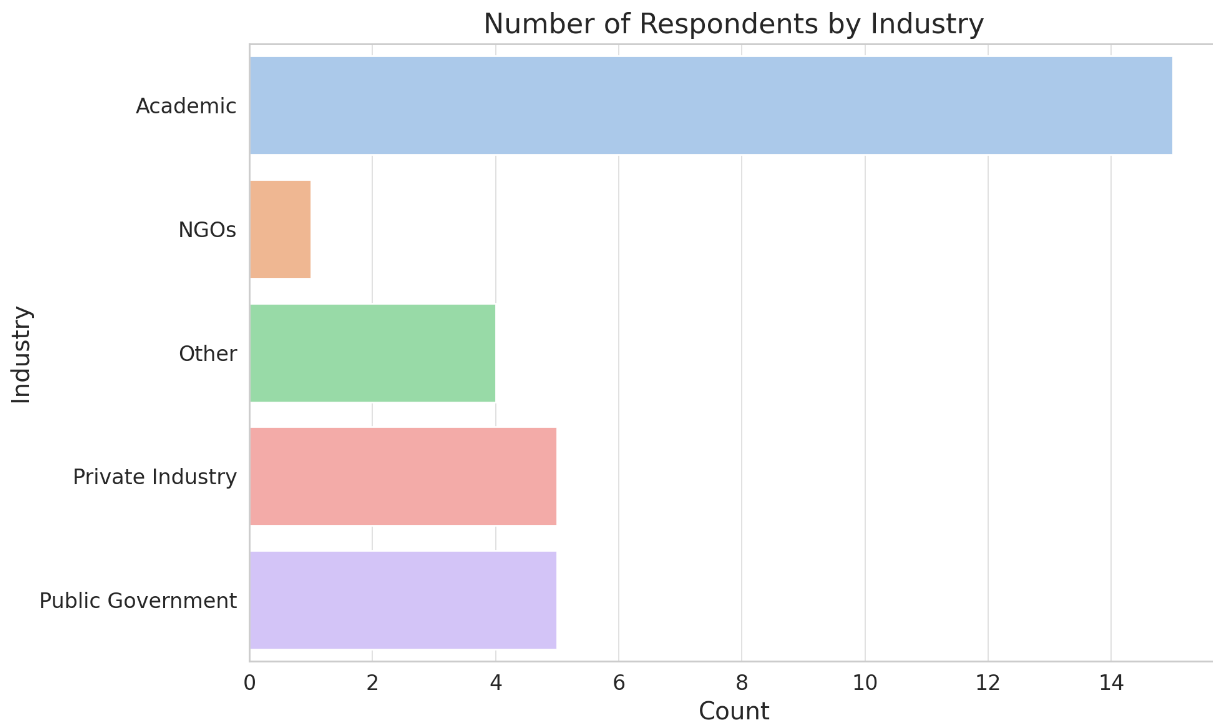
Additionally, a categorization of respondents by field (Fig. 2) unveils the varied domains where geospatial technologies find application. This diversity epitomizes the interdisciplinary nature of geospatial sciences and its resonance across distinct fields of inquiry.

Figure 2. Distribution of survey respondents' roles grouped by field of work.



The industry-wise distribution of respondents (Fig. 3) accentuates the number of sectors employing geospatial technologies. The data underscores the variety of geospatial applications across different industrial sectors, reflecting its role in contemporary professional landscapes.

Figure 3. Distribution of survey respondents' roles grouped by industry.



2.1.3 Participant Observation

Participant observation is a qualitative research method highlighted by the researcher's dual role as both observer and active participant in the group or community under study. This approach involves immersion in the participants' everyday life, enabling the researcher to gain a comprehensive understanding of their social context, behaviors, interactions, and beliefs. Such immersion, unlike more detached methods, yields rich, nuanced insights as the researcher experiences and interprets group dynamics firsthand.

In my research, especially in science-led projects, participant observation is instrumental in exploring the ethics of stakeholder engagement. Actively engaging in these processes allows me to grasp the dynamics within different stakeholder groups, transforming abstract concepts into tangible experiences. For example, I can directly observe the nuances of conversations, the diversity of perspectives, and how maps evolve into 'boundary objects' that facilitate communication across groups. Moreover, this method allows witnessing the practical application of GIS data in policymaking. It offers real-time insights into potential biases and misinterpretations and areas needing improvement. Direct interactions with individuals affected by policies provide a deeper understanding of the actual implications of these decisions, highlighting the importance of ground realities in shaping data-driven policies.

Conducting participant observation involves initial steps like gaining access to the community and establishing rapport, which requires building trust and understanding community norms. As a researcher, I maintain a balance between participation and observation, ensuring an objective collection of data. Detailed field notes documenting events and interactions, coupled with personal reflections, are critical for analysis. This methodology also includes informal interviews and gathering additional insights, enriching the understanding of the social context. Reflective practice is vital, where I continually assess my role and influence within the group. Data analysis then involves identifying themes and patterns, an iterative process where participant feedback is sought to validate findings.

Ethical considerations extend to reporting, ensuring confidentiality and anonymity of participants. Post-research, maintaining relationships with the community and sharing findings back are part of the ethical commitment. This comprehensive approach in participant observation not only enriches the research with in-depth, firsthand experiences but also significantly enhances our understanding of the complex processes involved in the journey from remote sensing and GIS data curation to policy implementation.

My participant observation is grounded in three prior case studies that demonstrate the practical application of remote sensing and GIS technologies for societal benefit. The first case study covers my involvement developing urban forestry tools for the Los Angeles County Department of Public Health, and is the main subject of my first two chapters. I focused on the development of an automated remote sensing model to alleviate the financial and logistical challenges associated with manually inventorying the county's public street trees, underscoring the potential of remote sensing technology in enhancing urban planning and resource management processes. Additionally, as part of my urban forestry research in collaboration with the County of Los Angeles, I surveyed users of the model discussed in the preceding chapters. This was an integral part of our process, aiming to understand and incorporate the perspectives of those who would directly interact with and benefit from the model. One of our project managers, an employee of NASA, devised a structured survey framework inspired by the agency's change traceability matrix. This framework allowed us to systematically record the needs and aspirations of stakeholders before embarking on the model development phase. Such a methodological approach ensures that the resulting model is tailored to address specific user requirements, thus enhancing its practical value and usability. The insights gained from these preliminary project meetings are also incorporated in this analysis. These findings provide a crucial understanding of the stakeholder landscape, shedding light on the diverse needs and expectations that guide the use and application of our urban forestry model. These data collection and analysis strategies illustrate the importance of stakeholder engagement and user-centric design in developing effective geospatial models.

The second case study was a joint venture between the University of California, Los Angeles (UCLA), and Arizona State University (ASU), that critically evaluated the efficacy of remote sensing-derived land surface temperature measurements as a tool for informing hyperlocal extreme heat interventions. This research provided a counterpoint to an existing collaboration between the City of Los Angeles and NASA's Jet Propulsion Laboratory (JPL), thereby emphasizing the need for empirical validation in such technologically advanced, data-driven initiatives.

The third case study is my ongoing involvement as a steering committee member for the Landscape Exchange Network for Socio-environmental Systems research (LENS), which seeks to advance a basic scientific understanding of integrated SES and the complex interactions--via dynamics, processes, and feedback--within and among the biophysical and social components of these coupled systems. LENS is a five-year research coordination network funded by the National Science Foundation and invites researchers interested in SES to work within the National Ecological Observation Network (NEON) Airborne Observation Platform (AOP) landscapes. Within this framework, members are encouraged to co-produce research with stakeholders working in and inhabiting these landscapes to push conceptual boundaries and theoretical constructs for SES by leveraging NEON AOP data. Members are encouraged to develop proposals that push conceptual boundaries and build new frameworks for understanding SES while deploying NEON AOP in their analysis. The intention is that network members will explore a connected and integrated SES within the NEON AOP landscapes or domains, including a detailed analysis of the processes and dynamics between the environmental and human components of the chosen landscape.

Three main objectives guide the steering committee while developing educational programs around NEON AOP data and analysis, practical steps for choosing and implementing an SES research framework, and providing an engaging and safe environment to share expertise and common language around the co-production of research questions between (non)experts. These objectives include 1) characterizing SES, 2) establishing remote sensing as an SES boundary object, and 3) utilizing translational ecology within the NEON context, fostering an interdisciplinary dialogue between stakeholders. Translational Ecology (TE) represents a shift in ecological research, emphasizing the integration of ecological knowledge into decision-making processes by melding scientific insights with social dimensions (Enquist et al. 2017). This approach recognizes the need for improved communication among different knowledge producers, including indigenous peoples, local communities, and academic researchers.

These case studies collectively exhibit a commitment to participatory research, demonstrating the importance of integrating cutting-edge technology, stakeholder engagement, and considerations of equity in addressing societal challenges. They also highlight the ongoing need for research and discourse to ensure the effective and inclusive use of these technological tools, as well as provide an opportunity to consider how these tools can be employed to 'socialize the pixel'. Much like the principles underpinning the mentioned case studies, 'socializing the pixel' advocates for stakeholder engagement, equity, and inclusivity in the use and distribution of digital resources. It is a call to action for researchers and policymakers alike to continue their efforts in making technology a tool for societal enhancement, rather than a source of disparity.

3 Results

My work reveals that Data Producers, Data Scientists, and Data Users generally view and interact with remote sensing data as an essential tool in their respective fields. These stakeholders recognize the potential benefits of using remote sensing, such as providing timely and accurate information for decision-making, environmental monitoring, or policy evaluation. However, they also acknowledge the challenges and limitations associated with remote sensing data, such as issues of data quality, accessibility, and interpretation. This section presents the results from the research by method, beginning with the semi-structured interviews, followed by the survey, and finalized with my participant observation across three case studies.

3.1 Semi-structured Interviews

My interviews with two remote sensing experts, both with advanced degrees and now working in forestry applications within the carbon credit market, revealed a highly structured interaction between Data Producers, Scientists, and Users within the private sector. The intricate protocols and meticulous auditing processes underscore the diligence applied in dealing with remote sensing data. My interviewee shared, "Mapping is one of the initial steps for any of our projects. We do a lot of remote sensing for mapping the area." This clearly demonstrates the rigorous standards that govern geospatial data production and interpretation. The interviewee also highlighted the use of advanced technologies like machine learning, specifically the Random Forest algorithm, for land classification, thereby reflecting the exactness involved in curating this data, "For our reforestation projects and Reducing Emissions from Deforestation and Degradation (REDD) projects, we used land classification using a machine learning algorithm called Random Forest."

Yet, within these codified interactions and stringent protocols, there's a glaring omission: the absence of Impacted Persons. The interviewee states, "Quite often, we don't go to the field to measure it ourselves because we try to hire local actors for that." This tacitly implies that while the local community is involved in the data validation process, they are not part of the larger discourse, a key determinant of their own environment. This trend is symptomatic of the broader issue of the systematic exclusion of Impacted Persons from the process of 'socializing the pixel.'

In reframing our perspective of 'socializing the pixel,' it becomes evident that the power dynamics involved in defining, interpreting, and applying pixel-based data are heavily skewed. While the existing protocols exhibit a high degree of rigor and precision, they fall short in engaging those most affected by the spatial realities these pixels represent. Thus, the process of 'socializing the pixel' calls for more inclusivity and a shift in power dynamics, ensuring that those impacted by the interpretation of pixels are not left on the sidelines but become active participants in the process.

3.2 Survey Work

In addition to the case studies from my work, this analysis included a dedicated survey with a total of 30 responses, providing a glimpse into the diverse demographic landscape of individuals engaged in geospatial technologies and related fields. The respondents represented a blend of experience, fields, and industries, from Urban Sociologists and Public Health Specialists to Environmental Scientists and Urban Planners, underscoring the multidisciplinary nature of geospatial applications. This section delineates the demographic contours of the survey respondents, setting a foundation for the ensuing discussion on the findings. Based on my analysis of the survey data, there are several key findings that reinforce the argument that Impacted Persons are relatively absent from any meaningful engagement in remote sensing applications.

- Interactions with Data Users: A significant portion of the respondents (n=11), including Urban Ecologists and Public Policy Analysts, shared insights related to their engagement with Data Users, or stakeholders as they described them. They highlighted the increasing significance of stakeholder interactions, which corroborates the central premise of my thesis that emphasizes community participation and stakeholder involvement. These responses shed light on the necessity of balancing scientific integrity and stakeholder interests. However, the lack of engagement with Impacted Persons directly could indicate a gap in the integration of qualitative, subjective narratives into remote sensing data analysis.

- Scale Mismatches: Concerns over scale mismatches were raised by a smaller group (n=3), comprising roles like Environmental Scientists and Geospatial Analysts, who stressed the importance of localized, culturally-sensitive approaches. This aligns with the study's focus on the relevance of hyperlocal data and its integration.

I found that only five respondents, including those in climate science, urban planning, and public health, reported working directly with Impacted Persons in various areas such as tourism, recreation, climate science, public health, and urban planning. They emphasized the value of communication and educational initiatives, suggesting potential avenues for increased community engagement.

The survey also reveals distinct perspectives based on professional roles in geospatial analysis. The curation and utilization of remote sensing and GIS data exert significant influence on the perspectives of various stakeholder groups. Additional survey results indicated that:

- Data Scientists (n = 18), the predominant group among our respondents, emphasized the importance of scientific rigor, often grappling with external pressures from other stakeholders. Most notably, they highlighted the challenge of explaining complex geospatial concepts and aligning project goals with various stakeholder demands.
- Data Producers and Data Users (n = 12), including roles like Remote Sensing Analysts and Geospatial Projects Directors, underscored the importance of contemporary geospatial technologies, with GIS being the most utilized. This

utilization underscores the centrality of GIS in shaping stakeholder interactions and perceptions.

- The absence of respondents categorized as Impacted Persons, with only a minority (n=5) regularly interacting with community members, indicates a crucial gap. These interactions, reported by professionals like Public Health Officials and Urban Planners, underscore the need for policies promoting clear communication and stakeholder engagement.

The responses collectively highlight a conflict between scientific rigor and end-user expectations across various roles. Government officials, for instance, often demand rapid results, leading to conflicts in objectives, as reported by Urban Planners and Environmental Analysts. Environmental concerns, such as climate change impacts, were frequently mentioned, particularly by roles like Climate Impact Forecasters and Ecologists. Balancing diverse interests, a challenge faced by nearly half of the respondents, including those in tourism and business sectors, points to the complexities of managing different stakeholder expectations. The frequent emphasis on scientific rigor, especially among Data Scientists, suggests that policies must prioritize the integrity and accuracy of geospatial data. The challenges faced in collaborating with non-experts, such as managing expectations, indicate a need for policies that foster clear communication and public engagement. The reliance on geospatial technologies, particularly GIS, highlights the need for policies that ensure the accessibility and usability of these tools across various stakeholder groups.

Despite the respondents' differences in focus, as shown above, the respondents agreed that there is often a conflict between scientific rigor and end-user expectations. The list below highlights this recurring theme in responses:

- Government Expectations: There appears to be a recurring theme of government officials wanting swift results or having different expectations, such as in the first response where urban planning was in conflict with the need for sociological studies (n = 7).
- Environmental Concerns: 80% of respondents highlighted conflicts arising from environmental concerns, such as climate change impacts or the fragility of alpine ecosystems.
- Balancing Interests: 47% of responses indicate that professionals often find themselves balancing diverse interests, such as tourist interests, business interests, and environmental conservation.

Challenges of collaborating with non-experts in geospatial projects:

- Balancing Different Objectives: As seen in the responses, two respondents struggle to balance urban development with sociological insights or business interests with conservation.
- Managing Expectations: Several respondents mention the challenge of managing the expectations of non-experts.
- Conservation vs. Development: The challenge of promoting conservation while addressing development or business interests recurs in the responses.

