Evaluating Uncrewed Vehicle Systems (UVS) for Regulatory Monitoring of Compensatory Mitigation Sites

Jocelyn Munoz

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2025

Committee:

Jonathan D. Bakker

L. Monika Moskal

Meghan Halabisky

Program Authorized to Offer Degree:

School of Environmental and Forest Sciences

University of Washington

Abstract

Evaluating Uncrewed Vehicle Systems (UVS) for Regulatory Monitoring of Compensatory Mitigation Sites

Jocelyn Munoz

Chair of the Supervisory Committee:

Jonathan D. Bakker

School of Environmental and Forest Sciences

In wetland resource management, drones, also known as uncrewed vehicle systems, offer unique advantages for monitoring and ensuring regulatory compliance at wetland mitigation sites. However, before introducing drone monitoring to evaluate performance standards on these sites, agencies need to understand the fundamental processes of working with drone-derived data, how these data compare to that collected via traditional field methods, and the respective advantages and disadvantages of drone-based monitoring. This study evaluated two critical components of the classification process: classification method and machine learning algorithm. I assessed overall accuracy and woody vegetation class accuracy at six mitigation sites in western Washington. These sites encompassed wetland, buffer, and riparian zones, and ranged in age from 3 to 10 years. Results from four classification trials indicated that object-based classifications consistently outperformed pixel-based approaches. The choice of machine

learning algorithms, whether support vector machines or random trees, had no significant impact on accuracy metrics. I then compared woody cover measurements from drone- and field-derived data and analyzed results at the mitigation zone and sample plot scales. At the zone scale, I represented drone cover as the percentage of woody cover across the entire zone and field cover as the average woody cover of all plots. At the plot scale, I expressed drone cover as the percentage of woody cover within each plot, and field cover as the raw data collected in each plot. Differences in mean woody cover estimates between methods ranged from marginally significant to significant at the zone and plot scales. While correlations were significantly positive at the zone scale, they were weaker and more variable at the plot scale. These findings suggest that drone methods underestimate woody cover, and further refinement is needed to improve agreement between methods at the plot level. This study highlights the potential for drones to help site managers meet monitoring demands, particularly when resources are limited. A key drawback is the upfront time investment required to learn processing and analysis techniques. Overall, these findings contribute to developing guidelines for government agencies to adopt drone-based monitoring using off-the-shelf equipment and user-friendly methods that prioritize practical implementation, affordability, and accessibility.

Table of Contents

List of Figures

List of Tables

Abbreviations

AAT	automatic aerial triangulation
AI	artificial intelligence
BBA	bundle block adjustment
CE	commission error
CPs	checkpoints
CWA	Clean Water Act
DSM	digital surface model
DTM	digital terrain model
EPA	Environmental Protection Agency
EPSG	European Petroleum Survey Group
GB	gigabyte
GCPs	ground control points
GNSS	global navigation satellite system
GPS	global positioning system
LWD	large woody debris
nDSM	normalized digital surface model
NWI	National Wetlands Inventory
OBC	object-based classification
OA	overall accuracy
OE	omission error
PA	producer accuracy
PBC	pixel-based classification
Pix4D	Pix4D Mapper Pro
rDSM	relativized digital surface model
RF	random forest
RGB	red-green-blue
RMSE	root-mean-square error
RT	random tree
SAR	synthetic aperture radar

SfM	structure-from-motion
SVM	support vector machine
TPA	Transportation Partnership Act
UA	user accuracy
USACE	United States Army Corps of Engineers
USFWS	United States Fish and Wildlife Service
UVS	uncrewed vehicle systems
WGS	world geodetic system
WSDOT	Washington State Department of Transportation

Acknowledgments

This thesis was a long, challenging, and fulfilling journey, and it would not have been possible without the support of several important people. First, I sincerely appreciate the WSDOT interns for their dedicated efforts in collecting field data at mitigation sites. Their contributions were valuable, as they provided the foundation for my method comparison study. I also appreciate WSDOT for granting me access to their data and drone equipment, which facilitated this research.

I extend my sincere appreciation to my advisor, Dr. Jonathan Bakker, for his constant support and problem-solving expertise. I came to him with an idea, and he helped me navigate the complexities, helping me bring together all the difficult pieces to make this research possible. He dove into the process with me, and together, we learned a great deal about drones, imagery analysis, and regulatory monitoring. I would like to specifically acknowledge his availability, his help with statistical analysis, and his patience as a teacher throughout this journey.

Pursuing a master's degree while working is never easy and requires a strong support system. My partner, Jonathan Fennell, was my greatest source of encouragement. His dedication to seeing me succeed matched my own, and without his unwavering support—through constant encouragement, revisions, and even assisting with fieldwork—I would not have completed this degree. Thank you for keeping everything running smoothly at home and shouldering so many responsibilities so I could focus on my thesis. To my best friend, Puck, thank you for your emotional support and for your fifteen-year-long friendship. Though you are no longer with me, your presence, love, and impact continue to be felt, and I carry your support with me always.

Chapter 1: Introduction

Remote sensing is an expansive and rapidly advancing field widely utilized for environmental research and natural resource management. It encompasses a range of platforms, including satellites, aircraft, uncrewed vehicle systems (UVS), and imaging sensors. The application of remotely sensed data is evolving from general environmental monitoring to addressing specific and localized management needs. In water resource management, particularly concerning wetlands, remote sensing offers significant benefits, including comprehensive monitoring for regulatory compliance of compensatory mitigation sites that are inaccessible or too large for traditional field methods. These tools improve the efficiency and feasibility of monitoring, providing critical insights into ecosystem conditions. By leveraging remote sensing technologies, water resource managers and biologists can optimize management practices.

1.1 Mitigation and Monitoring

While remote sensing has been effectively used for wetland inventories and mapping (Wu, 2017, p.13; Gage, Cooper, & Lichvar, 2020; Igwe, Salehi, & Mahdianpari, 2023), its application in monitoring mitigation sites has yet to be explored. Wetlands and aquatic resources are integral to maintaining healthy freshwater ecosystems by improving water quality, filtering pollutants, and providing wildlife habitat. This underscores the need for robust regulatory protection for wetland functions and services. Section 404[1] of the Clean Water Act (CWA) addresses this need by establishing regulatory oversight through a permitting process managed by the U.S. Army Corps of Engineers (USACE). Under Section 404, compensatory mitigation is required for any impacts

[1] Code of Federal Regulations, Title 33, Chapter 26, Subchapter 4, Section 1344

to wetlands, streams, riparian habitats, and other aquatic resources (U.S. Environmental Protection Agency [EPA], 2008). For instance, if a construction project impacts an existing wetland or wetland buffer, the project might require the development of a mitigation site with similar ecological functions.

Compensatory mitigation involves implementing a plan to monitor the newly created mitigation site over time, ensuring it develops the desired ecological functions. These sites may incorporate various mitigation strategies, including establishment, enhancement, restoration, rehabilitation, and preservation. Additionally, wetlands are often categorized into distinct communities, including forested, scrub-shrub, and emergent zones. Mitigation sites may also encompass protected non-wetland communities, including riparian zones and stream buffers.

Mitigation plans outline monitoring durations and performance criteria. These plans are prepared individually for each project, which may include one or multiple mitigation sites and zones. Performance criteria, often focused on vegetation cover, are established for all sites and may be tailored to specific vegetation communities (i.e. riparian, upland buffer, forested wetland). Common examples of vegetative performance standards written for forested and scrub-shrub wetlands include (Washington State Department of Transportation [WSDOT], 2017):

In Year 3, cover of native wetland trees and shrubs combined will be at least 20 percent in the combined scrub-shrub and forested communities.

In Year 5, cover of native wetland trees and shrubs combined will be at least 35 percent in the combined scrub-shrub and forested communities.

In Year 7, cover of native wetland trees and shrubs combined will be at least 50 percent in the combined scrub-shrub and forested communities.

In Year 10, cover of native wetland trees and shrubs combined will be at least 70 percent in the combined scrub-shrub and forested communities.

To assess the success or performance of mitigation, USACE mandates monitoring for site development, typically over a ten-year period at each mitigation site. Efficient monitoring methods are essential for determining whether a site can achieve its performance criteria. However, monitoring can be constrained by time and available resources, including the time needed for site visits and specific assessments.