The qualitative responses shed light on the intricate challenges faced by geospatial professionals. While they emphasize scientific rigor, they often grapple with external pressures from various Data Users. These pressures range from the demand for swift results, balancing diverse interests, addressing misinformation, to managing expectations, and "[o]vercoming climate change denial or misinformation.," as one climate scientist mentioned. The responses underscore the importance of effective communication across groups.

3.3 Participatory Work

3.1.1 Participatory Work: Urban Forestry in Los Angeles County

While building an automated remote sensing model for the optimization of Los Angeles County's urban forest, project objectives underscored the intersection of environmental justice and public health. It's crucial to note that while special interest groups often express a fondness for trees, their capabilities are sometimes overstated. A review by MIT unveiled a report by The National Academies of Sciences, Engineering, and Medicine which suggested converting up to four million hectares of land—roughly the size of the U.S. state of Maryland—into permanent forests to sequester 150 million metric tons of carbon annually (Temple, 2020). However, considering the US emits approximately 5.8 billion tons of carbon across various sectors, the required land would be almost 155 million hectares, more than double the size of Texas (Temple, 2020). The availability of such vast lands is minimal in most countries, including the U.S. Repurposing lands also has far-reaching implications for agriculture, logging, and other industries, but this does make urban forests an interesting alternative (Temple, 2020).

We conducted a survey among County Data Users spanning eight departments to identify their operational needs and refine our methodologies for providing tech solutions. A project manager on our team, who was formerly employed with NASA JPL, incorporated the Change Traceability Matrix approach from the agency into our project planning (Stavros, 2021). This matrix was adjusted to inquire about the stakeholders' project expectations, data requirements, and the most effective method of delivering our results. Moreover, we also asked how this project could produce societal benefits. The dominant themes included:

- Equity and Impact: Stakeholders underscored the significance of "mindful impacts related to equitable access," emphasizing the collective vision of ensuring urban forests benefits are accessible to all community members (n = 4).
- Historical Context: Stakeholders expressed concerns about addressing historical injustices, emphasizing that the project should contribute to a more inclusive urban environment (n = 3).

The social and cultural contexts shaped by historical land use patterns play a vital role in influencing public perceptions and values regarding urban green spaces and are acknowledged in the survey responses. This aspect is crucial for ensuring fair and equitable access to and distribution of urban forests, especially in areas that have historically been neglected or underserved. Understanding these multifaceted impacts of historical land use and cover across all three pilot sites (Appendix A) is essential for effective and inclusive urban forest planning.

Demographic Insights on Survey Respondents

The survey pulled in views from 15 individuals across a broad mix of groups involved in urban forestry within Los Angeles County. At the top of the list was The Nature Conservancy, a global group known for its work in protecting nature. Not far behind was TreePeople, an active group in Los Angeles, working hands-on to make the city greener. These two, along with others in the survey, ranged from community groups to government offices, highlighting the wide range of voices in the discussion on urban forestry. County departments like the Department of Parks and Recreation and Beaches and Harbors brought insights from the viewpoint of rules, guidelines, and city planning. The California Department of Forestry and Fire Protection (also known as CALFIRE) also participated in the survey, showing the project's special connection between city trees and fire.

The respondents, representing a cross-section of stakeholders in Los Angeles County's urban forestry landscape, provided diverse insights, shedding light on their unique perspectives and preferences. There were three specific insights: the need for community engagement, common tools, and diverse means of data collection. In the context of grant funding allocations, the emphasis was unmistakably on community engagement. Respondents repeatedly underscored the need for "meaningful community engagement" and active "community involvement". These responses, more than just a call for participation, highlighted the desire for deep-rooted and impactful involvement of the community in decision-making processes. When asked about the tools they employed in their work, the answers varied widely. The lack of commonality in the responses suggested that stakeholders in urban forestry use an array of tools tailored to their specific needs and tasks. This diversity of tools possibly mirrors the multifaceted challenges and objectives within urban forestry, from tree health monitoring to community engagement and policy planning.

Similarly, when it came to data format preferences, the respondents' answers did not converge on a single or few formats. The breadth of answers here again reflects the diverse nature of the stakeholders involved. Different roles and responsibilities within urban forestry might necessitate various data formats, from spatial GIS data for planners to tabulated data for researchers or community engagement professionals.

This survey, with its broad range of respondents, paints a picture of an applied remote sensing landscape that is both diverse in its stakeholder makeup and complex in its needs and preferences. Whether it's the tools they use, the data formats they prefer, or the emphasis on community engagement, the respondents' answers underline the multifaceted nature of urban forestry in Los Angeles County. Notably, three years after data collection, when the project findings were presented, the stakeholder feedback on how the analysis could be used shifted towards monitoring trees on private property for code enforcement purposes. This shift raises important questions:

- Alignment with Initial Expectations: The new objective seems significantly different from the initial Data Users' expectations of equitable access and addressing historical inequities. Monitoring trees for code enforcement on private properties introduces potential concerns related to privacy, property rights, and potential inequities in enforcement.

- Public Engagement and Trust: The shift in objectives might affect public trust due to the project's initial framing and the emphasis on equity and historical context (Appendix A). Data Users who expected a project focused on broader societal challenges might perceive this shift as a deviation from the project's original intent.

This evolution in project objectives highlights the complexities of public engagement. While projects must remain adaptive and responsive, significant shifts in objectives, particularly those diverging from initial stakeholder expectations, can have repercussions on community trust and project outcomes.

The urban forests project in Los Angeles County demonstrates the challenges and opportunities of public engagement in applied research. The significant shift in end-user objectives from the project's original framing serves as a reminder of the dynamic nature of stakeholder engagement and the need for research projects to remain both adaptive and aligned with community needs and expectations. As research projects evolve, maintaining transparency, continuous engagement, and alignment with initial commitments becomes crucial to preserve trust and achieve meaningful outcomes. The stakeholders' previous emphasis on equitable access and historical context suggests that they expected the project to provide data on urban forests and address broader social and environmental justice issues, but those became juxtaposed against wanting to use the results as a means to levy fines and citations.

The shift directly impacts the study. For instance, if the study is intended to be used for monitoring, an accuracy of 80% for a single species introduces some error and could send County personnel to a location under false pretenses, as would be the case for my model developed in Chapter 3. This is arguably a misallocation of County resources, however, when the model gets it right more often than it gets it wrong, the aggregate amount of time and money saved deploying resources outweighs the potential waste. This application provides a net positive in efficiency and effectiveness for managing the urban forest on public right-of-ways. However, if the model is adopted for code enforcement of private property to ensure property owners or renters are properly caring for a protected species, any error in classification or assessment could lead to falsely administered fines. The difference in impact demonstrates the need for continuous stakeholder engagement to reduce negative unintended consequences for scientific studies.

The survey responses, when viewed through the lens of public engagement, underscore the critical role of the voices of Data Users in shaping and guiding research projects. Users, from public works workers to data scientists and managers, bring diverse perspectives that enrich the project and ensure its relevance to community needs. Their concerns about equity and historical context highlight the importance of not just collecting data but also interpreting and using it in ways that address broader societal challenges. The survey responses provide invaluable insights into stakeholder expectations and concerns related to the urban forests project in Los Angeles County. Situating these findings within the broader context of public engagement emphasizes the need to ensure that research projects are not only scientifically rigorous but also socially relevant and responsive to community needs. By incorporating diverse stakeholder perspectives, projects like the urban forests study can contribute to more inclusive and equitable urban planning and policy-making. This would include broaching the topic of remote sensing as a surveillance tool for code enforcement and ethical considerations when private citizens are kept in the dark about such practices.

3.1.2 Participatory Work: Urban Form and Thermal Comfort in the Southwest U.S.

One case study that shows the need for the inclusion of Impacted Persons is my previous work studying the impact of urban form on thermal comfort across three Transformative Climate Communities in California. These communities were identified as vulnerable populations based on a number of socioeconomic, demographic, and environmental data, including exposure to extreme heat events. As a city located in a hot, dry, arid environment, the issue of heat in Los Angeles is not merely about temperature. It is also about thermal comfort and the urban design interventions that mitigate extreme heat. In fact, as our team found, shade is the most important factor when it comes to human comfort–more than air temperature, more than humidity, more than wind speed (Turner et al. 2021; Mars & Berube, 2023). This close relationship between human thermal comfort and mortality and illness due to heatstroke cannot be overstated. However, current interventions by NASA and the City of Los Angeles, through the use of remote sensing measurements of land surface temperatures (LST) (Hulley et al. 2020), are inadequate. Our findings suggest that shade is a better predictor of both simulated LST and mean radiant temperatures (MRT) than remotely sensed LST. Thus, these measurements are not sufficient for guiding heat mitigation at hyper-local scales in cities.

Our study underscored the dire need for incorporating the perspectives of Impacted Persons in the development and implementation of geospatial technologies and science applications. The cool pavement initiative, which was based on LST, was designed to mitigate urban heat by reflecting incident radiation back into the atmosphere. However, it failed to account for the lived experiences of those on the ground. By increasing the albedo of the pavement, the intervention inadvertently intensified radiation absorption for pedestrians, children, and commuters—Impacted Persons—who had to bear the brunt of an additional 10% radiation during the hottest parts of the day. This unintended consequence underscores the importance of including direct involvement and designing research questions around perspectives of Impacted Persons. Their inclusion would have provided crucial insights into the practical implications of the initiative and likely would have led to a more effective and human-centered approach to urban heat mitigation. Thus, this case study emphasizes the profound importance of diverse stakeholder engagement in shaping the application of geospatial technologies for societal benefit.

During the Surface Biology and Geology community workshop organized by NASA in Washington D.C., October 2022, a question was put forth to the urban heat group. The group, having just announced a tripling in financial backing from the City of Los Angeles for their cool pavement project, overlooked the query entirely. They responded, "The glint issue was effectively resolved by transitioning from white paint to another NASA-developed paint technology for tanks, which is gray and does not reflect sunlight into drivers' eyes." However, the question never raised the issue of glint, instead, it demanded attention to the problem of incident radiation being reflected back onto vulnerable populations - a concern that went unaddressed.

Arguably, the cool pavement project may have served as an exceptional promotional tool for NASA's science application team, demonstrating its efficacy in mitigating the urban heat island effect. However, this triumph may have inadvertently marginalized the very individuals it was designed to assist. The apparent disregard for these concerns suggests that both NASA and the City of Los Angeles are satisfied with the preliminary results of the cool pavement pilot project. Sam Bloch, a CityLab journalist, conversed with a resident of a street treated with cool pavement. This interaction revealed that, from their perspective, the initiative did not deliver a tangible cooling effect (Bloch, 2019).

3.1.3 Participatory Work: Landscape Exchange Network for Socio-environmental Systems

The use of remote sensing technologies plays a crucial role in policy-making and decision-making scenarios, particularly within the context of the LENS Research Coordination Network, where I serve on the steering committee. As a committee member, we facilitate SES research using NEON AOP data, which are pivotal for SES research. This data serves as a consensus-building tool for stakeholders and researchers, capturing landscapes through its high-quality imaging capabilities. AOP's sensor suite offers intricate measures of vegetation dynamics, biodiversity, and ecosystem functionalities. These metrics grant insights into how ecosystem characteristics reciprocate with human and natural system interactions, a core SES concern (Ordway et al., 2020). The challenge lies in ensuring data quality and consistency across platforms and years, making it accessible to everyone involved in the pixel pipeline. Achieving this paves the way for remote sensing to serve as a practical boundary object.

At LENS, Impacted Persons living and working within NEON AOP flight boundaries have been identified for a number of sites, but engagement remains low. Much of the discourse surrounding science applications through LENS involves mention of these persons, but lacks discrete protocols to engage with them. Additionally, integrating remote sensing within a translational ecology framework holds its own challenges:

- Remote sensing data can be challenging to interpret and understand.
- There is a lack of agreement on using remote sensing data to inform conservation decision-making.
- There is a need for improved communication and collaboration between remote sensing experts and other stakeholders.

This intensive engagement, while ensuring a rich and holistic understanding of ecological issues, can strain the bandwidth of researchers (James et al., 2022). The continuous back-and-forth, the need to address diverse stakeholder concerns, and the merging of scientific rigor with local insights can be taxing. This places a significant onus on researchers, potentially leading to burnout or diluted engagement quality over time (Safford et al., 2017; Goodrich et al., 2020). Leveraging boundary organizations offers a strategic solution to this limitation. These entities, designed to bridge the gaps between academia, communities, and policymakers, can act as intermediaries, easing the burden on individual researchers. With their expertise in engaging with (non)experts, boundary organizations can facilitate effective communication, streamline interactions, and ensure that researchers' insights and community feedback are harmoniously integrated. Moreover, they can provide logistical, administrative, and even emotional support, ensuring that researchers can focus on their core strengths without being overwhelmed by the multifaceted challenges of community engagement. In essence, while translational ecology's intensive engagement model can test individual researcher bandwidth, boundary organizations present a viable solution. By acting as intermediaries and providing necessary support, they can ensure sustained, effective, and enriching engagement with Data Producers, Scientists, Users, or Impacted Persons.

4 Implications for Future Geospatial Research and Applications

My research is a call to action for the remote sensing community to rectify a persistent information imbalance. It is clear from the nearly universal lack of engagement with Impacted Persons, with the exception of five survey respondents, we need advocates for the inclusion of Impacted Persons in every step of the data pipeline, to create a comprehensive, pixel-level narrative that truly reflects the human-environment relationship. By doing so, we can ensure that decisions are based on a more complete and nuanced understanding, thereby enhancing the accuracy and relevance of our investigations. Incorporating the perspectives of Impacted Persons into geospatial metadata is more than a nod to inclusivity; it's a transformative approach that elevates the quality and depth of geospatial data. By weaving in the lived experiences and insights of these individuals, we add a layer of qualitative richness to a domain that is predominantly quantitative. This interplay between qualitative and quantitative data adds a nuanced dimension to geospatial sciences, making the findings more holistic and contextually grounded. Including Impacted Persons' perspectives brings with it a fresh wave of interpretative depth, especially beneficial for socio-environmental research. By understanding the human implications and interpretations of spatial phenomena, researchers can draw more robust and comprehensive conclusions, bridging the gap between numbers on a screen and realities on the ground.

Furthermore, this integration is a significant stride towards data sovereignty, granting communities, often sidelined in geospatial dialogues, a voice and stake in the data that concerns their lands and lives. This not only reshapes the narrative around data ownership and authorship but also democratizes the geospatial discourse, extending its reach beyond the confines of labs and into the community gatherings. Such an inclusive approach is particularly vital for Data Users working in areas like public health, environmental justice, and equity, or like those of us at LENS. By having access to data that is both scientifically rigorous and socially informed, they can craft solutions and interventions that resonate with the community's needs and aspirations. Moreover, this approach can re-envision public engagement in geospatial sciences, fostering a collaborative space where community insights and scientific expertise coalesce, driving forward both the science and its societal impact. Some key themes and points include:

- Levels of Engagement: From "Inform" and "Communicate" to "Engage", highlighting the gradient of involvement and decision-making shared with communities, as we saw from five survey respondents working across fields that regularly interact with Impacted Persons, such as tourism, recreation, public health, urban planning, and education.

- Ethics of Community Engagement: With only one survey respondent discussing the importance of understanding cultural differences while working in the field, there is a clear and demonstrated need to emphasize the ethical considerations when engaging with communities among Data Producers, Scientists, and Users.
- Community-Based Scholarship and Public Engagement: Of the five survey respondents working closely with Impacted Persons, education programs like workshops or general community engagement activities and spaces of interaction were integral in communicating their expertise to a non-expert audience. Scholarship opportunities, or additional funding to create more of these spaces is needed.

Public engagement ranges from simply informing Impacted Persons to fully empowering them in decision-making processes. Within this spectrum, the curation and utilization of remote sensing and GIS data mold the perspectives and responsibilities of different stakeholder groups:

- Data Scientists: While many academic and professional spaces predominantly operate within the "Inform" and "Communicate" stages of public engagement, Data Scientists often find themselves navigating the complexities between maintaining scientific rigor and addressing diverse stakeholder demands. As one survey respondent highlighted, "Officials wanted to dismiss climate change impacts. I presented clear data and case studies to emphasize its importance," signifying the tension between expert knowledge and external pressures.

- Data Producers and Data Users: These groups, deeply engaged in data curation, play a critical role in informing public understanding. Their reliance on tools like GIS, as reflected in common acknowledgments within the survey that "GIS is the most frequently used geospatial technology," underscores their pivotal position in the public engagement process.
- Impacted Persons: While our data predominantly captures the perspectives of professionals, the indirect implications for Impacted Persons resonate through challenges like "balancing business interests with coastal conservation," according to one survey respondent who described themselves as a Coastal Policy Expert. In the broader context of public engagement, their voices and experiences are paramount in shifting from mere consultation to meaningful collaboration and empowerment.

The detailed narrative from my interviews offer an intimate lens into the nuanced ways professionals engage with geospatial data. Personal narratives, rich in detail and grounded in experience, serve as vital bridges in public engagement, translating high-level data into relatable insights and actionable knowledge. The global scope of the interviewee's work, spanning "Vancouver to Indonesia and Central America," emphasizes the universality of certain challenges and the need for a globally inclusive approach to public engagement.

Science-led projects like LENS serve as a site of encounter and action. They are full of unequal power relations stemming from the supposed status and prestige of the science being done, who is doing it, and who it is being done for (Sundberg 2004). How are these identities brought into being and enacted in time and place? Ethical community engagement, especially in academia and research, demands an alignment between data-driven insights and community needs. Policies must prioritize data integrity, as reflected in the sentiment shared by all survey respondents rating the importance in following scientific methods 4 or 5 out of 5, ensuring that community decisions based on this data are both reliable and representative. The challenges of "explaining complex geospatial concepts" point to the broader issue of effective communication in public engagement, emphasizing the need for education and awareness initiatives. And as the interviewee's reliance on proprietary software like ArcGIS suggests, public engagement frameworks should ensure widespread accessibility to essential geospatial tools, bridging the gap between expert knowledge and community understanding.

The interplay among the actors in all four groups shape the sociopolitical landscape of remote sensing applications, with each actor embodying a distinct role. Data Producers--originators of raw data--harness remote sensing technologies to capture geospatial information. The SES framework, intertwined with LEK, augments data production, fostering a richer understanding of environmental dynamics. The nexus of remote sensing and data science enables the transformation of raw data into actionable insights. Data Scientists, armed with computational prowess, delve into the complexities of geospatial data, extracting patterns, and trends that inform decision-making processes. The "Socializing the Pixel" framework exemplifies the imperative of a nuanced understanding and representation of social dynamics within this digital abstraction.

Spanning from policy-makers to conservationists, Data Users leverage the insights gleaned by Data Scientists to inform decisions, policies, and strategies. However, the challenges of data inequities and indigenous data sovereignty underscore the criticality of equitable data access and representation in this stage. Often relegated to the periphery of the data pipeline, Impacted Persons bear the ramifications of decisions made upstream.

"Socializing the Pixel" emerges as a potent narrative in redressing this disconnect. By fostering a more inclusive dialogue among the actors, particularly by integrating Impacted Persons in the co-production process, a more equitable, just, and sustainable data pipeline can be envisioned. This engagement not only enriches the data ecology but also engenders a more democratically constructed, socially aware, and contextually nuanced remote sensing praxis. Moreover, frameworks like the CARE principles for Indigenous Data Governance, emphasize responsible knowledge transfer and preservation of indigenous data rights, providing a blueprint for a more inclusive and equitable remote sensing data pipeline beyond indigenous collaboration.

Synthesizing these concepts calls for a paradigm shift in applied remote sensing—a transition from a technocentric to a more socio-ecologically conscious, inclusive, and equitable practice. Through such a lens, the data pipeline transcends mere technical processes, changing into a conduit that honors the multiplicity of knowledge, values, and impacts inherent in the realm of applied remote sensing. My work brings a fresh perspective to the field by integrating multiple narratives and disciplinary insights, effectively bridging the gaps between applied science, policy-making, and stakeholder engagement. It ventures beyond the conventional scientific discourse, incorporating personal narratives to highlight the complexities and nuances of interactions with boundary objects like remote sensing and GIS data. Furthermore, it addresses the often overlooked element of equitable stakeholder participation in the data life cycle, thus paving the way for more inclusive, grounded, and effective policy development. The nuanced critique of standard practices and the proposed strategies for improvement reflect a comprehensive understanding of the multifaceted nature of this field, marking a significant contribution to the discourse on integrating remote sensing and GIS technology into policy development. To move towards a more grounded and inclusive application of remote sensing and GIS information, several key recommendations emerge from my work:

- Promote Cross-Disciplinary Collaboration: Facilitating interaction between applied scientists, policymakers, and stakeholders from diverse backgrounds can enable a deeper understanding of how data is interpreted and used across different contexts. This can also help identify any biases or blind spots that may exist in current practices.
- Incorporate Experiential Learning: Hands-on workshops and training sessions can help stakeholders understand the technology and its implications, fostering a sense of ownership and participation in the data life cycle.
- Foster Transparent Communication: Clear and open dialogue about data collection, processing, and usage can build trust among stakeholders and ensure that any concerns or ideas they have are addressed.
- Ensure Equitable Access to Data: Policies should be put in place to ensure the data is accessible to all interested parties, reducing the potential for misuse or misinterpretation.
- Develop Context-Specific Applications: Understanding that the utility of remote sensing and GIS data varies across different contexts, applications should be developed with a keen focus on the specific needs and realities of each context.

These recommendations, while seemingly straightforward, encompass a myriad of experiences and nuances. They aim to capture the complexity of integrating remote sensing and GIS data into a holistic, inclusive, and grounded policy-making framework, taking into account the multitude of stakeholder perspectives, the intricacies of the technology, and the diverse societal contexts in which it is applied. By implementing these recommendations, we can strive towards a more just and inclusive use of remote sensing technology. As we continue to navigate the complexities and challenges inherent in human-environment interactions, it is crucial to remain critical, reflexive, and open-minded in our pursuit of knowledge and action. Ultimately, it is only through a collaborative, inclusive approach that we can harness the full potential of remote sensing to address the pressing social and environmental issues facing our world today.

4.1 Ongoing and Future Work

The initial findings presented in this chapter illustrate the early stages of an ongoing investigation, the scope of which is anticipated to widen in the future. The subsequent stages of my research will aim to delve deeper into the interactions with geospatial data by broadening the spectrum of my survey questions. This approach will allow me to examine the dynamics beyond applications and explore the intricate relationships among individuals that contribute to defining the pixel. In order to achieve a more holistic understanding, my future interviews will encompass all participants integral to the remote sensing data pipeline, rather than solely focusing on Data Scientists within the private sector. Of significant importance is my intention to incorporate the perspectives of Impacted Persons from previous case studies I have collaborated on. This approach offers an invaluable lens through which I can amalgamate participant observation with the real-world experiences of individuals impacted by our research. Future work in this direction is crucial to unravel the complexities inherent in this field.

4.1.1 Limitations and Challenges

Despite the substantial benefits of implementing these co-production recommendations, there are critiques which warrant thoughtful consideration. Firstly, cross-disciplinary collaboration, while ideal, may face considerable challenges due to the differing terminologies, methodologies, and priorities of distinct disciplines. Secondly, experiential learning may inadvertently exclude those who lack the time, resources, or physical capability to participate fully. Thirdly, fostering transparent communication may be inhibited by privacy concerns, proprietary interests, or fear of misuse. Fourthly, ensuring equitable access to data may inadvertently result in unqualified individuals misinterpreting or misusing complex data sets. Lastly, developing context-specific applications could potentially lead to an overemphasis on local needs and overlook broader, interconnected environmental issues. These critiques underscore the need to approach these recommendations not as definitive solutions, but as starting points for ongoing discussions and adaptations within the ever-evolving realm of remote sensing and GIS technology.

Additionally, this study's limitations include the potential biases in participant selection and the challenges associated with capturing a representative sample of stakeholder perspectives.

Moreover, the rapidly evolving nature of remote sensing technology means that some of the findings may not apply to emerging capabilities and applications:

- Cost of satellite imagery
- Access to hard- & software
- Lack of local capacity
- Uncertainty around free data & software
    - *Will it be discontinued? Will it be maintained and documented?*
- Difficulties carrying out fieldwork.
    - These challenges have been reported for > 20 years, and are largely the same today.

Engaging Impacted Persons in geospatial projects promises a richer, more holistic approach by incorporating local expertise. However, this integration isn't without complexities. The first challenge lies in the variability of local knowledge. While the depth and nuances of community insights are invaluable, they can vary widely, potentially introducing inconsistencies in data interpretation and application. This diversity, while a strength, can pose challenges in standardizing data collection and interpretation.

Cultural and language barriers further add layers of complexity. Effective communication is the bedrock of any successful engagement, but differences in language, customs, and traditions can sometimes create misinterpretations or misunderstandings. Addressing these differences requires sensitivity, patience, and often additional resources. Another significant concern is the logistical and resource constraints. Comprehensive community engagement might require extended timelines and increased budgets. The challenges can escalate when working in remote or challenging terrains, where accessibility itself can become a hurdle. Merging local insights with scientific methodologies is a delicate balance. While local knowledge provides context, scientific methodologies ensure rigor. Striking the right balance between the two, ensuring that one doesn't overshadow the other, can be intricate. Ethical considerations, from informed consent to data ownership and rights, are also paramount. The engagement process must be transparent, respectful, and must prioritize the rights and wishes of the community.

4.1.2 Impactes Persons

As our society becomes increasingly reliant on geospatial technologies, it is crucial to critically examine their broader implications for equity and inclusivity. To navigate the intricate landscape of community engagement in geospatial projects, a structured and empathetic approach is essential. Collaborative research frameworks can offer a solution to individual burdens. By fostering team-based research, tasks can be distributed, resources shared, and diverse expertise leveraged. Such collaboration becomes even more potent when boundary-spanning organizations are involved. These entities, designed to bridge academia, communities, and policymakers, can provide the necessary expertise and tools to streamline engagement, ensuring that it's effective and meaningful.

Local partnerships can further augment the engagement process. Institutions like local universities, NGOs, and community organizations bring with them a wealth of local knowledge and logistical support. They can act as bridges, helping researchers navigate the local context and ensuring that community voices are genuinely represented. Training initiatives are pivotal in this landscape. By equipping researchers with the skills needed for community engagement, they can navigate the complexities of local interactions with confidence and efficacy. These trainings can cover a range of topics, from cultural sensitivities to effective communication strategies.

The digital age offers tools that can significantly enhance community engagement. Digital platforms, from communication tools to data collection apps, can provide cost-effective solutions, making engagement more accessible and wide-reaching. Feedback mechanisms, both for researchers and the community, can ensure that the engagement process remains dynamic, adjusting to challenges and evolving based on insights. Finally, emphasizing the shared benefits of projects, from mutual learning to tangible project outcomes, can foster a sense of joint purpose. This mutual understanding and respect can lay the foundation for geospatial projects that are not only scientifically robust but also community-centric and inclusive.

5 References

Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J., ... & Vincent, G. (2019). A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests. *Remote Sensing*, 11(9), 1086.

Acevedo, M. F., Callicott, J. B., Monticino, M., Lyons, D., Palomino, J., Rosales, J., ... & Vilanova, E. (2008). Models of natural and human dynamics in forest landscapes: Cross-site and cross-cultural synthesis. *Geoforum*, 39(2), 846-866.

Adler, C., Hirsch Hadorn, G., Breu, T., Wiesmann, U., & Pohl, C. (2018). Conceptualizing the transfer of knowledge across cases in transdisciplinary research. *Sustainability science*, 13(1), 179-190.

Agrawal, A. (1995). Dismantling the divide between indigenous and scientific knowledge. *Development and change*, 26(3), 413-439.

Allan, A., Barbour, E., Nicholls, R. J., Hutton, C., Lim, M., Salehin, M., & Rahman, M. M. (2022). Developing socio-ecological scenarios: A participatory process for engaging stakeholders. *Science of the Total Environment*, 807, 150512.

Arnaiz-Schmitz, C., Schmitz, M. F., Herrero-Jáuregui, C., Gutiérrez-Angonese, J., Pineda, F. D., & Montes, C. (2018). Identifying socio-ecological networks in rural-urban gradients: Diagnosis of a changing cultural landscape. *Science of the Total Environment*, 612, 625-635.

Aval, J., Demuynck, J., Zenou, E., Fabre, S., Sheeren, D., Fauvel, M., ... & Briottet, X. (2018). Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 197-210.

Berger, C., Bieri, M., Bradshaw, K., Brümmer, C., Clemen, T., Hickler, T., ... & Twine, W. (2019). Linking scales and disciplines: an interdisciplinary cross-scale approach to supporting climate-relevant ecosystem management. *Climatic Change*, 156(1), 139-150.

Bloch, S. (2019, October 3). Reflective Pavement May Be Less Cool Than It Seems. CityLab. https://www.bloomberg.com/news/articles/2019-10-03/reflective-pavement-may-be-less-cool-than-it-seems

Blue, B., & Brierley, G. (2016). 'But what do you measure?'Prospects for a constructive critical physical geography. *Area*, 48(2), 190-197.

Carroll, S. R., Rodriguez-Lonebear, D., & Martinez, A. (2019). Indigenous data governance: strategies from United States native nations. *Data Science Journal*, 18.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., ... & Hudson, M. (2020). *The CARE Principles for Indigenous Data Governance*.

Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., ... & Mitchell, R. B. (2003). Knowledge systems for sustainable development. *Proceedings of the national academy of sciences*, 100(14), 8086-8091.

Caughlin, T. T., Graves, S. J., Asner, G. P., Tarbox, B. C., & Bohlman, S. A. (2019). High-resolution remote sensing data as a boundary object to facilitate interdisciplinary collaboration. In *Collaboration Across Boundaries for Social-Ecological Systems Science* (pp. 295-326). Palgrave Macmillan, Cham.

Chrisman, N. (1999, October). Trading zones or boundary objects: Understanding incomplete translations of technical expertise. In *4S meetings*, San Diego.

Cumming, G. S., Cumming, D. H., & Redman, C. L. (2006). Scale mismatches in social-ecological systems: causes, consequences, and solutions. *Ecology and society*, 11(1).C

Dennis, R. A., Mayer, J., Applegate, G., Chokkalingam, U., Colfer, C. J. P., Kurniawan, I., ... & Tomich, T. P. (2005). Fire, people and pixels: linking social science and remote sensing to understand underlying causes and impacts of fires in Indonesia. *Human Ecology*, 33(4), 465-504.

Dinko, D. H., & Nyantakyi-Frimpong, H. (2023). The prospects and challenges of using drone-based participatory mapping in human–environment research. *The Professional Geographer*, 75(3), 441-451.

Enquist, C. A., Jackson, S. T., Garfin, G. M., Davis, F. W., Gerber, L. R., Littell, J. A., ... & Shaw, M. R. (2017). Foundations of translational ecology. *Frontiers in Ecology and the Environment*, 15(10), 541-550.