As of 2024, the WSDOT Gray Notebook reports that the department owns 323 compensatory mitigation sites statewide, with 87 sites still within their regulatory monitoring period and 236 sites in long-term stewardship that have already met their permit requirements (Figure 1.1). In 2023, WSDOT monitored 87 mitigation sites covering 918 acres (WSDOT, 2024a). The large number of mitigation sites is largely attributed to the Nickel and Transportation Partnership Act (TPA) funding packages of 2003 and 2005, which introduced tax increases to fund transportation projects across the state over 16 years (WSDOT, 2024b). These funding packages led to more transportation projects and, consequently, greater wetland and aquatic resource impacts and a higher demand for mitigation. These figures only represent WSDOT's monitored sites and do not include sites managed by other government agencies or private entities, suggesting that the overall number of mitigation sites in Washington is likely higher. Therefore, techniques that offer rapid assessments of vegetation cover are highly valuable to resource managers, especially as monitoring demands increase.

Figure 1.1 Trends in Mitigation Sites and Acres Monitored by WSDOT (WSDOT, 2024a).

Currently, vegetation monitoring relies on traditional hands-on field methods, such as line-intercept or quadrats. These techniques can be laborious, time-consuming, and limited in spatial scope. They also pose challenges when site access is restricted due to safety hazards like high water levels, steep terrains, or the presence of unsheltered homelessness. As a result, relying solely on ground-based methods can be impractical for managing numerous mitigation sites of varying sizes and monitoring requirements. Integrating remotely sensed data could enhance both the quality and efficiency of monitoring. By providing a broader spatial context of site conditions and assessing vegetation coverage, remote sensing can support ground-based vegetation surveys. However, to fully leverage these benefits, agencies must first become familiar with remote sensing technologies and understand the trade-offs and limitations associated with remotely sensed data.

1.2 Remote sensing characteristics

A wide range of remote sensing platforms and sensors are used to quantify vegetation metrics for restoration projects (Gomez-Sapiens et al., 2021). However, the rapid expansion of these tools necessitates a thorough understanding of the trade-offs between their spatial, spectral, and temporal resolutions.

Spatial resolution is the size of the ground element represented by a single pixel and can vary from coarse (10m x 10m) to fine (sub-centimeter) (Jones & Vaughan, 2010). The altitude of the imaging sensor directly affects the spatial resolution. A sensor at a higher altitude will have a coarser resolution because it covers a larger area per pixel. Conversely, the same sensor positioned closer to the ground can capture finer details with a higher resolution because each pixel represents a smaller area. Mitigation sites, often varying in size from less than an acre to several hundred acres, benefit from higher spatial resolution as it allows for more accurate monitoring of small-scale features, such as young plantings.

Spectral resolution describes the number and width of individual bands the sensor can discern (Natural Resources Canada, 2019). Each spectral band has a specific range of wavelengths on the spectrum within the visible light and near-infrared region. The narrower the range of wavelengths for a given band, the finer the spectral resolution (Natural Resources Canada, 2019). A single-band black-and-white image will contain limited spectral data from one band, whereas a hyperspectral image will capture more detailed spectral signatures for features on the ground because they contain hundreds to thousands of narrow bands. Typically, multispectral sensors are equipped with 3-10 bands which can be advantageous in discerning objects with spectrally distinct characteristics, such as grass and trees. Three-band imagery, required for standard color imagery,

can differentiate between key vegetation types, making it valuable for monitoring vegetative performance standards, such as woody cover.

Temporal resolution defines how frequently a sensor can revisit and capture data for the same location on the Earth's surface. High temporal resolution allows for more frequent observations, which is essential for monitoring dynamic changes such as vegetation growth, flood progression, and urban expansion (Natural Resources Canada, 2019). Since mitigation sites are monitored for up to 10 years, having repeatable and reliable imagery is essential for consistent and effective monitoring throughout the monitoring period.

1.3 Remote sensing platforms and sensors

Different platforms and sensors offer varying spatial, spectral, and temporal resolution combinations, each suited for different monitoring needs. For example, factors such as the size of the site, target vegetation (groups or species), frequency of monitoring, and budget play a role in selecting the appropriate combination of platform and sensor. Users must align these choices with site-specific monitoring goals and constraints to achieve effective and efficient remote sensing outcomes.

Since 1972, space agencies have launched satellites such as Landsat, Terra and Aqua (MODIS), and Worldview for Earth observation and environmental monitoring. These satellites cover expansive areas and provide long-term global imagery (Ballari et al., 2023). The Landsat series (1-9) carry multispectral sensors that capture 4-11 spectral bands across visible, near infrared, and thermal regions (Wulder et al., 2019). Providing the longest record for Earth observation, applications for Landsat data range vastly, from crop phenology mapping to water use and evapotranspiration quantification. The MODIS sensor on the Terra and Aqua satellites

capture data across 36 bands, from visible to thermal infrared, and are used for large-scale climate studies, vegetation health monitoring (Garcia-Mora et al., 2012), and regional fire assessments (Joyce et al., 2009). The Worldview satellites (1-4) carry high-resolution multispectral sensors that can capture up to 8 bands, including visible, near infrared, and shortwave-infrared regions with higher spatial resolutions than Landsat and MODIS. Common applications for Worldview data include bathymetry, forest management, and urban studies (Zhao et al., 2022). While these satellites vary widely in spatial, spectral, and temporal characteristics, making them suitable for diverse Earth observation applications, they share common limitations, including relatively coarse spatial and temporal resolutions due to the time it takes for them to orbit the Earth and revisit the same area.

Satellite platforms and sensors serve as effective tools for larger studies focused on understanding broader environmental processes at the regional or global level. The remote sensing data they provide tends to be multispectral and offers a wide range of potential uses, but the primary limitation is the scale. The spatial resolution of MODIS (250 - 1000 meters), Landsat (250 - 100 meters), and Worldview (sub-meter to 2 meters) is too coarse for capturing small mitigation sites.

To map smaller vegetation units in mitigation areas, users require a higher spatial resolution. Aircraft and UVS, more commonly referred to as drones, can capture aerial imagery with sub-centimeter spatial resolution using a variety of mounted sensors, including optical (red-green-blue [RGB], multispectral, and hyperspectral), LiDAR, and synthetic aperture radar (SAR). Aircraft surveys generally cover targeted areas and provide greater flexibility in sensor selection compared to satellite platforms, which retain their original sensors once deployed. Researchers commonly use multispectral and hyperspectral sensors for mapping vegetation communities, with

hyperspectral sensors offering greater capabilities to differentiate vegetation at the species level (Du et al., 2021). Aircraft-based surveys frequently capture LiDAR data, which provide elevation and point cloud data for applications such as topography and terrain mapping, canopy height and structure analysis for forest management, and stream and river mapping (Li et al., 2020). Additionally, aircraft-based surveys utilize SAR sensors, a microwave remote sensing technology independent of solar illumination, to monitor soil moisture and measure surface properties in conditions where optical sensors are limited, such as during nighttime or under cloud coverage (Palmann et al., 2008; Tsokas et al., 2022).

Unlike drones, aircraft can carry larger payloads, enabling them to mount larger or multiple sensors simultaneously. This capability allows for the integration of different sensors (optical, LiDAR, and radar) in a single survey. However, a key drawback of aircraft surveys is their higher operational cost compared to drones from required maintenance, fuel, and human-operated flights (Zhang & Zhu, 2023). The expense of a single aerial survey often exceeds the budgets of many natural resource agencies, particularly when repeat flights are necessary for ongoing monitoring.

In contrast, drones are a much more affordable option for obtaining sub-centimeter resolution imagery due to their closer proximity to the ground. Drone surveys are relatively inexpensive, easy to deploy, and allow for the greatest control over temporal resolution. Drone surveys are highly customizable and can utilize the same range of sensors found on satellite and aircraft platforms. However, non-optical sensors, such as LiDAR and SAR, are typically impractical for lower-budget studies because they require larger drones with the capacity to carry heavy sensors. Most commercial drone manufacturers equip their drones with a standard RGB camera, which reduces costs but also limits spectral resolution. Drones can be the best choice for agency-level management needs. In combination with object recognition algorithms and

photogrammetric principles, drones can produce high-quality three-dimensional outputs beneficial for mapping and evaluating performance of vegetation groups over time, such as woody vegetation.

1.4 Drones and photogrammetry

Advances in photogrammetry, specifically the rise in accessibility and usability of structure-from-motion (SfM) algorithms, have driven the demand for drone-derived imagery. SfM aims to reconstruct 3D geometry from a sequence of standard uncalibrated images (Carrivick, 2016). With the emergence of photogrammetric software, such as Pix4D, and computer vision algorithms, users can process drone images to derive high-quality 3D products almost instantaneously (Carrivick, 2016). These advancements have enabled federal and state-level natural resource agencies to achieve SfM with as little information as raw images from drones containing GPS and altitude data.