Enqvist, J. P., West, S., Masterson, V. A., Haider, L. J., Svedin, U., & Tengö, M. (2018). Stewardship as a boundary object for sustainability research: linking care, knowledge and agency. *Landscape and Urban Planning*, 179, 17-37.

Escobar, A. (1996). Construction nature: Elements for a post-structuralist political ecology. *Futures*, 28(4), 325-343.

Foster, A., & Dunham, I. M. (2015). Volunteered geographic information, urban forests, & environmental justice. *Computers, Environment and Urban Systems*, 53, 65-75.

Galvin, K. A., Boone, R. B., Smith, N. M., & Lynn, S. J. (2001). Impacts of climate variability on East African pastoralists: linking social science and remote sensing. *Climate Research*, 19(2), 161-172.

Garmestani, A., Twidwell, D., Angeler, D. G., Sundstrom, S., Barichievy, C., Chaffin, B. C., ... & Allen, C. R. (2020). Panarchy: opportunities and challenges for ecosystem management. *Frontiers in Ecology and the Environment*, 18(10), 576-583.

Geoghegan, J., Pritchard, L., Ogneva-Himmelberger, Y., Chowdhury, R. R., Sanderson, S., & Turner, B. L. I. (1998). "Socializing the pixel" and "Pixelizing the social" in land-use and land-cover change. In *People and pixels*. Washington, DC: National Academy Press.

Goodrich, K. A., Sjostrom, K. D., Vaughan, C., Nichols, L., Bednarek, A., & Lemos, M. C. (2020). Who are boundary spanners and how can we support them in making knowledge more actionable in sustainability fields?. *Current Opinion in Environmental Sustainability*, 42, 45-51.

Haraway, D. (2016). ‘Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective'. *In Space, Gender, Knowledge: Feminist Readings* (pp. 53-72). Routledge.

Hartmann, H., Adams, H. D., Anderegg, W. R., Jansen, S., & Zeppel, M. J. (2015). Research frontiers in drought-induced tree mortality: crossing scales and disciplines. *New Phytologist*, 205(LA-UR-14-29321).

Heider, K., Lopez, J. M. R., & Scheffran, J. (2018). The potential of volunteered geographic information to investigate peri-urbanization in the conservation zone of Mexico City. *Environmental monitoring and assessment*, 190(4), 1-17.

Hoffmann, S., Pohl, C., & Hering, J. G. (2017). Methods and procedures of transdisciplinary knowledge integration: empirical insights from four thematic synthesis processes. *Ecology and Society*, 22(1).

Hulley, G. C., Dousset, B., & Kahn, B. H. (2020). Rising trends in heatwave metrics across southern California. *Earth's Future*, 8(7), e2020EF001480.

Isager, L., & Broge, N. H. (2007). Combining remote sensing and anthropology to trace historical land-use changes and facilitate better landscape management in a sub-watershed in North Thailand. *Landscape Research*, 32(2), 147-169.

James, C. A., Francis, T. B., Baker, J. E., Georgiadis, N., Kinney, A., Magel, C., ... & Wright, C. W. (2022). A boundary spanning system supports large-scale ecosystem-based management. *Environmental Science & Policy*, 133, 137-145.

Jones Jr, J. W. (2011). Using LiDAR Data and Geographical Information System (GIS) Technology to Assess Municipal Street Tree Inventories. Mississippi State University.

Kerr, J. T., & Ostrovsky, M. (2003). From space to species: ecological applications for remote sensing. *Trends in ecology & evolution*, 18(6), 299-305.

Keshkamat, S. S., Kooiman, A., van Maarseveen, M. F. A. M., van der Veen, A., & Zuidgeest, M. H. P. (2012). A boundary object for scale selection—Moderating differences and synergising understanding. *Ecological economics*, 76, 15-24.

Kramer, D. B., Hartter, J., Boag, A. E., Jain, M., Stevens, K., Nicholas, K. A., ... & Liu, A. J. (2017). Top 40 questions in coupled human and natural systems (CHANS) research. *Ecology and Society*, 22(2).

Kugler, T. A., Grace, K., Wrathall, D. J., de Sherbinin, A., Van Riper, D., Aubrecht, C., ... & Van Den Hoek, J. (2019). People and Pixels 20 years later: the current data landscape and research trends blending population and environmental data. *Population and environment*, 41(2), 209-234.

Lave, R., Wilson, M. W., Barron, E. S., Biermann, C., Carey, M. A., Duvall, C. S., ... & Van Dyke, C. (2014). Intervention: Critical physical geography. The Canadian Geographer/Le *Géographe canadien*, 58(1), 1-10.

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., ... & Taylor, W. W. (2007). Complexity of coupled human and natural systems. *Science*, 317(5844), 1513-1516.

Liverman, D. M., & Cuesta, R. M. R. (2008). Human interactions with the Earth system: people and pixels revisited. Earth Surface Processes and Landforms: The Journal of the British *Geomorphological Research Group*, 33(9), 1458-1471.

Rindfuss, R. R., Stern, P. C., Liverman, D., Moran, E. F., & Rindfuss, R. R. (1998). People and Pixels: Linking Remote Sensing and Social Science. *People and Pixels: linking remote sensing and social science*.

Mars, R., & Barube, C. (Hosts). (2023, July 18). Shade Redux (Ep. 545) [Audio podcast episode]. In *99% Invisible*. Radiotopia. https://99percentinvisible.org/episode/shade-redux/

McGinnis, M. D., & Ostrom, E. (2014). Social-ecological system framework: initial changes and continuing challenges. *Ecology and society*, 19(2).

Michaels, S., & Tyre, A. J. (2012). How indeterminism shapes ecologists' contributions to managing socio‐ecological systems. *Conservation Letters*, 5(4), 289-295.

Niemiec, R. M., Asner, G. P., Brodrick, P. G., Gaertner, J. A., & Ardoin, N. M. (2018). Scale-dependence of environmental and socioeconomic drivers of albizia invasion in Hawaii. *Landscape and Urban Planning*, 169, 70-80.

Oestreich, W. K., Chapman, M. S., & Crowder, L. B. (2020). A comparative analysis of dynamic management in marine and terrestrial systems. *Frontiers in Ecology and the Environment*, 18(9), 496-504.

Olbrich, P. (2019). Technological expectations and global politics: three waves of enthusiasm in nongovernmental remote sensing. *Space Policy*, 47, 107-116.

Ordway, E. M., Elmore, A. J., Kolstoe, S., Quinn, J. E., Swanwick, R., Cattau, M., ... & Wilson, C. (2021). Leveraging the NEON Airborne Observation Platform for socio‐environmental systems research. *Ecosphere*, 12(6), e03640.

Peek, L., Tobin, J., Adams, R. M., Wu, H., & Mathews, M. C. (2020). A framework for convergence research in the hazards and disaster field: The natural hazards engineering research infrastructure CONVERGE facility. *Frontiers in Built Environment*, 6, 110.

Perz, S. G. (2019). Introduction: Collaboration across boundaries for social-ecological systems science. In *Collaboration across boundaries for social-ecological systems science* (pp. 1-33). Palgrave Macmillan, Cham.

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H., & Turner, W. (2014). Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology*, 51(4), 839-848.

G Pricope, N., L Mapes, K., & D Woodward, K. (2019). Remote sensing of human–environment interactions in global change research: A review of advances, challenges and future directions. *Remote Sensing*, 11(23), 2783.

Ramage, S., Carroll, S. R., Hudson, M., David-Chavez, D. M., Kenote, T., Anderson, J., ... & Shanley, L. (2020, December 7-9). *Indigenous Data Sovereignty* [Conference session].

GEO Indigenous Summit, Virtual. https://www.earthobservations.org/indigenoussummit2020.php

Rattling Leaf Sr., J., Lenkokwai, R., Sudlovenick, C., Schmidt, C., DeRoy, S., Zwartjes, L., ... & Prados, A. (2020, December 7-9). *Education / Intergenerational Knowledge Transfer* [Conference session]. GEO Indigenous Summit, Virtual. https://www.earthobservations.org/indigenoussummit2020.php

Rattling Leaf Sr, J. (2022). What is Traditional Ecological Knowledge and why does it matter?. *Frontiers in Ecology and the Environment*, 20(1), 3-3.

Rissman, A. R., Owley, J., L'Roe, A. W., Morris, A. W., & Wardropper, C. B. (2017). Public access to spatial data on private-land conservation. *Ecology and Society*, 22(2).

Robbins, P. (2001). Fixed categories in a portable landscape: the causes and consequences of land-cover categorization. *Environment and planning* A, 33(1), 161-179.

Robbins, P. (2003). Beyond ground truth: GIS and the environmental knowledge of herders, professional foresters, and other traditional communities. *Human Ecology*, 31(2), 233-253.

Roy, R. D. 2018. Decolonise science–time to end another imperial era.*The Conversation*.

Safford, H. D., Sawyer, S. C., Kocher, S. D., Hiers, J. K., & Cross, M. (2017). Linking knowledge to action: the role of boundary spanners in translating ecology. *Frontiers in Ecology and the Environment*, 15(10), 560-568.

Sakakibara, C. (2010). Kiavallakkikput Aāviq (Into the Whaling Cycle): Cetaceousness and Global Warming among the Iñupiat of Arctic Alaska. *Annals of the Association of American Geographers*, 100(4).

Stavros, E. N. (2021). Wicked Problems need WKID Innovation: Innovation as a Process to Develop a Disruptive Technology Product. *Research-Technology Management*, 65(1), 39-47.

Sundberg, J. (2004). Identities in the making: conservation, gender and race in the Maya Biosphere Reserve, Guatemala. *Gender, Place & Culture*, 11(1), 43-66.

Svoray, T., Perevolotsky, A., & Atkinson, P. M. (2013). Ecological sustainability in rangelands: the contribution of remote sensing. *International Journal of Remote Sensing*, 34(17), 6216-6242.

Swartz, T. H., Palermo, A. G. S., Masur, S. K., & Aberg, J. A. (2019). The science and value of diversity: closing the gaps in our understanding of inclusion and diversity. *The Journal of infectious diseases*, 220(Supplement_2), S33-S41.

Tadaki, M., Brierley, G., Dickson, M., Le Heron, R., & Salmond, J. (2015). Cultivating critical practices in physical geography. *The Geographical Journal*, 181(2), 160-171.

Tellman, B. (2018). Clandestine Political Economic Activity in Landscape Dynamics-Linking Pattern to Process. *People Pixels Revisted*, 1-15.

Temple, J. (2020). A Trillion Trees" is a Great Idea–that could Become a Dangerous Climate Distraction. *Technol Rev*.

Turner, V. K., Benessaiah, K., Warren, S., & Iwaniec, D. (2015). Essential tensions in interdisciplinary scholarship: navigating challenges in affect, epistemologies, and structure in environment–society research centers. *Higher Education*, 70(4), 649-665.

Turner Ii, B. L., Esler, K. J., Bridgewater, P., Tewksbury, J., Sitas, N., Abrahams, B., ... & Mooney, H. (2016). Socio-Environmental Systems (SES) Research: what have we learned and how can we use this information in future research programs. *Current Opinion in Environmental Sustainability*, 19, 160-168.

Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. *New Phytologist*, 186(4), 795-816.

Walker, P. A., & Peters, P. E. (2007). Making sense in time: remote sensing and the challenges of temporal heterogeneity in social analysis of environmental change—cases from Malawi. *Human Ecology*, 35(1), 69-80.

Walters, B. B., Rönnbäck, P., Kovacs, J. M., Crona, B., Hussain, S. A., Badola, R., ... & Dahdouh-Guebas, F. (2008). Ethnobiology, socio-economics and management of mangrove forests: A review. *Aquatic Botany*, 89(2), 220-236.

Wilson, N. (2020). The cost of living with wildlife. *Frontiers in Ecology and the Environment*, 18(6), 314-314.

Whyte, K. 2017. What do indigenous knowledges do for indigenous peoples?.Forthcoming in *Keepers of the Green World: Traditional Ecological Knowledge and Sustainability*. Cambridge University Press, UK.

Vurdubakis, T., & Rajão, R. (2022). Envisioning Amazonia: Geospatial technology, legality and the (dis) enchantments of infrastructure. *Environment and Planning E: Nature and Space*, 5(1), 81-103.

Zhou, T., Popescu, S. C., Lawing, A. M., Eriksson, M., Strimbu, B. M., & Bürkner, P. C. (2018). Bayesian and classical machine learning methods: a comparison for tree species classification with LiDAR waveform signatures. *Remote Sensing*, 10(1), 39.

Zubrow, A. A. (2003). Mapping tension: remote sensing and the production of a statewide land cover map. *Human ecology*, 31(2), 281-307.

Chapter 5.

Conclusion

This dissertation serves as a valuable contribution to the field of urban forestry management and highlights the need for interdisciplinary collaboration in addressing complex challenges facing urban environments. By incorporating both technical and sociopolitical perspectives, my chapters provide insights that can inform practical applications and recommendations for future research. The sociopolitical dimension of science-based solutions, as explored in the fourth chapter, underscores the imperative need to include Impacted Persons in the data exchange continuum. Their inclusion not only democratizes the decision-making process but also augments remote sensing studies with richer ground truthing, recognizing and harnessing local knowledge. Technical investigations in chapters two and three have offered valuable tools to urban forestry managers in Los Angeles County. The LiDAR processing analysis revealed the advantages of using the Dalponte2016 algorithm to segment individual tree crowns, which will provide technical assistance for future LiDAR acquisitions, ensuring that County managers are equipped with the most effective tools for the task.

Furthermore, the development of the Random Forest algorithm in Google Earth Engine showcased the potential of integrating spectral and structural data for tree species classification. With an overall accuracy rate of 89% for 20 species across three pilot sites, this model showcases the power of combining robust classification methods with cloud-based platforms, eliminating the need for cumbersome data downloads. In sum, this dissertation, through its methodical investigations and critical analyses, offers a holistic view of remote sensing in urban forestry. It emphasizes the symbiotic relationship between technical prowess and sociopolitical understanding, ultimately aiming for more sustainable and equitable urban forest management practices in Los Angeles County and potentially beyond.

Moving forward, my research will continue to focus on enhancing interdisciplinary collaboration in urban forestry management, with a keen awareness of the diverse challenges that cities of varying sizes and management practices face. A key priority will be advancing tree crown segmentation methods and refining remote sensing technologies. These improvements are crucial for providing more accurate and detailed data on urban forests, which is fundamental for effective management and policy-making. However, it is important to recognize the limitations and potential obstacles in this area of research. One major challenge is the disparity in data access and quality between different urban areas. Larger municipalities like Los Angeles County often have more resources and advanced technologies at their disposal, enabling more sophisticated data collection and analysis. In contrast, smaller cities may lack such capabilities, leading to less detailed or accurate data. This disparity can result in uneven urban forestry management and policy implementation.

Additionally, future work must account for varying management practices across different urban settings. Urban forests in different cities are managed under diverse policies and guidelines, influenced by distinct sociopolitical and environmental contexts. This diversity necessitates tailored approaches to urban forestry management, ensuring that solutions are effective and relevant to each unique urban environment. Another critical aspect is the continued integration of technical expertise with socio-political understanding. My work has underscored the importance of this integration in providing a comprehensive understanding of the dynamics at play in urban forestry management. Future research should continue to bridge this gap, ensuring that technical advancements in remote sensing and tree management are informed by and responsive to the needs and dynamics of the communities they serve. While future work holds significant potential for enhancing urban forestry management, it must be approached with a realistic understanding of the varying capabilities and contexts of different urban environments. Balancing technical innovation with socio-political awareness will be key to developing sustainable and equitable solutions in urban forestry across diverse urban landscapes.