Several commercial and open-source software packages, such as DroneDeploy, Pix4D, and AgiSoft, are readily available for image processing. Each software employs a different photogrammetric procedure, but the general workflow includes image matching, feature extraction, triangulation, and post-processing. As part of the initial processing, image matching identifies the common points, or key points, that appear in several different photographs, allowing the software to reconstruct the scene (Carrivick, 2016). Through camera model optimization, the system resolves the camera location and orientation (Durgan, 2020). Most SfM software also allows the input of ground control points (GCPs) for georectification (Durgan, 2020). Once the software completes feature matching and determines the camera's relation, it computes 3D points from the position of two or more images through the process of triangulation (Carrivick, 2016).

The most standard drone-derived products that Pix4D generates are orthomosaics, point clouds, and digital surface models (DSMs). Orthomosaics are high-resolution 2D models containing RGB and spectral data, created by combining smaller aerial images into a map. Point clouds refer to the 3D visualization of a scene, consisting of millions of georeferenced points that contain the X, Y, and Z position and color information for each point. A raster DSM models the mapped surface, including vegetation, buildings, and terrain. Together, these outputs support image classification to identify different surface types, such as vegetation, and provide useful information about growth and cover.

1.5 Drone use for vegetation monitoring and mapping

Drone imagery, combined with photogrammetry and image classification, increasingly supports the acquisition of fine-spatial data for ecological monitoring (Gomez-Sapiens et al., 2021). However, the spectral quality of the imagery impacts the results of a classification analysis, the process of assigning a specific land cover category to each pixel in an image based on its features. While off-the-shelf drones that provide 3-band RGB data are cost-effective and appealing to resource agencies, their limited spectral resolution can hinder species-level vegetation classification. Even with high spatial resolution, these classifications often yield low accuracy and unreliable maps because their broad spectral resolution is insufficient to distinguish between similar vegetation classes (Prosek & Simova, 2018). To improve classification accuracy, the percentage of correctly classified pixels out of all the pixels in an image (Gonzalez-Perez et al., 2022), analysts can generalize classifications from the species level to broader categories, such as combining shrubs and trees into one class representing woody vegetation.

Harris (2020) demonstrated this approach by evaluating the restoration success of five coastal wetlands using a drone with a standard optical camera to map and quantify the cover of

dominant vegetation and two land cover groups. Harris reported overall classification accuracies ranging from 75% to 90%, with a mean accuracy of 83%. The restoration site with the least diverse vegetation community achieved the highest classification accuracy, while the restoration site with the most diverse vegetation community achieved the least accurate classification. These findings are typical when using an RGB camera, especially for mapping species that appear spectrally similar, such as herbaceous or woody vegetation. Ultimately, the spectral resolution of the sensor plays a major role in determining the level of detail achievable, whether the focus is on species-specific mapping or broader group classifications.

Another method to address the limitations of low spectral resolution data is to incorporate vertical data. Although generating 3D models, such as DSMs, require additional processing times, studies have demonstrated that integrating drone-derived 3D data enhances mapping performance for both generalized classes and specific wetland species. For instance, Prosek and Simova (2018) reported an increase in overall classification accuracy from 73.3% to 88.2% using a multispectral-SfM fused dataset for shrubland vegetation. Their study achieved high-accuracy classifications for six classes, including four woody plant species and two meadow community types.

Harris (2020) and Prosek and Simova (2018) highlight the challenges of mapping vegetation with low-spectral resolution data and demonstrate that fusing data while shifting from species-level to community-level classifications can significantly enhance mapping accuracy. This approach is particularly relevant for compensatory mitigation sites, where performance standards are defined at the community level. High classification accuracies, especially for woody vegetation, which constitutes the majority of these standards, are required over detailed species differentiation. Given these challenges, it is increasingly important to study and standardize classification methodologies for drone-derived data in the regulatory monitoring of these sites. To

identify the factors that influence classification model accuracy, we must understand the entire process, from capturing individual drone photos and stitching them into an orthomosaic, to assigning land cover classes to pixels in the image. Then, comparing these cover estimates to field data will show if drones can be effectively used in regulatory monitoring. This process is essential for ensuring that these mapping techniques reliably meet regulatory monitoring and reporting requirements, particularly as much of the current drone research is geared toward non-regulatory monitoring applications.

1.6 Research objectives and limitations

In this study, I used drone imagery to identify woody vegetation and quantify woody cover at six mitigation sites in western Washington. These sites included forested and scrub-shrub wetlands, buffers, and riparian zones. The overarching goal of my research was to assess the potential for drones to monitor performance standards at mitigation sites using affordable, off-the-shelf equipment and user-friendly processes and analyses. This study assumed that woody vegetation at the sites consisted predominantly of native species. Additionally, a key limitation of this methodology was its inability to classify vegetation at the species level, which restricted the measurement of species-specific cover.

1.6.1 Research questions

My research aimed to address the following questions:

1. Which components of the classification process influence accuracy: the classification method and/or the choice of machine learning algorithm?
    a. Does the integration of vertical data improve the accuracy metrics of this study: overall accuracy, and user and producer accuracies of the woody vegetation class?

2. How do woody cover estimates from drone imagery compare to field estimates obtained using the line-intercept method, at the zone and plot scales?
    a. Do different classification trials yield woody cover estimates that are more comparable to field estimates?
    b. How do zone and age classes affect woody cover estimates derived from drone imagery?

1.6.2 Chapter summaries

In Chapter 2, I explored the image classification process to identify woody vegetation in drone imagery. Specifically, I tested the performance of two classification methods (pixel-based and object-based) and two machine learning classifiers (support vector machines and random trees). The classification methods were applied to each mitigation site with a target accuracy of at least 85%. Additionally, I examined a fused classification approach at two mitigation sites, integrating spectral (RGB) data with vertical data derived from DSMs (Figure 1.2).

In Chapter 3, I compared woody cover estimates from the most accurate classification trials in Chapter 2 with field measurements at two scales: zone and plot. At the zone scale, I represented drone-derived vegetation cover as the percentage of woody cover across the entire zone and field-derived data as the average woody cover of all sample plots. At the plot scale, I expressed drone-derived vegetation cover as the percentage of woody cover within each plot, and field-derived data as the raw data collected in each plot.

In Chapter 4, I summarized the broader applications of this research, particularly for implementing drone surveying in regulatory monitoring. This chapter highlights ways to improve the study and suggests directions for future research.

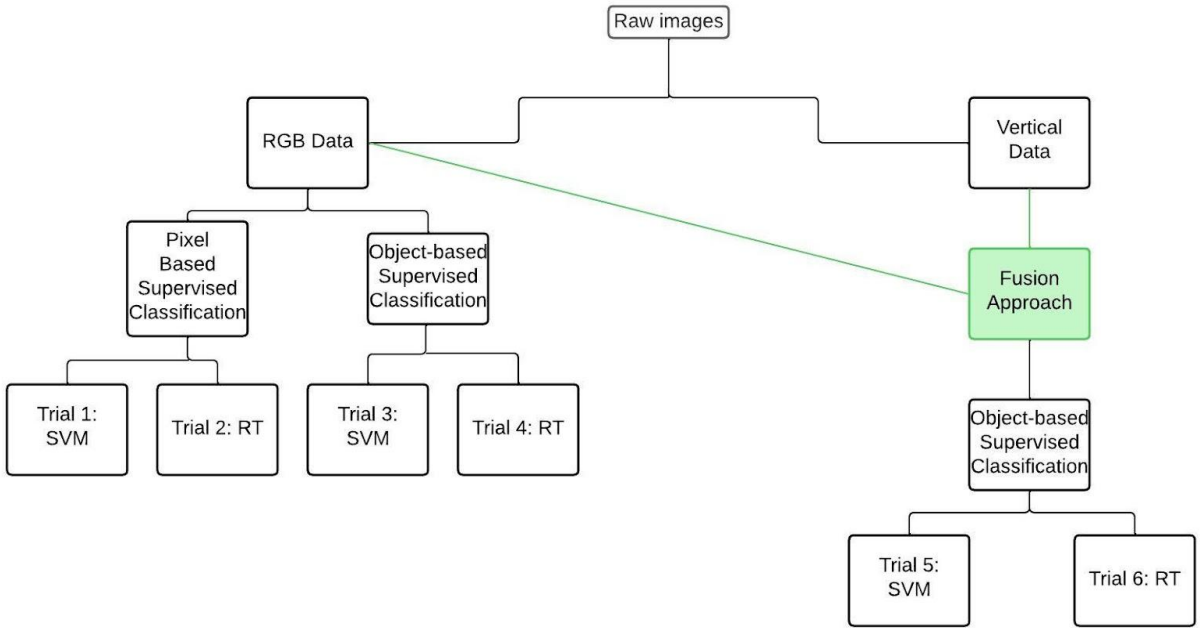

Figure 1.2 Summary of the classification trials. I conducted Trials 1–4 at all sites and Trials 5–6 at two sites.