Appendix A. Historical Aerial Photographs of Los Angeles County Pilot Sites

A.1 Aerial photographs of Altadena in 1933 (a) and 1966 (b). Spence Air Photos Collection, UCLA Department of Geography Air Photo Archives.

a)

b)

A.2 Aerial photographs of East Los Angeles in 1931 (a) and 1951 (b). Spence Air Photos Collection, UCLA Department of Geography Air Photo Archives.

a)

b)

A.3 Aerial photographs of Marina del Rey in 1938 (a) and 1951 (b). Spence Air Photos Collection, UCLA Department of Geography Air Photo Archives.

a)

b)

A.4 References

Historical aerial imagery courtesy of The Benjamin and Gladys Thomas Air Photo Archives, Spence Air Photos Collection, UCLA Department of Geography.

Appendix B.

B.1 Applied Remote Sensing Survey: Survey Questions and Anonymous Responses

Questions

1. Your role or position.
2. Years of experience in geospatial science.
3. Which geospatial technologies do you most frequently use?
4. On a scale of 1 to 5, how important is it to follow scientific methods and standards in your projects?
5. Which type of interested parties do you usually collaborate with?
6. Have you ever had to modify your methodology to cater to an interested party's request?
7. If yes or sometimes, did this change impact the scientific validity of the results?
8. Briefly describe a situation where there was a conflict between scientific rigor and client's expectation, and how you resolved it.
9. On a scale of 1 to 5, how effectively do you believe you communicate complex geospatial data to non-experts?
10. What do you believe is the most challenging aspect of collaborating with non-experts in geospatial projects?

Responses

Respondent	Answers
Response 1	Urban Sociologist
Response 1	3-5 years
Response 1	GIS
Response 1	5
Response 1	Government, NGOs
Response 1	Yes
Response 1	Negatively
Response 1	In my recent collaboration with city planners, government officials were pressing for insights into specific urban development plans. I had to assert the importance of conducting comprehensive sociological studies, which are crucial for sustainable urban planning. Quick results could overlook vital community needs and lead to ineffective or detrimental policies. By integrating detailed social research, we ensure that development plans are equitable and aligned with the actual requirements of the residents. Despite the push for speed, I advocated for a balance between expediency and meticulous, community-focused planning. I've found that

	integrating quantitative data with qualitative community feedback is vital for a holistic understanding.
Response 1	4
Response 1	Balancing urban development with demographic and socioeconomic insights I aim to ensure that both historical trends and future projections inform urban planning initiatives.
Response 10	Public Policy Analyst
Response 10	5-6 years
Response 10	Remote Sensing
Response 10	5
Response 10	NGOs, Government
Response 10	Sometimes
Response 10	No impact
Response 10	The pushback on recognizing climate change impacts from some officials was concerning. To counter this, I compiled longitudinal temperature records and frequency analyses of extreme weather events, which clearly documented the changing patterns. I supplemented this with case studies from similar regions where climate change mitigation strategies had successfully been implemented, highlighting both environmental and economic benefits. By presenting this evidence in a series of workshops, I was able to foster a more informed dialogue and emphasize the critical need for proactive measures. This approach led to a gradual shift in perspective, with officials eventually acknowledging the significance of the issue and considering policy adjustments. It's critical to bridge the gap between scientific data and public perception to foster informed decision-making.
Response 10	4
Response 10	Overcoming climate change denial or misinformation I'm committed to engaging with the community through educational programs to demystify scientific data.
Response 11	Ecologist, Mountain ecosystems
Response 11	1-2 years
Response 11	GIS
Response 11	4
Response 11	Academics
Response 11	No

Response 11	No impact
Response 11	Tourism stakeholders wanted more mountain activities. I highlighted the fragile nature of alpine ecosystems and proposed sustainable alternatives. I'm passionate about finding the equilibrium where human enjoyment doesn't compromise ecological sustainability.
Response 11	4
Response 11	Addressing tourist interests while conserving ecosystems My strategy includes developing environmental education programs for tourists and stakeholders alike.
Response 12	Coastal Policy Expert
Response 12	7-8 years
Response 12	Remote Sensing
Response 12	5
Response 12	Government
Response 12	Yes
Response 12	Negatively
Response 12	Local businesses opposed certain coastal policies. I organized community meetings to address concerns and find a middle ground. My goal is to advocate for policies that are informed by long-term ecological forecasts, not just short-term economic gains.
Response 12	5
Response 12	Balancing business interests with research findings I prioritize collaborative projects that align business models with sustainable practices.
Response 13	Remote Sensing Analyst
Response 13	5-10 years
Response 13	Remote Sensing, GIS
Response 13	5
Response 13	Private firms
Response 13	Sometimes
Response 13	Positively

Response 13	Farmers wanted to use older, less accurate data for land monitoring. I demonstrated the benefits of newer technology for better crop yields. I believe in leveraging the latest in technology to provide the most accurate data to support sustainable farming practices.
Response 13	4
Response 13	Managing expectations I focus on communicating the long-term benefits of precision agriculture to stakeholders.
Response 14	Geospatial Intelligence Analyst
Response 14	11-20 years
Response 14	GIS, GPS
Response 14	5
Response 14	Government
Response 14	Yes
Response 14	Negatively
Response 14	During a flood event, there was a push for quicker results. I had to balance speed with accuracy to provide reliable flood maps. I had to leverage not only my technical expertise but also my crisis management skills.
Response 14	4
Response 14	Balancing between quick results and accuracy I strive to maintain precision without sacrificing the urgency required in emergency situations.
Response 15	Geospatial Projects Director
Response 15	11-20 years
Response 15	GIS
Response 15	4
Response 15	Private firms
Response 15	No
Response 15	No impact
Response 15	A client wanted to develop a piece of land identified as having high ecological value. I provided alternatives that were less impactful. This required not just technical know-how but also diplomacy and strategic negotiation skills.

Response 15	3
Response 15	Explaining the implications of geospatial data My approach involves translating complex GIS data into actionable insights for all stakeholders.
Response 16	Ecosystem Restoration Adviser
Response 16	5-10 years
Response 16	Remote Sensing, GIS
Response 16	5
Response 16	Public stakeholders, NGOs
Response 16	Yes
Response 16	No impact
Response 16	An NGO wanted to highlight deforestation in an area, but the data showed mixed results. We collaborated on a more nuanced report. I found myself in the midst of a moral dilemma, balancing ecological integrity with the needs of the community.
Response 16	5
Response 16	Ensuring data is interpreted correctly I navigate these challenges by advocating for the voiceless environment, ensuring its story is told and understood.
Response 17	Machine Learning Insights Architect
Response 17	Less than 5 years
Response 17	Remote Sensing, GPS
Response 17	4
Response 17	Private firms
Response 17	Sometimes
Response 17	Positively
Response 17	The company wanted real-time data integration, but this would have compromised data quality. I proposed a slight delay for better accuracy. The challenge was to innovate at the speed of thought, making data sing in real-time.
Response 17	3
Response 17	Translating technical jargon to understandable insights I translate algorithms into business strategies, bridging the gap between data science and practical application.

Response 18	Strategic Operations Overseer
Response 18	More than 20 years
Response 18	GIS, GPS
Response 18	4
Response 18	Private firms
Response 18	No
Response 18	No impact
Response 18	There was pressure to expedite shipments at the expense of environmental concerns. I used data to show long-term benefits of eco-friendly routes. It was a test of my ability to remain unflappable under pressure and make data-driven decisions swiftly.
Response 18	3
Response 18	Making long-term decisions based on data My focus remains on aligning immediate action with the overarching strategic vision for sustainable growth.
Response 19	Spatial Data Strategist
Response 19	5-10 years
Response 19	Remote Sensing, GIS
Response 19	5
Response 19	Public stakeholders
Response 19	Yes
Response 19	Positively
Response 19	Park officials wanted to open a new trail, but satellite images showed it could disrupt wildlife. We found an alternative path. Devising a new trail required a balance of ecological foresight and visitor satisfaction.
Response 19	4
Response 19	Balancing recreational use with conservation My approach is to synthesize environmental stewardship with community access.
Response 2	Marine Biologist
Response 2	3-5 years

Response 2	Remote Sensing
Response 2	5
Response 2	Academics, Government
Response 2	Sometimes
Response 2	Positively
Response 2	Funding agencies wanted to focus on popular marine species. I advocated for studying lesser-known but ecologically crucial species. I had to champion the less charismatic species that play key ecosystem roles.
Response 2	5
Response 2	Highlighting the importance of lesser-known marine species I strive to shine a light on all marine life, ensuring a holistic approach to ocean health.
Response 20	Predictive Geospatial Analyst
Response 20	More than 20 years
Response 20	Remote Sensing
Response 20	5
Response 20	Government
Response 20	No
Response 20	No impact
Response 20	During a storm event, the media wanted early predictions. I emphasized the need for more data before making accurate storm path predictions. The key was to deliver accurate predictions while managing the chaotic nature of storms.
Response 20	5
Response 20	Managing public expectations during high-stakes events It's about balancing precision with the urgency demanded by real-time events.
Response 21	Innovation Pathway Leader
Response 21	1-2 years
Response 21	Remote Sensing
Response 21	5

Response 21	Academics, Private firms
Response 21	Yes
Response 21	No impact
Response 21	Investors wanted quicker results from our tech platform. Collaborated with academics to ensure scientific accuracy. Steering a startup means aligning investor timelines with real product development cycles.
Response 21	4
Response 21	Balancing tech innovation with scientific accuracy I prioritize transparency in innovation, ensuring stakeholders understand the development journey.
Response 22	Social Impact Coordinator
Response 22	2-3 years
Response 22	GIS
Response 22	4
Response 22	Academics, NGOs
Response 22	Sometimes
Response 22	Negatively
Response 22	Donors wanted specific results to support their cause. I emphasized the importance of unbiased research and data. The challenge was to align donor desires with on-the-ground realities and longer-term goals.
Response 22	4
Response 22	Navigating between donor expectations and scientific rigor I endeavor to mediate between immediate results and the need for sustainable impact.
Response 23	Urban Planner
Response 23	4-5 years
Response 23	Remote Sensing
Response 23	5
Response 23	Government, Academics
Response 23	No

Response 23	No impact
Response 23	City officials wanted quick solutions for traffic. Used remote sensing data to propose long-term effective solutions. I used data-driven analytics to illustrate the potential outcomes of various traffic solutions.
Response 23	5
Response 23	Communicating the importance of long-term planning My focus is on marrying immediate fixes with the broader vision of urban sustainability.
Response 24	Professor of Environmental Studies
Response 24	More than 20 years
Response 24	Remote Sensing, GIS
Response 24	5
Response 24	Public stakeholders, Private firms
Response 24	No
Response 24	No impact
Response 24	Had to convince local authorities about the importance of preserving a local forest. Presented empirical evidence from satellite images to show the forest's ecological significance. I leveraged my extensive publication record to underscore the validity of my arguments.
Response 24	5
Response 24	Explaining technical details I aim to distill complex research findings into actionable insights for non-academic audiences.
Response 25	Ph.D. Student
Response 25	5-10 years
Response 25	Remote Sensing, GIS, GPS
Response 25	5
Response 25	Public stakeholders
Response 25	Sometimes
Response 25	Positively

Response 25	A project sponsor wanted quicker results, but this would have required using lower resolution images. I demonstrated the value of higher resolution data using past research. Patience and persistence were key as I balanced sponsor expectations with academic rigor.
Response 25	4
Response 25	Managing expectations
Response 26	Associate Professor
Response 26	11-20 years
Response 26	Remote Sensing, GIS
Response 26	5
Response 26	Public stakeholders, Individual clients
Response 26	Yes
Response 26	Negatively
Response 26	Dealt with a tourism company wanting to minimize the impact of coastal erosion on their reports. Demonstrated the long-term value of sustainable coastal management. I drew upon my fieldwork to provide a nuanced perspective that transcended commercial interests.
Response 26	5
Response 26	Balancing between scientific accuracy and stakeholder interests
Response 27	Senior Research Fellow in Urban Studies
Response 27	More than 20 years
Response 27	Remote Sensing, GPS
Response 27	4
Response 27	Government, Private firms
Response 27	Yes
Response 27	No impact
Response 27	Faced pressure from a local government to underreport the severity of a drought. Collaborated with international NGOs to present a unified and accurate assessment. I navigated political pressures while maintaining academic integrity in my urban development research.
Response 27	4

Response 27	Ensuring the data is not misused or misinterpreted I emphasize the ethical use of data to inform sustainable urban planning.
Response 28	Assistant Professor
Response 28	5-10 years
Response 28	Remote Sensing, GIS
Response 28	5
Response 28	Interest groups, Public stakeholders
Response 28	Sometimes
Response 28	Positively
Response 28	Had to balance between providing detailed satellite images for a defense project and ensuring that sensitive information is not compromised. Used data encryption and secure channels. Precision was paramount, as the satellite data I provided would inform critical environmental policies.
Response 28	5
Response 28	Security concerns Ensuring the security of sensitive data is a top priority, as it has far-reaching implications.
Response 29	Environmental Analyst
Response 29	Less than 5 years
Response 29	GIS, GPS
Response 29	4
Response 29	Private firms
Response 29	No
Response 29	No impact
Response 29	City officials wanted a more optimistic portrayal of urban heat islands. I organized a workshop demonstrating the health impacts of heat islands.
Response 29	3
Response 29	Translating research into actionable insights
Response 3	Climate Impact Forecaster

Response 3	5-10 years
Response 3	Remote Sensing
Response 3	5
Response 3	Government
Response 3	Sometimes
Response 3	No impact
Response 3	Media wanted alarming reports on ice melts. I ensured data accuracy and presented a balanced view. I aimed to communicate the seriousness of climate trends while avoiding unnecessary alarmism.
Response 3	5
Response 3	Avoiding sensationalism in climate data I am committed to providing a balanced view that is informative yet measured.
Response 30	Researcher/Academic
Response 30	Less than 5 years
Response 30	Remote Sensing, GIS (Geographic Information Systems)
Response 30	5
Response 30	Public stakeholders, Interest groups
Response 30	Sometimes
Response 30	Sometimes
Response 30	During a forest conservation project, there was a push to use certain satellite images that were readily available but lacked the resolution required for accurate forest density mapping. I had to explain the importance of using higher-resolution data to ensure the reliability of the study. After some discussions and demonstrations, we decided to invest in acquiring the appropriate satellite data to maintain scientific rigor. I focused on integrating indigenous knowledge with scientific research to inform conservation efforts.
Response 30	4
Response 30	Explaining technical details I demystify scientific data for stakeholders to aid in effective policy-making.
Response 4	Robotics Engineer
Response 4	4-5 years

Response 4	GIS
Response 4	5
Response 4	Private firms
Response 4	No
Response 4	No impact
Response 4	Investors wanted flashy robotics demos. I stressed the importance of foundational research. I had to align investor expectations with the practical realities of robotics engineering.
Response 4	3
Response 4	Balancing tech innovation with foundational research I maintain a balance between showcasing cutting-edge robotics and ensuring technological reliability.
Response 5	Mineral Resource Analyst/Geologist
Response 5	6-7 years
Response 5	GIS
Response 5	5
Response 5	Academics, Private firms
Response 5	Yes
Response 5	Negatively
Response 5	Private firms wanted quick mineral studies. I stressed the importance of understanding geological structures. I insisted on comprehensive studies to avoid costly errors in mineral exploration.
Response 5	4
Response 5	Communicating the importance of thorough geological studies My approach is to clarify the value of meticulous geological surveys in guiding investment.
Response 6	Public Health Specialist
Response 6	8-10 years
Response 6	GIS
Response 6	5
Response 6	NGOs, Government

Response 6	Yes
Response 6	Positively
Response 6	Funding agencies wanted large-scale interventions. I emphasized the need for localized, culturally-sensitive health campaigns. I focused on sustainable, evidence-based interventions rather than quick fixes.
Response 6	5
Response 6	Ensuring public health data is culturally sensitive I ensure that health initiatives are not only effective but also resonate with local cultural values.
Response 7	Environmental Scientist
Response 7	3-5 years
Response 7	GIS
Response 7	5
Response 7	Academics, NGOs
Response 7	Sometimes
Response 7	Positively
Response 7	NGOs wanted immediate action against deforestation. I emphasized the need for thorough studies before interventions. I championed a methodical approach to tackle deforestation without overlooking community needs.
Response 7	5
Response 7	Balancing conservation with economic interests I aim to harmonize the dialogue between ecological preservation and development agendas.
Response 8	Urban Ecologist
Response 8	4-5 years
Response 8	Remote Sensing
Response 8	4
Response 8	Government
Response 8	No
Response 8	No impact

Response 8	Government officials wanted to reduce green spaces for development. I showcased the importance of biodiversity in urban settings. I advocated for the integration of green spaces as vital urban infrastructure.
Response 8	4
Response 8	Explaining the importance of urban biodiversity I translate the ecological significance of urban greenery into urban planning language.
Response 9	Arid Region Environmental Coordinator
Response 9	2-3 years
Response 9	GIS
Response 9	5
Response 9	Academics, Government
Response 9	Yes
Response 9	Negatively
Response 9	Local communities resisted certain conservation measures. I collaborated with them to understand their needs and adapted my approach. I worked closely with local communities to find mutually beneficial conservation strategies.
Response 9	5
Response 9	Navigating local cultural and environmental needs I navigate the complex intersection of environmental science and cultural dynamics.