1.7 Significance of this thesis

Despite the wide emergence of drone technology, drones are rarely used by regulatory agencies as a natural resource monitoring tool. Their narrow use likely stems from uncertainties in methodologies for processing imagery, a lack of standardized guidelines around using drone-derived data to classify vegetation, and uncertainty about how cover estimates from drones compare to those from field-based methods. Before introducing drone monitoring to address regulatory performance standards in compensatory mitigation sites, agencies must understand the fundamental processes of working with drone-derived data.

This research seeks to provide reference for image processing and classification to compare woody cover data between traditional field methods and drone imagery. Resource managers and biologists can apply the results from this study to determine the feasibility of using

drones to support vegetation monitoring under a regulatory scope. By making remote sensing data more accessible and user-friendly, this research aimed to enhance decision-making and improve monitoring efficiency for wetland managers and biologists.

References

Ballari, D., Vilches-Blázquez, L. M., Orellana-Samaniego, M. L., Salgado-Castillo, F., Ochoa Sánchez, A. E., Graw, V., Turini, N., & Bendix, J. (2023). Satellite Earth observation for essential climate variables supporting sustainable development goals: A review on applications. *Remote Sensing, 15*(11), 2716. https://doi.org/10.3390/rs15112716

Carrivick, J. L., Smith, M. W., & Quincey, D. J. (2016). *Structure from motion in the geosciences*. John Wiley & Sons.

Du, B., Mao, D., Wang, Z., Qiu, Z., Yan, H., Feng, K., & Zhang, Z. (2021). Mapping wetland plant communities using unmanned aerial vehicle hyperspectral imagery by comparing object/pixel-based classifications combining multiple machine-learning algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 8249–8258. https://doi.org/10.1109/JSTARS.2021.3100923

Durgan, S. D. (2020). *Evaluating unmanned aircraft system photogrammetry for coastal Florida Everglades restoration and management* (Unpublished master's thesis). Florida Atlantic University.

Durgan, S. D., Zhang, C., & Duecaster, A. (2020). Evaluation and enhancement of unmanned aircraft system photogrammetric data quality for coastal wetlands. *GIScience & Remote Sensing, 57*(7), 865-881. https://doi.org/10.1080/15481603.2020.1819720

Gage, E., Cooper, D. J., & Lichvar, R. (2020). Comparison of USACE three-factor wetland delineations to National Wetland Inventory maps. *Wetlands, 40*, 1097–1105. https://doi.org/10.1007/s13157-019-01234-y

García-Mora, T. J., Mas, J. F., & Hinkley, E. A. (2011). Land cover mapping applications with MODIS: A literature review. *International Journal of Digital Earth, 5*(1), 63–87. https://doi.org/10.1080/17538947.2011.565080

Gómez-Sapiens, M., Schlatter, K. J., Meléndez, Á., Hernández-López, D., Salazar, H., Kendy, E., & Flessa, K. W. (2021). Improving the efficiency and accuracy of evaluating aridland riparian habitat restoration using unmanned aerial vehicles. *Remote Sensing in Ecology and Conservation, 7*(3), 488–503. https://doi.org/10.1002/RSE2.204

Gonzalez-Perez, A., Abd-Elrahman, A., Wilkinson, B., Johnson, D. J., & Carthy, R. R. (2022). Deep and machine learning image classification of coastal wetlands using unpiloted aircraft system multispectral images and lidar datasets. *Remote Sensing, 14*(16), 3937. https://doi.org/10.3390/rs14163937

Harris, J. M. (2020). *Evaluating coastal wetland restoration in Louisiana using drones* (Unpublished master's thesis). University of Louisiana at Lafayette.

Igwe, V., Salehi, B., & Mahdianpari, M. (2023). Rapid large-scale wetland inventory update using multi-source remote sensing. *Remote Sensing, 15*(20), 4960. https://doi.org/10.3390/rs15204960

Jones, H. G., & Vaughan, R. A. (2010). *Remote sensing of vegetation: Principles, techniques, and applications*. Oxford University Press.

Joyce, K. E., Wright, K. C., Samsonov, S. V., & Ambrosia, V. G. (2009). Remote sensing and the disaster management cycle. *Advances in Geoscience and Remote Sensing, 48*(7), 317-346.

Li, X., Liu, C., Wang, Z., Xie, X., Li, D., & Xu, L. (2020). Airborne LiDAR: State-of-the-art of system design, technology, and application. *Measurement Science and Technology, 32*(3), 032002. https://doi.org/10.1088/1361-6501/abc867

Natural Resources Canada. (2019). *Fundamentals of remote sensing: A Canada Centre for remote sensing tutorial*. Natural Resources Canada. https://natural-resources.canada.ca/maps-tools-publications/satellite-elevation-air-photos/fundamentals-remote-sensing-introduction

Palman, C., Mavromatis, S., Hernandez, M., Sequeira, J., & Brisco, B. (2008). Earth observation using radar data: An overview of applications and challenges. *International Journal of Digital Earth, 1*(2), 171–195. https://doi.org/10.1080/17538940802038317

Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., & Wilhelm, V. L. (2017). Object based classification of wetland vegetation using very high-resolution unmanned air system imagery. *European Journal of Remote Sensing, 50*(1), 564-576. https://doi.org/10.1080/22797254.2017.1373602

Prošek, J., & Šímová, P. (2019). UAV for mapping shrubland vegetation: Does fusion of spectral and vertical information derived from a single sensor increase the classification accuracy? *International Journal of Applied Earth Observation and Geoinformation, 75*, 151–162. https://doi.org/10.1016/j.jag.2018.10.005

Tsokas, A., Rysz, M., Pardalos, P. M., & Dipple, K. (2022). SAR data applications in Earth observation: An overview. *Expert Systems with Applications, 205,* 117342. https://doi.org/10.1016/j.eswa.2022.117342

U.S. Environmental Protection Agency. (2008). *Compensatory mitigation factsheet*. U.S. Environmental Protection Agency. Retrieved March 17, 2024, from

https://www.epa.gov/cwa-404/background-about-compensatory-mitigation-requirements-under-cwa-section-404

Washington State Department of Transportation. (2017). *Environmental and wetland mitigation writing standards*. [Unavailable online].

Washington State Department of Transportation. (2024a). *Wetland mitigation acreage*. Washington State Department of Transportation. Retrieved September 15, 2024, from https://wsdot.wa.gov/about/data/gray-notebook/gnbhome/environment/wetlands/default.htm

Washington State Department of Transportation. (2024b). *2003 Nickel funding package*. Retrieved December 19, 2024, from https://wsdot.wa.gov/construction-planning/funding/legislative-funding-packages/2003-nickel-funding-package

Wu, Q. (2018). GIS and remote sensing applications in wetland mapping and monitoring. In B. Huang (Ed.), *Comprehensive geographic information systems* (pp. 140–157). Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.10460-9

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., ...& Zhu, Z. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment, 225*, 127–147. https://doi.org/10.1016/j.rse.2019.02.015

Zhang, Z., & Zhu, L. (2023). A review on unmanned aerial vehicle remote sensing: Platforms, sensors, data processing methods, and applications. *Drones, 7*(6), 398. https://doi.org/10.3390/drones7060398

Zhao, Q., Yu, L., Du, Z., Peng, D., Hao, P., Zhang, Y., & Gong, P. (2022). An overview of the applications of Earth observation satellite data: Impacts and future trends. *Remote Sensing, 14*(8), 1863. https://doi.org/10.3390/rs14081863

Chapter 2: Classification methods and machine learning classifiers for drone imagery

2.1 Introduction

The technological advancements, versatility, and accessibility of uncrewed vehicle systems (UVS), commonly called drones, have significantly enhanced environmental monitoring. To expand drone accessibility for natural resource agencies, it is essential to understand the fundamentals of working with imagery data. Key aspects of imagery analysis, such as image classification and accuracy assessments, differ from traditional ground methods. These differences necessitate relevant and tangible examples of their application. Given the broad scope of image classification, approaches can vary widely from custom applications like PyTorch to out-of-the-box applications such as image classifiers available in ArcGIS Pro. The former provides greater flexibility in tailoring products to specific agency needs while preserving control over the final product. The latter offers usability advantages that align with the goals of this research, which seeks to find achievable ways to integrate drone monitoring with natural resource agencies.