B.2 Applied Remote Sensing Semi-structured Interview

B.2.1 Interview Questions

1. In your own words, how would you describe the significance of remote sensing in contemporary geospatial applications?
2. Could you tell me more about the types of projects you have recently worked on? How did you incorporate geospatial technology, specifically remote sensing, into these projects?
3. What methods or standards do you adhere to when using geospatial technology?
4. How do you ensure that your clients' needs are met when incorporating geospatial technology into your projects?
5. Could you provide an overview of the key challenges you face when working with remote sensing data in your projects?
6. Are there any specific tools or software platforms you prefer when handling and analyzing remote sensing data? Why?
7. In what ways do you think the advancements in geospatial technology have influenced the expectations of your clients or stakeholders?
8. How do you maintain the balance between cutting-edge technologies in remote sensing and the reliability or familiarity of older methods?
9. How do you ensure the ethical collection and use of data, especially when handling sensitive geospatial information?
10. Can you describe a project where integrating remote sensing posed unique challenges, and how you navigated through them?
11. Are there instances where you had to make trade-offs between data accuracy and other project requirements?
12. How do you stay updated with the rapidly evolving standards and best practices in remote sensing and geospatial sciences?
13. Can you shed light on the interdisciplinary nature of your work? How does collaborating with professionals from other fields influence the application of remote sensing in projects?
14. Are you familiar with feedback loops in an environmental context? What role do feedback loops play in your projects, especially when refining geospatial methodologies?
15. Could you explain the peer review or validation processes you undergo to ensure the scientific integrity of your geospatial applications?
16. How do you handle uncertainties or errors in remote sensing data, and how do you communicate these to clients or stakeholders?
17. Can you describe a time when you had to adapt your approach based on the specific cultural, environmental, or social context of a project?
18. In what ways do you think public perceptions of geospatial technologies, particularly remote sensing, have changed over the years?
19. How do you envision the future of remote sensing, and what role do you see yourself playing in it?
20. Are there lessons or insights from your academic journey in geospatial science that you often find relevant or applicable in your professional projects?

B.2.2 Anonymized Interview Transcript: Interviewee #1

I have about 20 questions. I don't know if we're going to hit them all. I would like to keep this, you know, something to your volunteering, your time here. And so i'm going to try and keep keep this. Within reason and there is a chance that i could ask one question and in your response depending on how detailed you want to be, you may answer the next five questions that i had planned.

Um, so we could end up, you know, only covering five questions and 20 minutes or we could cover all 20 in about 45 minutes. So so i don't feel like that. I, you, i fail. Same way information gathering from an expert like yourself and i'm really grateful that that you're spending time with me today.

So for sure, I guess my first question is in your own words. Oh and i'm sorry if you you know you've noticed i've been reading off my tablet here. Yeah. If at any point you're talking and you see me looking down, i'm not trying to be rude and just taking notes in addition reporting.

Okay, okay. In your own words. How would you describe the significance of remote sensing in contemporary geospatial applications?

Okay, my question, my answer is going to be completely biased to the To the field. Of work that i specialize, which is the carbon market. So, It is of the Of great importance because we, we are going to advance with this technology interview, you're going to be able to monitor.

And verify and report, the aboveground biomass or the carbon that's being either removed, or Uh, stored in in In different porn stars from all over the world. So, I think it is a great important that we start developing that the certificate the the registers start developing tools and methodology, just so we can apply that tool.

Um, could you tell me more about the types of projects you have recently worked on and how did you incorporate you a spatial technology specifically remote sensing into those projects? Sure. So i am a forest carbonado this basically. So all I to you, and i have a forestry background, so i'm a forest engineer with the undergrad.

Masters inclusi enforce management. So yes, mapping is one of a of our, like, i would say steps zero for any of our projects, try. You do a lot of the most sensing for mapping for mapping the area. Uh, when we are doing when we first seeing, if that project is Is worth of our time.

The first thing i do is open a google earth or go to my ArcGIS. And literally, just Analyze that or is that area? That could be here around vancouver but also in indonesia or the areas that i've been working more soft American Central America. So, laryn while i'm looking for is to see How is the poorest covert, how is the land cover and the utilization of the classification of the land water?

Soil, pasture and material forest. So, this is one of the first things that i use that for and then for sure, for the projects, i basically written before station projects, in red projects, so projects, therefore called conservational, force. Um, we did use for one of our projects. Now, these lend classification we are using uh Machine learning algorithm called random forest, but one of my colleagues has that was training that model and developing.

The model is to talk to the public. That was the second first person that i would, Our suggest you So one of the things that we need to prove when we developing a carbon project is okay, if you're doing reforestation and we starting off with pasture, then you need to prove that 10 years before the project started.

Low force was cut. And how do we do that? Well, that's one of the first use of remote sensing getting an image from 10 years prior and then use the land classification model with random or is to distinguish between what was for and what was known forest and show to the auditors show to the That's water.

So the registry look no forest was cold was cut or this is how the area looked like 10 years ago. This is how the area, you know, looked like before the project is studied, And, We're not very far away. But sooner, were you going to be using that even more from monitoring that this person?

So after we validate and verify the credits and being Uh, retired. We need to keep mind during those projects. Talking about projects that last 40 years. So every five years you would need to verify again. So i think in it's very, it's an assumed feature, we will be Uh, using remote sensing for that monitoring too.

So anytime we have our monitoring plots in our inventory, well done. We're just going to use your non-sizing to keep like, seeing how the above ground balance is growing. That's great. I think. Jumping off what he said here. This isn't a question and down, but you did mention art, gios and google earth.

Are there any other software or platform that you utilize to do some some of that initial visual assessment?

No. Yeah i said google earth pearl. Okay. That would be the the easiest ones because we usually get kml files. Okay and articha. Yeah. See google earth pro. And our js are the main softer that i use yet and then google earth engine to get some old image for sure.

Okay. Engine as well. And then Onto assuming anything, but do all the subsequent methods that you use stay within sort of these software, architectures arcGIS google earth engine. Yeah. Okay. And then, Kind of expanding from sort of the initial visual assessment that you can duct. What methods or standards do you hear to when using remote tempting?

Okay, so for project disability again suppose you're my client. You were my client that comes to the company and say look i have this farm, my parents have this area. Could we do a carbon project? Um, Depending on the size and how undecided the people that. What we call the front line here, the frontliners, right?

So those are you wouldn't. You wouldn't come and talk to me as because i'm the technical. So i stay in the basement, crunchy numbers, just so my boss and people from the finance working, look good. So, yes. But then they, they get in doubt very like, uh, This is worth that's when it comes to my head.

We don't have, i don't apply the land classification for that because again We are not making money out of the project yet. So, how much of my time are going to spend into doing a detail analysis, and getting, because we need to get some sort of, like, some acid.

It's some rough assimilates of how much carbon would With that area generate. And then i spend it back to the financial team meeting on the combination models and decide it's worth not worth it. Uh, we don't, so i usually i just do like a visual, it's visual And what i'm looking for is construction, so buildings Um, And basically, Seeing the land cover the difference in.

Forest area, what i'm seeing as that. Usually, about the shade by the texture of the image. Um, Yeah, and then i go to my google, to rgs to basically create shape files of this areas. Estimate the area. The projected area. And then have some sort of accurate estimates of that, how much area we would have for each type of project.

So, okay, of course, station or are we going to do conservation here and based on that area? I do the rough estimates. I wish we could do a more detail using, maybe now that we have this land classification model with random floor is ready. That we could use for our feasibility, for doing the visibility.

Um, because again it's something fast i shouldn't be A lot of work and energy that you put on doing it. It's almost the same amount of work that i would have done for any other project. Because when i'm doing those estimates, i want to be as accurate. As possible.

I'm not looking for precision and looking for accuracy. Well, accuracy and precision blood. Yeah, we want to be accurate because we don't want to be Penalized for, Having some uncertainty and yes, in those Those carbon. Okay. Because we, we gap. Analyzed that. Yes, some errors. So i feel like the devote sensing application and i feel it's going to come to increase that accuracy and transparency.

And doing some trust to our clients and all the stakeholders. Say of how things are done, how things are estimated, how the monitoring is done. And i'm going to talk about monitoring is not just how the forest is growing. But is there any sort of leakage So have the pores being cut or we can monitor.

Natural disturbance was a fire, like trees were blown down. Landed with the monsters who we have access, much more feasible to do this at them. That's great. No i think in the in your answer, there, you kind of touch on a few of my future questions that Like to jump to right now.

Um one of them you're talking about, sort of building trust and accuracy and the fact that you would get penalized And so, i guess, my question here is, how do you handle uncertainties or errors and remote hunting data? And how do you communicate these to clients and stakeholders? And i guess if you could also expand on sort of how you get penalized and foods making, The decision as to whether you would be penalized.

Okay. Yeah. So when i was talking about the penalization is related to begin certainty when we estimating the carbon stalks, So when we estimating in the above ground, by the mass or above ground below ground, It it just, i don't know, depending on the tools that you can include soil.

So, we use a The main registries that we've been working on. The certifications is Vera. Is yes. So PCS has a modules and tools. That tells you how how you get a calculate those the common stocks and how are you going to calculate that uncertainty in those estimates on those askings?

The the module that touches on remote sensing is being prepared. Now, it hasn't been released. So so far There's no. The the use of remote sensing, at least for the carbon word, it's not. I'm gonna call regulated, but it's not, i shouldn't that shouldn't say regulated. What i want to say is like there's no tour methodology.

That. Yeah, decorates that sets. The procedures, the standard operating procedures for using remote sensing than calculating their background by. We could use, then we've been using by, in the end, we will need to come back to the exhaust spreadsheets and for all monitoring mentors. So the uncertainty comes or the penalization comes with the uncertainty when assimating those average Uh, cabin so when you're asking any, so if your standard deviation is pretty high which is pretty common.

If you've talking about flexibility for a station, you have small tree, they have big truth. So how you simply that? Can lead you to. A higher uncertainty. So if you pass an asserting threshold of 10 percent, you're going to start getting finalization. So, the amount that you estimating will be deducted by 15, 20 percent due to that uncertainty.

Wow. Okay. So, i feel that with the use of remorsements Again, all we By our goal is to be accurate, is to be as close as possible to the true value, to actually, what's up there. So i feel that the remote system can help us there. But that doesn't mean that we are giving up on the monitoring.

Network of plots and going to the field and things that. As of course, A lot of experience is on measuring, dbh and deviation hides. So i feel we we are dissenting in media going fast with that, but for our projects, we will always start to some sort of like launcher blocks calibrating those removal since you models making sure that they are, you know, the deer accurate the in making and even understanding that the air that we coming out of that.

Right? So we can control that for the management that's going to come after. Again, i'm talking about projects that it's going to last 40 years. I don't know if i'm going to be involved in the project anymore, but my role is to set up the rules and conditions and all the procedures that it should have.

So i feel the remote sensing comes, at least for the carbon market to help with that monitoring and being more active. Um, The methodology that outlines the procedures for using that in assuming the background mask. At least for vera for VCS, it's not yet released or it should be released pretty soon.

Very we already announced.

Okay.

I guess going into sort of how those penalties work in coming outside certain I'm certainties you had a certain rash cold and then how you end up getting penalized there and in your role setting up sort of the rules and regulations for. Like you mentioned a 40 year project in which you might personally not be.

A part of by then, could you explain sort of maybe the peer review or validation process? You undergo to ensure the in scientific integrity of your applications for your client. Just to Of correct. Something that you said i don't set up the rules, i obey the rules and look at it set up by the cdn by the clean development.

And even by the byvera, by the rest of, it could be Uh, the American standards, or the, the car, the climate action reserve. I've been working a lot this data, and we see us, So it's they set up the methodology and as a project developer i'm close, we are responsible to get that project.

Out of the whatever is happening on the ground already. My row is okay. Translate to vera and put it into the project design, right? That document, that outlines all the details of the project. That's so that's kind of my room my role. And when i say, well, the procedures, i think my role is also to outline the procedures of how we are gonna.

For example, how are we going to monitor? How are we going to do this inventory? Or how we're gonna, how to calculations are going to be performed, so, which equations. The following the methodology that was already outlined by their, Um, the mineralization so we have The finalization comes with the higher uncertainty when i'm assuming that average above ground biolence.

So we have one way that we can calculate that that equation is established by there. And actually not by there. It's Statistics that basically, what makes? Um, Uh, But and then, but what happens first when you have a project, REDACTED comes in my team comes in. We're helping you to post that project and translate what you're doing on the ground to put that on a dockman.

And show that to vera so that we can start. Indicating the verified carbon units, right? So we are reviewed and the first level of review is called the auditors or the dvd. Uh, so that's independent. Company third-time company that is listed. On the vera website. So, There is some conditions, for those, someone to become an auditor.

But they come to the site. And they. The evaluate a proportion of the monitoring law. That's And they calculate. What they call it. It's The percentage differences. So whatever i presented to them, as the result of my inventory, they come back and be very beside that. So they we measure a percentage of plots.

And get some sort of error out of that. So that's what you show. That's what we got. Those errors could be associated and basically, what they want to try to do is to catch that Welcome to six. That's systematic errors, right? So the errors there, come from It could be a number that you put on long, the way that you're doing the measures, Say wrong.

So the verification wants to catch that. Uh, but for example, we got one project that the way we were doing, the inventory was criticized, in the sense of Uh, we should be doing an asset therapist approach. Instead of doing just the simple five because of the high variability. So we have big individuals.

And we haven't natural regeneration happening. Yeah, by using an ass to the privilege. And those areas, we extrapolate, That biomass to the area so you have smaller interviews being measured in smaller areas but big individuals being imagined bigger use Then your standard deviation is it's pretty low. Yeah, because of the way that you assembly, i mean, that way the uncertainty is going to be creative.

The uncertainty is higher than 10 percent. We're getting. Be getting realized. Okay, 10 or 15 percent of the, the credits are going to be discounted. Uh, into that. Again, that's without using any sort of remote sensing. That's that's how it is. But i feel like, even if we use remote sensing, the otters would need to go back then.

And verify the work that was done. So whatever we do with remote sensing from now on at least with the youths. In the carbon line it needs to be very they need is to be transparent. It needs to do with the business. Because the the audience comes as for us with peer reviews, when you're trying to publish, But they're gonna question.

Yeah, and they're going to try to repeat your work. So having shareable codes. So that's been using, you know, google earth engine or something so arctic could be an issue. Because it's a paid software. So not all the orders. Have that allow license. So you need to be using either kgs or something or just google or pro everything you can.

Now it's easier for them to verify right? The maps. Um, But exactly where my less. How the peer you happen? Yeah. No that's very good. I'm actually interested in In that one approach in which he would critique the name, they offered up a new methodology. Is that now Your standard moving forward or is, you know, when a new or different methodologies proposed to reduce uncertainties, that's something you adopt going forward, or is it project-based?

Yeah. So Each project you realize that each project's quite unique? Because, It depends on many factors. It could be that it's starting from the zero. We had a project this year. That people have been doing Google Forestation since 2013. So the implant ventures the NGOs that are working on the ground.

They highly experience. But we and then we when you go and you decide to certify that to turn, Whatever that they're doing. To a carbon project just different. The fact that they're doing with more stationery Now we got to translate that into we're going to be done a carbon project under the on the vcs Uh, certification.

Um so we can do what they call the allow vcs and most of those magistry allow what they call the methodology deviation. So if they propose something but you've done something different, As long as you. Justify and show the logic. And then it's ability to convince, right? So if you got your references in everything, good references, a good large day.

So, Just like in academia, how lensing, the And when you're defending your or even like, for your comprehensive things, and so yeah. Um yeah the ability to convince and show what you did and why did you do? Now, did that? To that critique in which method was recommended, where you given the chance to Sort of redo your estimates before getting penalized using a new recommendation, or where you given the chance to defend the method in which you chose them.

Yeah, so for this project will continue for No, we couldn't use the estimates bone that came from that monitoring inventory, because the monitoring inventory gave up an air. Verification that was race. A period that we couldn't explain and that was because it was not using permanent plots. Okay, an asset approach.

So And again when when REDACTED in my team, came to the project, all those things were already done. So we were, we could Are we we try to explain to the other, but it's one thing is related to the other. If you cannot, if you're monitoring inventory, cannot be verified and that will be the same with if we using remote senses.

Yeah. If it's not very fileable it's like if they cannot reproduce what they did and come up with the same results. And for sure, when we're talking about forest, Trees grow. In the tropics tree grow trees, grow really fast. So six months, there's going to be a difference, they do expect it, but that difference cannot be higher than 10%.

But, And you can justify. Okay. If before was standing now, high is 12 which would be good but they cannot be cannot be six. I cannot be 20. So even if we're dealing with students or whatever tool that we're using, It got to be very gotta be reproducible.

Great. So i guess There's one question i have here that you might have already explained but i'm going to ask it just in case. Are there instances where you had to make trade-offs between data accuracy and other project requirements?

Okay, i think i'm not gonna answer exactly what you're saying but i can just find something that happened and i see that it could be.

It could restrain the use of, for example, we need it, i told you that for one of our products, we need a high quality image to prove that no trees were removed. No forest was came back down. In the last two years. The image that we're available for free.

That we could have retained from google engine. Yeah. I have resolution was awful. And the auditor was not on convinced. We put an even distinguish. What wasn't true, like the boundaries.

Try to get a better resin, shouldn't age and enter the price of it came, as a I'm gonna call it as an issue, but it wasn't necessarily an issue. Oh yeah. But made for many projects without purpose or i'm seeing from my field, right? It can be. Because then, The threshold is how much of your life, the scientific approach.

How much are we going to, you know, specialize tools and being very Um, Very scientific. For those projects. They're going to become very expensive.

To get out to, to get certified. And then in the end, That their price is going to be pretty hot. And then my questions i is it going to be sustainable in the long run. I would get able to sell the credits that come from of forestry projects compared to the To the plastic projects or even from the technology.

The cost type of projects that you just switch or some sort of material and that's much easier to To catch it. So i see, for example, the price of those image or, for example, the user rgs like the license of the software. Well, these things could come as a As a As a struggle and yeah there could influence in the end in the accuracy of the data in the in the actress of how you estimated this.

Even though that's a that i know that's what their This tracking for this is how you're going to build trust. Yeah. What we doing? This is not cryptocurrency. Even though it sounds like, because we cannot see it, Yeah. Right. But Yeah, i see. We are constantly debating that how much of this, how much science we should be using because it's not a lab, you're not associated to university, be it?

On those projects and how much of the accuracy You're losing and just using, i'm going to call a rudimentary, you know, like simple things. Yeah. Simple to yes, free tools. Free image.

There's, This brings up a really new question in my mind. And this comes that You know, you're hitting on a really important topic here of sort of trade-offs between, you know, cutting at the science versus what's tried to tested. And maybe a just certain point becomes a bit easier to implement.

Um, And i'm wondering if with your clients that Um, you go through this entire process and everything. How many of them are mandated to sort of have a science-based solution? And do you believe in your association with these projects that they understand that this might not be the cutting edge of science that this might not be.

Um, a method or an approach that You know, when When it's reported on in bloomberg or New York Times and it says you know, science has estimated this and that Um that we are using simpler cheaper more affordable, open source method versus maybe the cutting edge of science which you alluded to might be expensive.

All each project is what you need, right? And it depends on. Again, it comes to show to show so that we can have evidence. I'll come back to that example because it just happened. That was something that we were discussing last week, right? And then the price of the image came as a Yeah.

As okay. Let's push on the brakes now here. Um, i'm not talking about problems about thousands of dollars, right talking to indian 500 of dollars, maybe. But that comes as an element surprise, it's like, Oh, i did not see that. Is this really Why. Now during the second round of water is of audit of finding that you bringing that up, You know, and Um, and again, you need to limit manage the clients expectations.

Sometimes you're dealing with people that are very intelligent too, And they want to know everything that we just and And all the airs and say they want to be a part of that. Yeah. Uh, i think you imagine something that about that reminds me of the I'm going circles, i'm not, i'm Not falling.

A logic here, but let's see. The, the guardian article that came out this year. Which the government criticized in the assets and most of those credits are The ghosts, they did not exist. And the question, the way, the transparency, the how the How the data was disclosed. So yes as project developers as a company.

Lean and make sure in my team. Most of us. We either have PhD or masters. So that's to say, we always choosing to use the scientific approach, so we always questioning each other, like, where's the evidence? Like, what's what's the reference for this? Seem to logic. Seems that makes sense? The results that we get.

We always going to go for like that. Um, A solid. Scientific approach. Of that, once it's global implementable very father. And then it's also the two of them. As i swear, i could not share a shape files. With the auditors. I share KML. Because that's something that you can open any one of those tools.

I cannot share very complex models with them. Are, you know, our Package. Well, because they may not and they may not have the technical expertise. So, That's something. Don't live it up for you guys. And the most interesting to think and solve that, okay? Uh, i know that if we need something more specialized, we realize We had some people for another project doing some sort of like lighter.

So we subcontract those Um, But again, it's that discussion how much Um, Science. And How much money? I will an invest in this. The price. The credits that can.

This has been incredibly informative. I loved spending some time and hearing about. Yeah, some of you are data thing blind with your projects pain points and getting an insight into sort of the process and how it relates back to your own academic training and everything and Um, where it goes from here.

So, And i know this for a minute over and, um, I think covered all the questions i had and i'd love to follow up in case you had anyone else that you'd recommend, i talk to. And i just wanted to thank you again.

B.2.3 Anonymized Interview Transcript: Interviewee #2

Okay, so Um, So i've had the privilege like i mentioned before to witness transformative impacts the geospatial technologies. Um, recognized this dynamic. It's a dynamic landscape. A lot of people involved in the curation of applied Uh, geospatial methods, in your words, how would you describe the significance of remote sensing in contemporary geospatial applications?

Um, I mean there is a really big significance of a special in our felt. I cannot really tell for any other fields. Um, but in forester and especially in the forest carbon projects, I can even see the trend of being like, pushed father. And father is just So, we have a lot of, for example, of projects that are located like in the very isolated places and quite often, it's not even possible to get there to measure their background on themas to do like, any other field measurements.

So i feel like js, or remove something allows you to measure it and also reduce your carbon emissions while you're doing that project. So it is highly significant and it's gonna get more and more significantly mistakes. That's interesting. Um, you mentioned sort of isolation of some of these field sites.

How are you able to validate the remote sensing data or the remote sensitive imagery that that is collected for these sites? So quite often what would do is, um, We use lidar data. Okay, which is a little bit more precise and for example using a sentinel or Lancet which is like go max to 30 meter resolution Uh definitely allows you to measure by my small precisely because for example you can Are quite often measure there at the height of the trees, you can measure the crown closure, you can measure the slope and from there you can estimate the biomass.

Um but you definitely need to center this data to something. So, for example, if you collect the lighter data, Um, And then you say, like, okay, this product already has this amount of card one of this amount of biomass, but there's of course, going to be something variation, some of this places going to be less than we're others.

Are going to have, for example, like a A higher. Like species, richness. So for that what we're usually do, you still send the crew? Sometimes unfortunately to isolated places and you put a GPS point. Um, so the plot center is going to be there. Uh, and once you collect the data, at least it knows where the plots are located.

So you can like equally put it or like mop layer two layers on each other. Oh, that's interesting. Um, In, i'm guessing that you're working in landscapes that Can change season overseason, i think prevalence of were then forestry. There's forest fires. How often is the lighter data collected. How often are crews going out there to get that x y or To sort of validate the data.

Do you have any issues with biomass estimation in such? Or issues with fire in general, especially with these isolated areas. Yes, certainly Um, so once the data is collected, it's unfortunate not quite cheap to apply lighter or even like, even if you want to get a better resolution box.

Um, so once you start a project and once you estimate biomass, You kind of stay with that. Because it's going to be a start of the project but then you supposed to remember everything. Um, Every five years. So that's going to be money to work. Yes. Um, you're supposed to remember it.

And then there is going to be some buffer that would let you like. For example, the biomass is still there. Uh, you're gonna get like more credits. In a future if biomass unfortunately, there was affected by fire or by natural disturbances or even by illegal logging, Uh that those credits, those buffer credits are going to be taken from your project so you kind of lose them.

Um, how often it happens? Unfortunately you see it quite a lot of fun. There is definitely going to be a fire there. There's definitely going to be um natural disturbances depending for example They were a couple of projects that we worked on like even not our projects but we saw that they were acting like hurricane, So if that happens, they have to speed up the money touring.

Like reward of the monitoring red measurements like once it happens and once they know the ear, it was affected that they will measure everything from the start. They will send a team evaluate the The fact of it, they will get the new data and they will, they will have to report it.

An ideal word. Nice interesting. You mentioned this timing of every five years, you go back very luncher. Is that a standard is the client codes. Standard operating, where is that five years coming from? It's coming from the registry and from them. Yeah, that you decided to use. But most of them methodologies, still use like five years but like um, period.

Okay. Great. And then Bit of a personal curiosity here. When he said, Things like illegal logging could affect. Um, client's ability to to gain or lose credits. Um do these credits get insured for things like that? Or are there protections against things like illegal operations? Oh yes, certainly. Um So, for example, if there is a project that works with So there are two different projects, for example, of forestry, if we're talking only about like, Their frustration, deforestation, there's avoided plan, and there is avoided unplanted for a station.

So quite all find these days people prefer not to work with avoided unplanned, the first station just because It's really hard to justify it. It's really hard to measure it and It's been getting more and more criticism these days. So what we prefer and what many other companies prefer to work instead is with avoided plant deforestation and that's where you, for example when The.

The. The land owner, for example, he would have a Illegal paper saying that if it wasn't for the project i have applied, for example, for like a pyramid to deforest or my neighbor has this forced the area and applying to do something like that, something similar. So you kind of like measure what would happen And to protect, you still need to protect that area because quite often like even if it's a private property even if he's doing everything he can or she is doing everything, um they can To protect the error, they're still going to be some of the legal login.

So what they need to do as a part of the product, they need to hire Rangers that would go around the theater once in a while. And look for the signs of deforestation or look of the sign for the people being present there.

So, it sounds like there's Multiple aspects to one of these credit projects. And if the client is, Tasks with hiring rangers. Check on everything i'm wondering REDACTED doing the measurements. They have the client client has other obligations in responsibilities. How involved is clamco in the entire process? Or is it just on sort of the Measurement.

Um, Side of things. All either came. It depends on a project. Um, But if, for example, Was started from the scratch. Then we're gonna be involved in everything that includes like project setup doing a literally literature research on. What has happened around the land. I do, for example mapping i check how the land looked 10 years ago.

I'm checking um The historic and land use. So you're trying to estimate like what was, for example, the main agent of deforestation Um, that would be other culture. Um, that could be just timber harvesting, that could be any other potential. Waste, uh, two layout together for i guess to deforest that.

Um, so we would be included in that. We also would be included in calculating carbon Um, quite often, we don't go to the field to measure it ourselves because we try to hire a local actors for that. But we go with them to. Check on the results and we will go there with auditing team to make sure that everything is correct.

And then also includes also like registering the project selling the credits and Although other. Okay. Awesome. Thanks for that overview. That's super interesting. If Sounds like there's A lot of work, especially when client goes brought in in the beginning, it seems like there's a lot of work to look at sort of social contexts of sites as well.

Um, How does any of of those i guess social impacts on the land that you might come come across in your lit review or your Uh, Sort of site history. How then do those play into the more quantitative measurements and registries and standards that you do with GIS implements and how do you account for for historical content?

So i could just lay clefrite again. Yeah, so you mentioned the Um, you might do a lit review on the history of the site. Uh, you might hire local actors to go in and kind of look at Aside to see what has been done. What's currently happening, what are some of the land uses around there?

So there's Some social aspect. There's some human involvement on the landscapes, and i'm wondering how how do you measure those? And how do they play into some of you or more credit specific or forestry application? When? When registering the project and and Your screen credits. Yeah. Um, so you mean like the people that we hire or you mean like, uh, how do we?

Like, i guess. Yeah, sorry, it's confusing. I guess. When you conduct the lit review, when you look at the history of the site, how to how is that informing. All the subsequent work that comes after that. Okay, so you're trying to like find the relation. Okay, my bad i was like, oh no, no that's that's totally on me.

Yeah. Yeah. No no it also misunderstood it. Um, so yeah, you definitely need to Check the historical land use. It's because it helps you to identify and calculate deforestation rates. So, for example, if i go back, using the sentence telling you, as a go, NAC that like, 20% of the forest was deforested.

And then they Check the server accounting, iris and i do year by years. So, for example, i check to sell 2010. I'll check this out on 11 to sell 2012 and i usually, i believe them at theology, ask you to check it for the 10 years of for 10 years.

You see how much of the land was used and how or how was, how much of the force was lost? How much of the land was used for private purposes etc? Um that way you can calculate deforestation rate and knowing that number you can calculate how much carbon would be preserved.

Because the project is there. Interesting. Okay. So the 10 years is set by the registry and then you use remote sensing to get those measurements and then do a comparison deforestation rate, which then plays into Accrediting. Okay. Paper, interesting. Okay. Um, In what ways do you think the advancements in geospatial technology have influenced the expectations of your clients or stakeholders?

I definitely improved that. Um, in a way that clients are now, like are quite familiar with the data set available there. So quite often they don't even want to work was a 30 meter resolution which is like a while ago was considered to be like good. Now like you hear project having like for example nastly it is having a project if force carbon project right now.