Land classification, a specific type of image classification, is the process of labeling pixels in an image to identify land cover classes. Land classification consists of several successive stages (Figure 2.1): creating a multispectral dataset from raw images, choosing a classification method (either pixel-based or object-based), selecting a machine learning classifier (such as support vector machine or random trees), and delineating training and validation data to train the model to recognize different classes and evaluate its performance. Each step can greatly affect the overall quality and interpretation of the results, requiring in-depth knowledge of the methods and input imagery.

2.1.1 Classification methods (pixel-based and object-based)

The first consideration in image classification is selecting between pixel-based and object-based methods. Pixel-based classification assigns labels to individual pixels based solely on their spectral information. However, this approach can be insufficient for high-resolution imagery, as noise from individual pixels can degrade classification accuracy, leading to a "salt and pepper" effect (Zhou et al., 2021). Since pixel-based methods depend on spectral data, they perform best with imagery that includes spectral bands extending into the near-infrared region (Halabisky, Moskal, & Hall, 2011). Pixel-based classification can also be effective in homogeneous landscapes (Whiteside & Ahmad, 2005) and requires fewer processing steps than object-based classification, making it more user-friendly.

In contrast, object-based classification groups pixels into segments or objects based on spectral data, spatial relationships, and shape (Jones & Vaughan, 2010). Although object-based classification achieves higher accuracy in heterogeneous landscapes (Sibaruddin et al., 2018; Zheng et al., 2022), it involves extra steps, such as segmentation, and relies on a more complex process that integrates both spectral and spatial data.

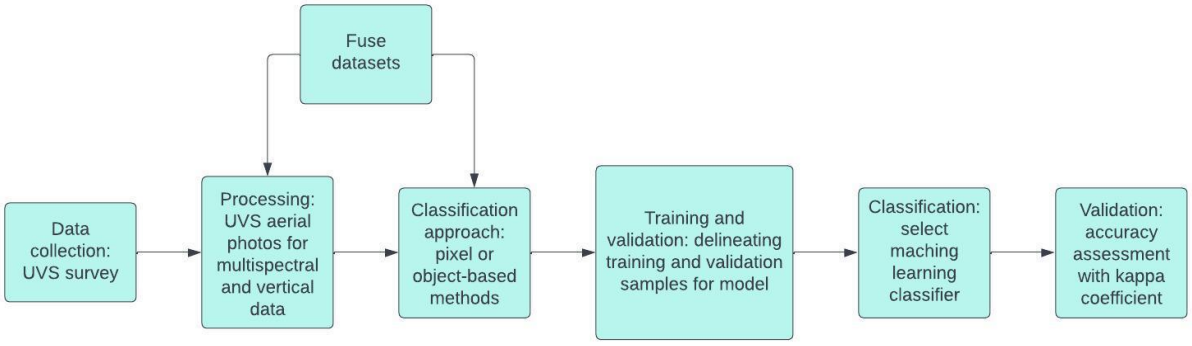

Figure 2.1 Flowchart of land cover classification steps

2.1.1.1 Image classification methods in wetlands

In wetland classifications, most studies have predominantly adopted object-based methods to achieve higher accuracy (Durgan, 2020; Gomez-Sapiens et al., 2021; Pande-Chhetri et al., 2017). Halabisky et al. (2011) highlighted several advantages of object-based image analysis (OBIA) for mapping wetlands in semi-arid regions, including improved classification accuracy, more precise delineation of wetland boundaries, reduced salt-and-pepper effects, and the ability to integrate multiple data attributes such as texture and spatial relationships. Notably, their research acknowledged that pixel-based approaches are less suitable when features such as the shape of an object are essential for classification.

However, Berhane et al. (2018) found that while object-based classification can improve accuracy by incorporating spatial context and multiple feature attributes, pixel-based methods, particularly those using a Random Forest (RF) classifier, also performed well in wetland-dominated landscapes. Their study showed that a pixel-based RF classifier achieved an overall accuracy of 87.9%, compared to 84.6% for an object-based RF approach using the same input layers. When additional spectral and texture features were included, the object-based RF accuracy increased to 90.4%, but this improvement required higher computational resources and greater user input for segmentation tuning. The study concluded that while object-based methods offer advantages in certain contexts, pixel-based methods remain a viable alternative, particularly when there are limitations to computational efficiency and resource availability.

2.1.2 Machine learning classifiers

The second consideration in image classification is selecting an appropriate machine learning classifier. The support vector machine (SVM) algorithm is a machine learning method

that maps training data points into an n-dimensional space, where each feature corresponds to a specific coordinate. The classifier identifies a hyperplane that optimally separates the classes (Bottou & Vapnik, 1992). One of SVM's key advantages is its robustness to the distribution of underlying data, making it more adaptable than other methods. Additionally, SVM is particularly effective in remote sensing applications, as it can achieve high classification accuracy even with small training datasets (Sheykhmousa et al., 2020). Its ability to efficiently handle high-dimensional feature spaces and detect complex patterns makes it especially useful for drone imagery, which often includes multiple spectral bands.

The random trees (RT) classifier, widely used in remote sensing for land cover classification, is valued for its transparent decision-making process and high classification accuracy (ESRI, n.d.d). For clarity, ESRI's RT classifier is equivalent to the Random Forest (RF) algorithm; however, because I used ESRI software in my analysis, I will refer to it as the RT classifier throughout the discussion of my research. The RT classifier consists of multiple decision trees, each generated from different subsets of the training data. For every pixel classified, a series of ranked decisions is made, and the entire image is classified multiple times based on a random sub-selection of training pixels. The algorithm selects the final classification using the tree output that appears most frequently. RT is robust to outliers, can handle large numbers of input variables, and offers high computational efficiency, though it requires greater processing capacity compared to SVM (Sheykhmousa et al., 2020).

2.1.2.1 Machine learning classifiers for wetland imagery analysis

Image classification in wetlands present unique challenges due to the spatial and spectral complexity of these ecosystems, which include diverse vegetation types and heterogeneous

landscapes. SVMs are widely used in remote sensing due to their ability to draw clear decision boundaries, even in complex environments, which improves classification accuracy (Sheykhmousa et al., 2020). Wetland vegetation can have clear distinctions between different types of plants based on spectral reflectance patterns, such as aquatic plants, emergent vegetation, and forested wetlands. SVMs can classify vegetation types by distinguishing between spectral signatures of different plant species or vegetation communities.

Decision tree classifiers effectively classify wetlands by integrating multiple spatial variables and handling complex, non-linear relationships. For instance, Halabisky et al. (2023) applied a decision tree model in their Wetland Intrinsic Potential (WIP) tool, which utilized multi-scale remote sensing proxies of wetland indicators to enhance wetland identification. Their model, incorporating vegetation, hydrology, and soil variables, achieved 91.97% accuracy. This study highlights the effectiveness of decision tree models in accounting for spatial complexity in wetland mapping, particularly for detecting cryptic wetlands that may not be easily identifiable in conventional imagery. Similarly, wetlands with a mix of spectrally overlapping land cover types can be challenging to classify, but decision tree algorithms, like RT, are well-suited for mapping wetland vegetation due to their ability to handle large and complex datasets.

2.1.3 Fusing DSMs with spectral data

Researchers are increasingly exploring the fusion of vertical data, such as digital surface models (DSMs), with image classification to enhance the accuracy of wetland vegetation mapping (Durgan, 2020; Gomez-Sapiens et al., 2021). Vertical data can help differentiate between vegetation classes based on height, effectively distinguishing herbaceous plants from trees and shrubs. However, unlike lidar, photogrammetric techniques cannot reliably reconstruct digital

terrain models (DTMs) below the vegetation canopy, because imagery cannot penetrate the canopy. This limitation can be problematic, particularly because vertically fused data is typically normalized, especially in areas with variable terrain. A normalized digital surface model (nDSM) is derived by subtracting a DTM from a DSM, representing the relative height of features above the ground surface. Normalization allows for height comparisons between features, whereas DSMs and DTMs represent absolute height, and cannot be reliably used for comparisons between features (Husson, Reese, & Ecke, 2017).

2.1.4 Objectives, questions, and hypotheses

In their review, Dronova et al. (2021) analyzed 122 case studies from 29 countries to assess the current status and emerging opportunities in the use of drones for wetland applications. They found that drones are increasingly employed for various wetland monitoring and management objectives, including vegetation mapping, hydrological assessments, and wildlife monitoring. However, despite the widespread use of drones for ecological assessments, their application for *regulatory monitoring* remains largely unexplored. Given that image classification has yet to be extensively studied for regulatory monitoring, this work represents an important first step toward developing guidelines for drone monitoring.