I believe in indonesia. Maybe malaysia somewhere in a celsius to Asia and they using 30 centimeters resolution bombs. So, like the bar is very high. Yeah. So quiet, definitely ask you to provide you better moms and then the auditing team also comes in like they were cases with auditing team questioned our use of landset which is quite understand because with the data available, Like these days.

A lot that is really bad. That's so interesting. I've been One of my own projects, we had access to 10 centimeter data. I'm working on forestry applications, and over a really large. Space. Client thought, this is amazing. We're going to be able to do some lunch and Coming from remote sensing perspective.

I'm thinking this is too fine of a resolution for our project. We really don't need this fine resolution if anything. It's a bit cumbersome. Do you have any sort of Conflicts with clients or stakeholders in or auditors, in trying to justify your methods and being like, no, in this case, landsat is good for this application.

And how do you go about that? Yeah, um, yeah. One the audit like Clients. Usually don't question it as long as like okay and like that would be happy to have good resolution maps and i think they Often wanted like for the personal use because then they can look at put a picture of the project.

And like it's very fine resolution. You can see every tree, they can see like what's happening there where the 30 meters resolution. Looks like like something blurry, right? But auditing team quite often questioned that and There were times where we used, we used landsats for land classification. Uh an auditing team question.

How we were able to like get the training data with such a bad quality mouse? Oh interesting. And which is a like a fair point. And how we justified it is we use the arcGIS default map as a reference Um, because like, for example, I believe it's i don't remember what is default map of RJS.

It is definitely 50 centimeters resolution but i don't remember the The region of it, but anyways, Quite often you can see if it's like, You can check the year of the place surface, for example, something 2014 and the lands are dated some of the 15 and if there is a clearing like, of course, declaring is still would be there.

So like you can still see the like the the color is going to be different, you would definitely try to apply spectral indexes. Um, Like see if there is even identified bear crown more but yeah like auditing auditing team is still pushing. Forward high resolution loves these days. Got.

Interest. That's good to know. But um, Really still helpful. With your historical content, the the expectation of using the archgis default map, you know, keeping things of the same resolution and needing to provide that justification. I think it's interesting. But the client kind of goes. About your recommendations on questioning, like, not questioning them or Or anything.

But i did want to and sort of on the same vein, are there any instances where you had to make trade-offs between data accuracy, or data, resolution and other project requirements?

Um, I feel like we're still go for. Data accuracy over deer, like the higher resolution. Like yes, client clients don't usually like they do one like gain better moms, but of wanting it like a day, you explain them that you were able to Like to estimate the force that era that would be fine with that.

The my like fight over a little bit like, oh, if you use the better resolution months, there's going to be more forest. Oh it is an on really true? Yeah of yourself. Like sometimes like even higher resolution like it just creates to my noise. Yeah. And, Which is. Yeah.

We we still can use Lance and like we're still trying to use not like landsat is probably not our top choice. Like sentinel two is still the one that will work them with but i give the project. Like, for example, if the project started in 2013, then we have no choice but to use a landset.

So yeah, we try to still like go over accuracy, for accuracy, over resolution and of course, once do the like classification, you like you do the confusion matrix, you do accuracy assessments and when they see, like, good results. If if statistically speaking everything is, okay, they're okay with that, okay?

And so on that note, um, one of the things that i found in my own work was meeting to sort of explain those statistics and accuracy results and user and producers, and what these meant, and why they were differing numbers on the most part, in your clients, just stakeholders is that a basis of understanding, or do you have to find or do you find your Your, your team explaining?

Um, Enter to get detailed, technical details to clients all the time. No, no, we don't, we don't go over like into details. I still, i usually give them like a brief overview of what happened. How we did it and show them the confusion matrix. Uh, i try not to go for the user accuracy and like, producer accuracy, because that might be a little bit too confusing.

I just use it overall one. Uh, and then you show them like how many pixels will miss identified just in general, just like show the matrix because i feel like it's easier for them to connect with numbers like when it's in the table. Um, and that's pretty much sad.

And then, of course you show them like Like visually. How correct? Where results? Okay. Yeah. So yeah. You yeah. Um, Also, in a similar vein. You know, you're working with people with different expert. Expertise backgrounds. How does collaborating with professionals from other fields, influence the application of remote sensing in your projects?

There were times where We worked a couple of times with people from ubc, so university, british columbia. Um, because Our company. Like, People from my company, from my team used to work there. So they know people. So you definitely have access to like The most recent updates in the field and like quite often they recommended us on the best way to approach, for example.

Um, like classification again. So yeah, like we tried to stay updated, we try to chat with them. And um in general we try to like go to webinars whenever it's possible. We try to pursue our team members to like finish little courses that as we provides that archet provides just to stay updated.

Okay, okay.

Um, Oh, you touched on this one? So, we've mentioned auditors. Mentioned. But there's this group of people that are asking for justification on methods. Um i guess my question here is, can you explain a little bit more sort of the peer review or validation processes that you undergo to ensure the scientific integrity of your applications?

Yeah, so they're gonna Usually two to three circles with auditing team. Okay, so first thing they go to the field. Um, they randomly select a couple of plots, a couple of permanent forestry, a forest plots. Uh, do your measure everything, including a wolf ground by moss depending like what it was included in the project.

Like see his diversity. Sometimes slope like tree height, siding index, etc. They were measure it. And i for everything is dean, correct. They okay with that even if they're not happy with like some of the things that will measure it and sometimes they like, it gets very like, very many peaking, um, really go very into details.

Um, like for example, the equation could not be correct. Or like they did not use that. Like the most updated like, for example equation to extrapolate the biomass numbers. Or like, the dbh was not measured from like the the How do you say the high that they seem to find the best?

Um, for? So for example, for mangrove Um, projects. Um, there, for example, whatever you measure, dbh of the three, they're going to be some of the like higher and lower like, Like parts of the tree and if you just like rotate around, then you're losing quite a lot. So you need to be very respect to actually like hold it here, like move it there to make sure you get into every priest.

They will read and they will know if you hit them, right? Interesting. And i'm That would be the first round, then they give you time to respond, you can try to find it, but some of the things like the auditing team, there is for you to make the project better.

So usually you don't want to find them if they find something wrong in calculations, you just want to fix it. You like it's something weird to find them on that. Yeah. Um, So, you try to fix it the new standard again, they find something else then Uh and the end of the day, they don't have any comments and they find with that.

They give you a paper saying, like they think you fix everything they found and then you go with this feeder to the registry. Okay. As there ever been any conflict in the sense that they found something that you needed, correct? You follow their their recommendations, you've tried to correct it, but you keep coming to a difference between project and auditor.

Or does it usually? Their recommendation was usually what you needed to do. That recommendation is usually what you need to do, and like, Um, not like when you set up a project again, we have the connection to UBC, you like some of the people, you know, our team of like come with a heavy, like four strip background.

So i'd like to think they know where they're doing. So we usually don't get like, That of a bad comment. If they find something incorrect, we're a measure it, or like, we like, we apply it. But there were times where we took the projects that were non-developed by us.

And, We saw their auditing. The reviews and they were bad, like they were really bad and you could see they were like, trying to fight it over and like sometimes Sometimes they were not even trying to fight it. Sometimes there was like A simple miscommunication or a simple problem with a language barrier.

Um, The auditing team would tell them to do something and they would not tell them to like they would not do it. Then again, they said me that for the second round auditing team again, would tell them to fix it. They would not do and then they said me to the gate and we saw that auditing team at some point.

We were like losing their mind and they would like, put it in the cups. Yeah. And like less so like, yeah.

That's there, that's interesting. It's almost like a peer review, paper process.

Great. I guess my last question here is we have about a little less than five minutes left and we got through the meat of the questions that i wanted to ask you. Can you describe a specific project, where integrating remote sensing posed, unique challenges, and how you navigate it through that?

Uh, sure. Um, There were times where we had to estimate forested area in some of the Places. And Sometimes there's like a game but data set, which was 10 meter resolution, which is, i guess it's okay. Yeah, but we did not only take into the account that the place was very hilly, very sloppy that like the sloperation was insane.

So every time like even with 10 meter resolution, I've ever try and we try to do line classification and train the data. I like all of the sides, all of the like shady sides or all of the trays that were on a side of the hill, they were identified.

As clearing. Oh because yeah, like the the spectral signature for some reason was like, working more for the bigger grounded, it could not see that. It was forest. Um, And i feel like, They're like, how we got over it. In the end of the day, is that? Yeah. Like we applying the slope again into the lung qualification that didn't prove a little bit results.

We collected more data, but the more data if the date like the more data we added, the more variation will we're getting and then of the end like the such as typical results were dropping. So what we did and we actually consolid was auditing team If we could clip those areas out so like you just do parts you clip.

Uh they're like the One area out. You do like classification there? And then you do the shady heels on a different classification. And then you kind of like whatever thing together which is not the most elegant way out. Yeah. But you're a good results and auditing team was fun with that, okay?

Fine. With that fun out. We'll see you. They'll get back to us. You might have to say you said it was fine. Yeah i mean we're past perfect first auditing like circle. When i was fine. No sometimes they don't notice something in the first round and they look i'm back to you with the third one.

How long does that process usually take for each round? About with like three four months. Oh, okay. So then if you're doing it, two to three rounds, you've already gone a year into the project. Okay. Three months, three. Um, The leftovered. All the questions i wanted to ask you did anything pop-up that you wanted to ask me at all.

Oh, Up. Don't think so. Yeah. Not at 25. Any questions you can always, don't you? Yes, definitely. Please do if you think of anything and thank you again so much for your time. I really enjoyed chatting learning a lot. Um, And yeah, if anything pops up at all, shoot me an email if you can.

B.3 Change Traceability Matrix: Los Angeles County Tree Inventory Project

WKID Innovation (Stavros, 2021)

Decision Context				Application Traceability			
Policy, Economics and Sociocultural factors that govern people and the Technology (processes) that drive their interactions				**Process Knowledge Mapping**			
What is the expected impact or change that you want to mitigate?	**What is the decision that needs to be made? Who is responsible for making that decision?**	**What are the driving motivations (policy mandate, $, etc.) for making that decision?**	**What current technology /tools are used to inform those decisions?**	**On what information on those decsions rely and the conditions under which its useful (e.g., latency, accuracy, etc.)?**	**What data exists from which to create the necessary information to inform the decision?**	**What are the current limitations of the existing aproach? What would improve this process?**	**Stakeholders**
1. Enhance tree canopy to achieve all the benefits of trees: reduced heat, improved water quantity & quality, job training & job creation, improved air quality, etc.	Research and Education to provide tree management guidelines and assign interdepartmental responsibilities as often as necessary.	The County Board of Supervisors and the County Chief Executive Office (CEO), accountable to residents, need to address the near-term priority action in the "Our County Sustainability Plan". Motivations include equity, resilience (e.g., tree health), and resource efficiency. Coastal Commission	Collector App, ArcGIS (w/ capability to read Tree Assets data), Davey Treekeeper, Inventory management software, LA County Tree Canopy Map Viewer, UFMP Toolkit, i-Tree, GHG calculations,	Accurate, up-to-date (i.e., annual) tree ownership (private vs. public) by location	LA County Parcel Level GIS Layers	Tree Inventories require site verification which is a huge human resource constraint. Some sites require identification of only one species (e.g. oak trees, joshua trees)	USFS; DPR; TreePeople; The Nature Conservancy

<table>
<tr>
<td rowspan="4">2. Improve tree canopy equity, improve urban forest resilience, streamline tree establishment and maintenance

3. Preservation and promotion of healthy urban tree canopy while ensuring public safety

4. Support biodiversity and endangered species

5. Improve corridors for wildlife movement</td>
<td rowspan="4">Urban Forest Management Planning (UFMP): Government agencies (e.g., the City), land owners, residents, non-profits, and others with the jurisdiction of the land/trees need to regularly determine maintenance of trees (Pruning, removing & replanting trees) AND determine what should be planted, when, and where to maximize urban forest health, succession plantings that are climate-appropriate and pest-resistant species diversity</td>
<td rowspan="4">Public Works Administration

Specifically: OurCounty Goal 2, Strategy 2D, Action 43, the Marina del Rey Land Use Plan, Health and Safety Plans

City Council specific Private Tree Ordinances

Federal Clean Water Act under the Los Angeles Safe Clean Water Program</td>
<td rowspan="4">Certified arborists, Urban foresters, Grants (Minimum Data Collective Attributes-MDCA), Plan it Geo, TreePeople /LMU tree viewer, Google Tree Canopy Lab CalEnviroscreen GIS-Net Google Earth, Google Earth Engine, R, TNC-derived data products, CNDDB</td>
<td>Accurate, up-to-date (i.e., annual) information on tree management</td>
<td></td>
<td rowspan="4">and could facilitate more expeditious review of permits where only one species type needs to be verified

Existing approach is fragmented driving needs for bigger budgets and more people in urban forestry to update tree inventory.

There is a need for a coordinated effort for UFMP across entities responsible to</td>
<td rowspan="4">California Dept. of Forestry and Fire Protection (Urban & Community Forestry Program); LA County Chief Sustainability Office; CALFIRE; DPR; TreePeople; L.A. County Beaches & Harbors; Conservation Corps; UCLA / LA Urban Cooling Collaborative; County of Los Angeles Public Works Road Maintenance Division; The Nature Conservancy</td>
</tr>
<tr>
<td>Accurate, up-to-date (i.e., annual) tree health condition (disease/insect type/soil conditions, etc.) of trees, needed annually or every 3-5 years.</td>
<td>Data fusion of Sentinel 2/1/Landsat/AVIRIS/LiDAR/VHR Airborne</td>
</tr>
<tr>
<td>Potential tree planting sites</td>
<td>Maxent Modeling/Million Tree Plan</td>
</tr>
<tr>
<td>Canopy cover</td>
<td>Lidar + Optical</td>
</tr>
</table>

	Funding allocations for grants, conceptual studies, construction documents, installation and maintenance; decisions made by various stakeholders			Priority neighborhoods for planting and maintenance	Census track data + Land Surface Temperature	identify common datasets and needs to overcome funding limitations through collaborative planning and dedicated funding. Sharing urban forest data from private lands Inventories are too few and far between to assess meaningful change-over-time data Technology that allows the inventories to be done accurately & remotely to	Los Angeles County Public Works; CALFIRE; TreePeople; Los Angeles Conservation Corps
	Staff biologists at the LAC Department of Regional Planning and TNC identify Sensitie Ecologocal Areas of unincorporated LA County to facilitate tree preservation in project design phase	Various biological preservation programs, General Plan/Area Plan policies, and Title 22 regulations implemented by the Department of Regional Planning to protect native trees		Species (of interest) Type and Location	Data fusion of Sentinel 2/1/Landsat/AVIRIS/LiDAR/VHR Airborne		LAC Department of Regional Planning; The Nature Conservancy
	Recommendations for areas for habitat restoration to agencies, policymakers, NGOs, funders, and others with the authority and funds to carry out such projects	Regulatory and public comment/input requirements/expectations		Finer scale resolution of the "California Wildlife Habitat Relationships System" data products for the urban environment			The Nature Conservancy

	Prioritized sites for green infrastructure recommendations to agencies, policymakers, NGOs, funders, and others with the authority and funds to carry out such projects					provide map-driven data at a municipal or jurisdictional level A Taskforce dedicated to securing funds and channeling them into tree planting projects for underserved communities or park poor areas Ease of use in the Field	The Nature Conservancy

B.4 References

Stavros, E. N. (2021). Wicked Problems need WKID Innovation: Innovation as a Process to Develop a Disruptive Technology Product. *Research-Technology Management*, *65*(1), 39-47.