This chapter examined two primary components of the classification process: the classification method and the machine learning algorithm. I conducted four classification trials, comparing overall accuracy and accuracy for woody classes across each trial. My study expands on previous research that focused on individual components within a single study area by investigating these components together across various sites. I also explored the usability of drone-derived DSMs for vertical data fusion, especially when lidar cannot reliably produce terrain

models. By relativizing DSM values, I assessed whether this approach improves classification accuracy in two study sites.

This chapter addressed the following questions:

1. Which principal components of the classification process influence accuracy: the classification method or the choice of machine learning classifier?
2. Does the integration of vertical data improve classification accuracy, specifically, overall accuracy, and user and producer accuracy for woody vegetation?

I hypothesized that object-based classification would outperform pixel-based methods because it accounts for image segments, and that the SVM classifier would outperform the RT classifier by clearly distinguishing boundaries between woody vegetation and non-target classes when those differences exist. Additionally, I hypothesized that integrating vertical data would enhance overall classification accuracy at two mitigation sites by further distinguishing woody vegetation from non-target classes such as herbaceous vegetation.

This chapter explores the practical application of image analysis using ArcGIS Pro to streamline workflows and make the process more accessible for natural resource management. For classification products, I established an acceptable accuracy threshold of 85%, covering overall accuracy as well as user and producer accuracies for the woody vegetation class. Although no formal guidelines define an accuracy standard for regulatory monitoring, the U.S. Geological Survey (USGS) and the Intergovernmental Panel on Climate Change (IPCC) recognize an 85% accuracy level as an acceptable benchmark for land cover classifications (Anderson et al., 1976; USGS, 2023; IPCC, 2019).

2.2 Methods

2.2.1 Study sites

I conducted this study at six mitigation sites managed by WSDOT in western Washington (Figure 2.2). I selected these sites based on three major criteria: vegetation zones; age; and performance standards. The sites consist of wetlands, buffers, and riparian zones. In 2023, three sites were in year 3, while the other sites ranged in age from year 7 to year 10. For all sites, the performance standards focused on the establishment and growth of native woody vegetation.

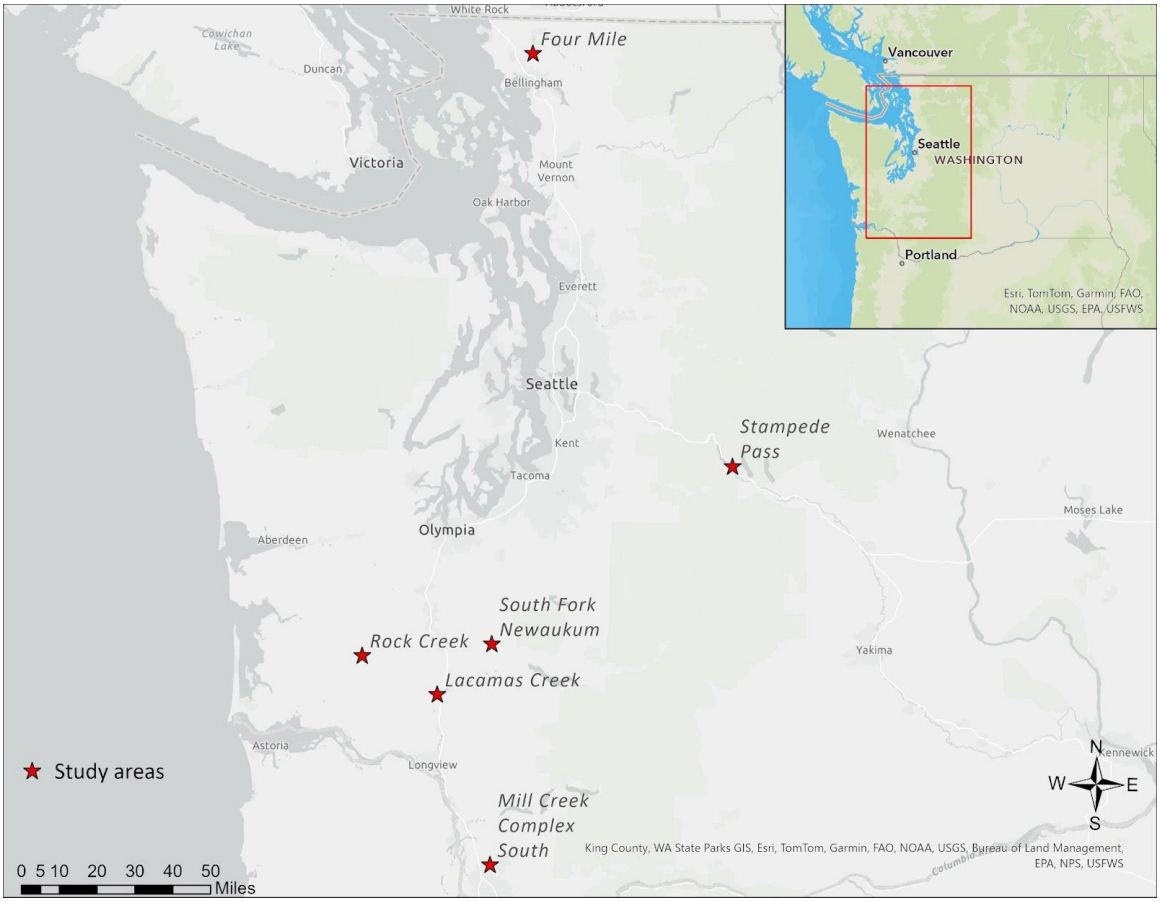

Figure 2.2 Location of the study sites in western Washington.

2.2.1.1 Four Mile Creek

Four Mile Creek (Figure 2.3) is a 3.61-acre compensatory mitigation area located in Whatcom County, near the intersection of SR 539 and Ten Mile Road (WSDOT, Northwest Region, 2020). The site includes 0.21 acres of created scrub-shrub wetland and 1.38 acres of rehabilitated scrub-shrub wetland along the banks of Four Mile Creek. It also includes 2.02 acres of buffer. The wetland community primarily consists of twinberry honeysuckle (*Lonicera involucrata*), rose spiraea (*Spiraea douglasii*), and various willows (*Salix* spp.). The buffer community consists of a mix of upland and wetland woody vegetation, such as twinberry honeysuckle, willows, snowberry (*Symphoricarpos albus*), red elderberry (*Sambucus racemosa*), paper birch (*Betula papyrifera*), and Sitka spruce (*Picea sitchensis*).

2.2.1.2 South Fork Newaukum

South Fork Newaukum (Figure 2.4) is a 4-acre compensatory mitigation site located five miles east of Onalaska, Washington (WSDOT, Southwest Region, 2016). The site features 3.72 acres of enhanced riparian buffer adjacent to South Fork Newaukum River in Lewis County. It is dominated by snowberry, Oregon grape (*Mahonia aquifolium*), and oceanspray (*Holodiscus discolor*). Other common woody plants include red elderberry, osoberry (*Oemleria cerasiformis*), and salmonberry (*Rubus spectabilis*). The remaining 0.28 acre was planted as a wetland buffer under the bridge, inaccessible to the drone and therefore was excluded from the study.

2.2.1.3 Rock Creek

Rock Creek (Figure 2.5) is a 19.98-acre compensatory mitigation site located three miles west of Pe Ell, Washington (WSDOT, Southwest Region, 2013). Rock Creek contains 2.31 acres

of created forested and scrub-shrub wetland and 3.40 acres of upland buffer adjacent to a flashy creek system and Willapa Hills Trail. The remaining preservation acreage was not part of this study, as it did not contain regulatory performance standards. The primary woody vegetation in the wetland consists of young Oregon ash (*Fraxinus latifolia*), and various willows. Herbaceous vegetation in the wetland includes slough sedge (*Carex obnupta*), tufted hairgrass (*Deschampsia cespitosa*), common rush (*Juncus effusus*), and small-fruited bulrush (*Scirpus microcarpus*). In the buffer, the dominant woody vegetation consists of snowberry, Douglas fir (*Pseudotsuga menziesii*), and roses (*Rosa* spp.).

2.2.1.4 Lacamas Creek

Lacamas Creek (Figure 2.6) is a 5.1-acre compensatory mitigation site located two miles east of Vader, Washington (WSDOT, Southwest Region, 2021). Lacamas Creek features an enhanced riparian buffer dominated by snowberry, Oregon grape, osoberry, bigleaf maple (*Acer macrophyllum*), and Western red cedar (*Thuja plicata*). The mitigation site is situated between upland pasture to the east and Lacamas Creek to the west.

2.2.1.5 Stampede Pass

Stampede Pass (Figure 2.7) is a 1.92-acre compensatory mitigation site located within the Okanogan-Wenatchee National Forest near Swamp Creek (WSDOT, South Central Region, 2016). The site contains 1.58 acres of scrub-shrub wetland re-establishment, 0.34 acres of buffer restoration, and three ponds. The wetland is primarily populated by willows and Sitka alder (*Alnus viridis*), while the buffer includes Douglas fir and red alder (*Alnus rubra*), along with smaller

shrubs such as snowberry and roses. A portion of the buffer has developed into a non-native field dominated by ragworts (*Senecio* spp.).

2.2.1.6 Mill Creek Complex South

Mill Creek Complex South (Figure 2.8) is a 5-acre compensatory mitigation site adjacent to Mill Creek in Battle Ground, Washington (WSDOT, Southwest Region, 2012). The site provides 1.63 acres of created forested and scrub-shrub wetland along with 0.11 acres of wetland enhancement. It also includes 3.26 acres of oak woodland and buffer. The wetland community is composed of Oregon ash, Pacific ninebark (*Physocarpus capitatus*), Pacific crabapple (*Malus fusca*), cluster rose (*Rosa pisocarpa*), and red osier dogwood (*Cornus sericea*). The forested and scrub-shrub wetland also supports a robust native emergent community, primarily consisting of slough sedge. The oak woodland preservation and upland buffer is characterized by mature Oregon white oak (*Quercus garryana*) in the canopy, with snowberry, osoberry, and oceanspray in the understory.

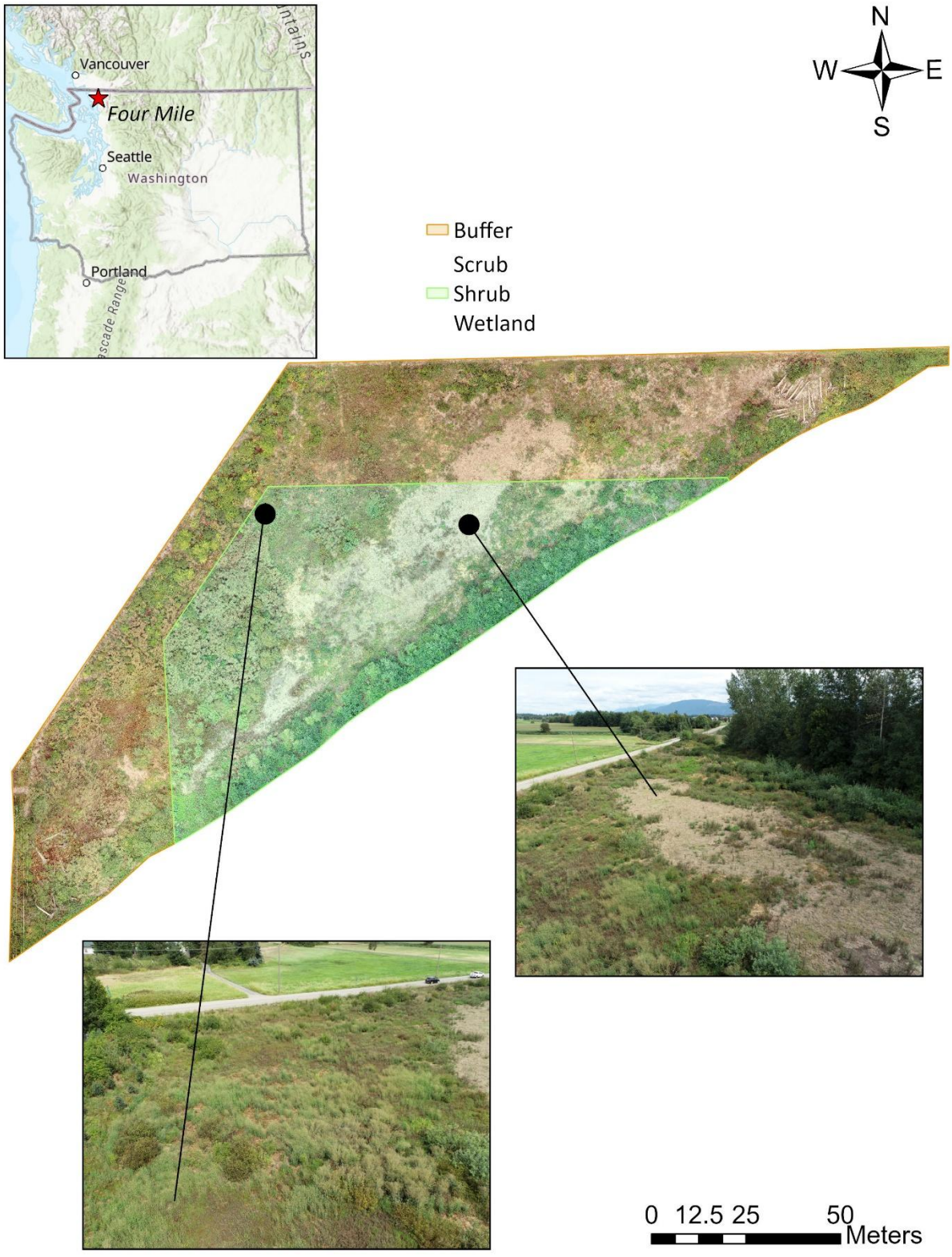

Figure 2.3 Four Mile study area, showing buffer and scrub shrub wetland vegetation zones. WSDOT captured drone images for this site on August 23, 2023.

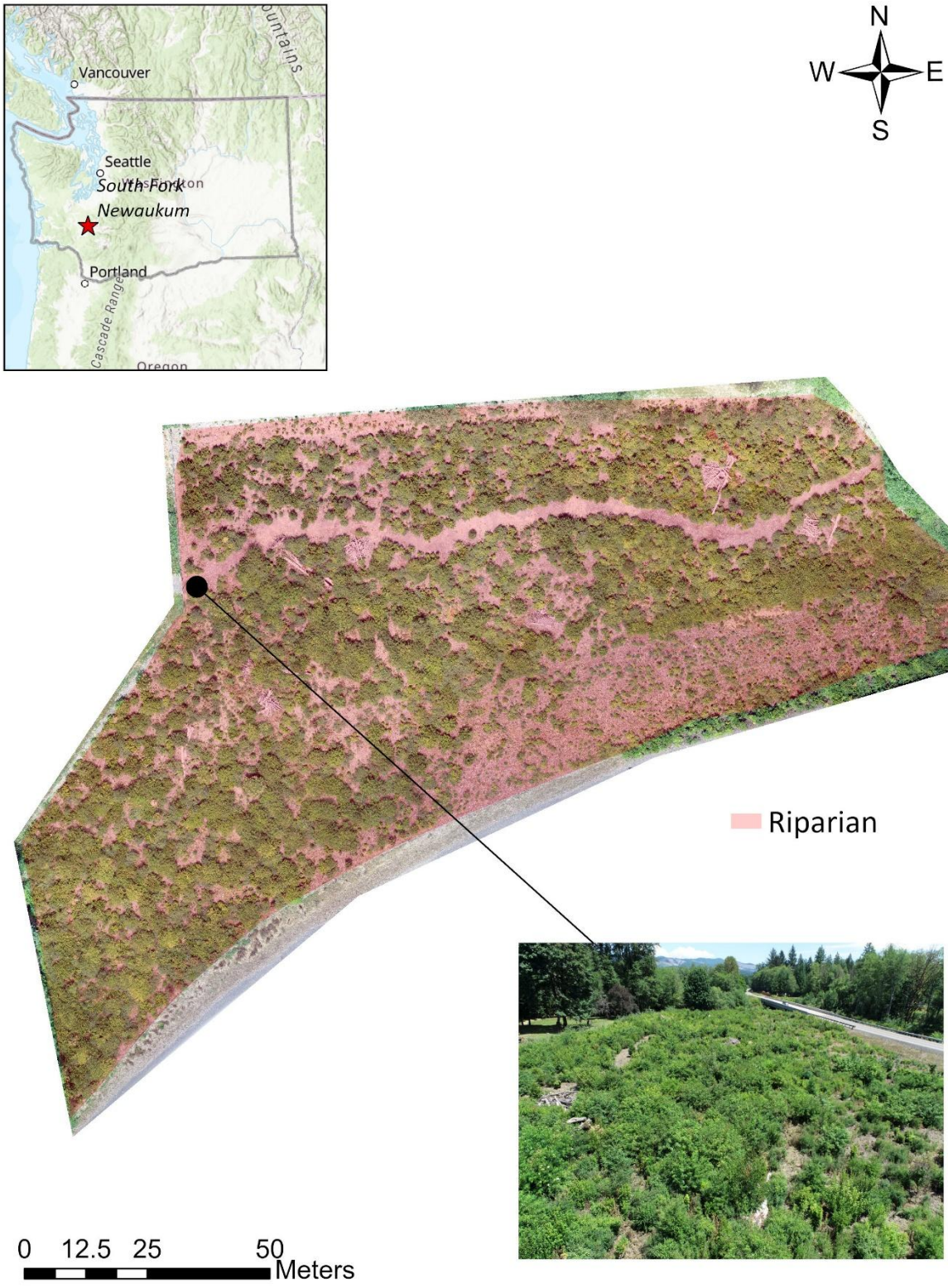

Figure 2.4 South Fork Newaukum study site, showing riparian vegetation zone. WSDOT captured drone images for this site on July 12, 2023.

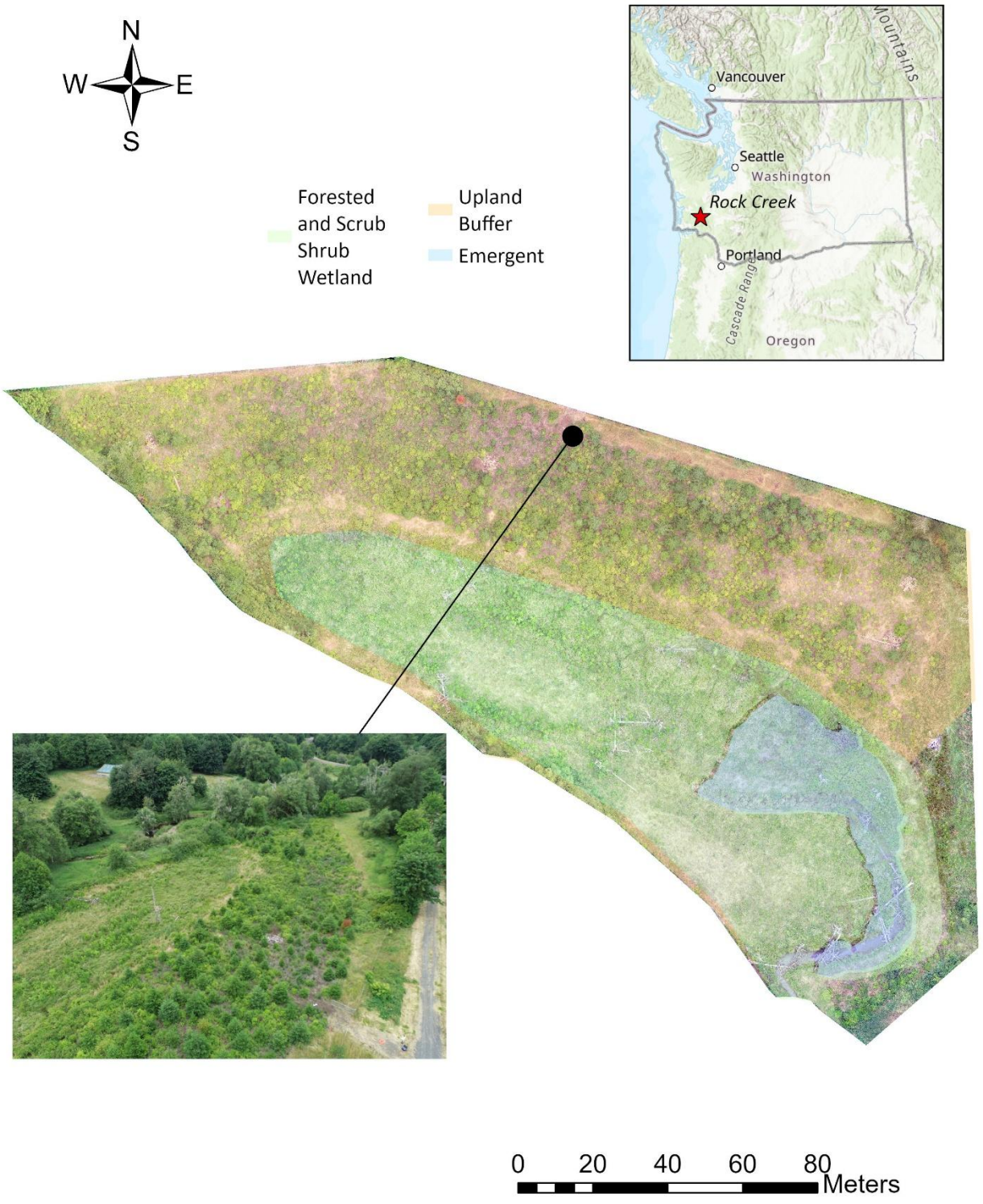

Figure 2.5 Rock Creek study site, showing forested and scrub shrub wetland, buffer, and emergent vegetation zones. WSDOT captured drone images for this site on July 12, 2023.

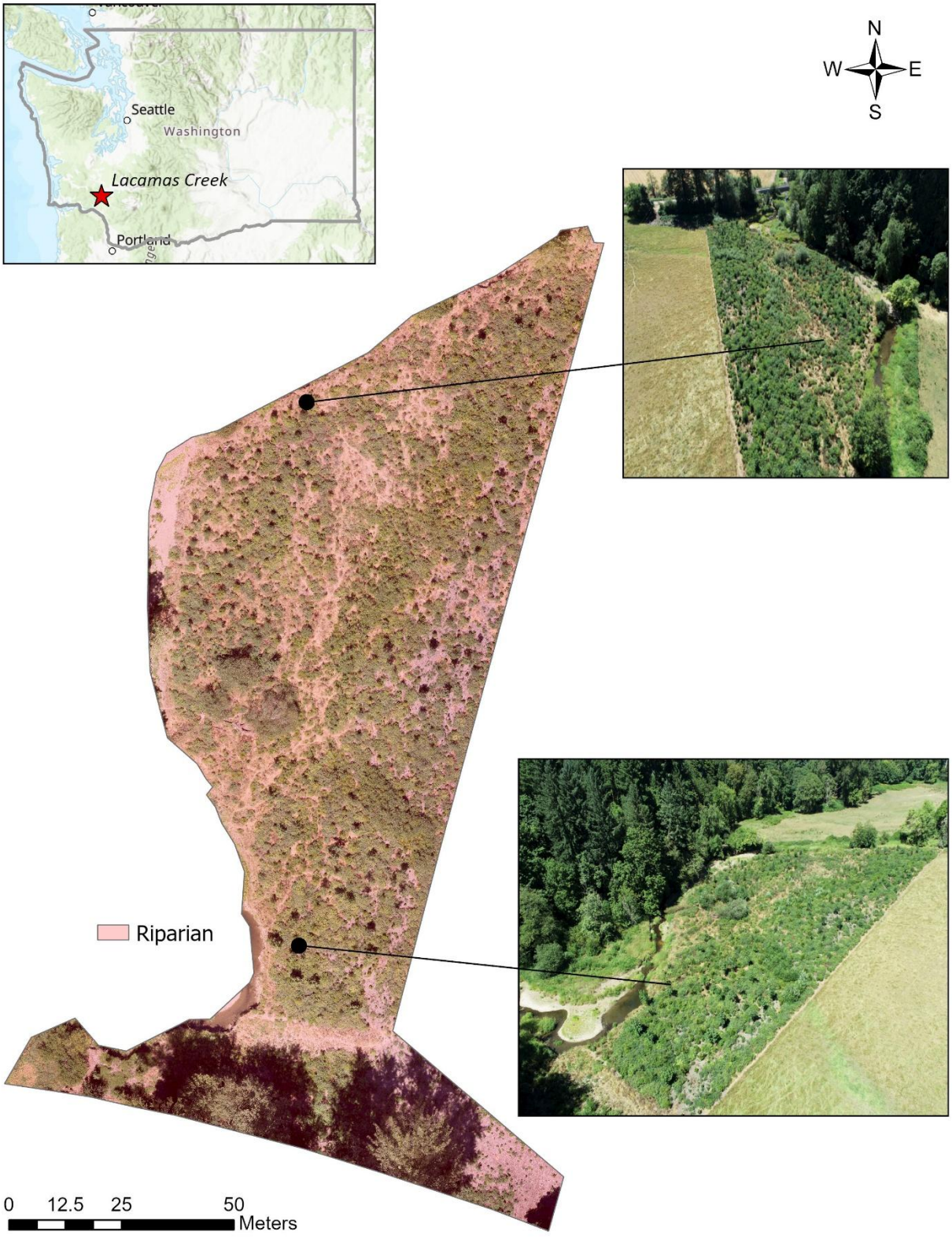

Figure 2.6 Lacamas Creek study site, showing riparian vegetation zone. WSDOT captured drone images for this site on August 2, 2023.

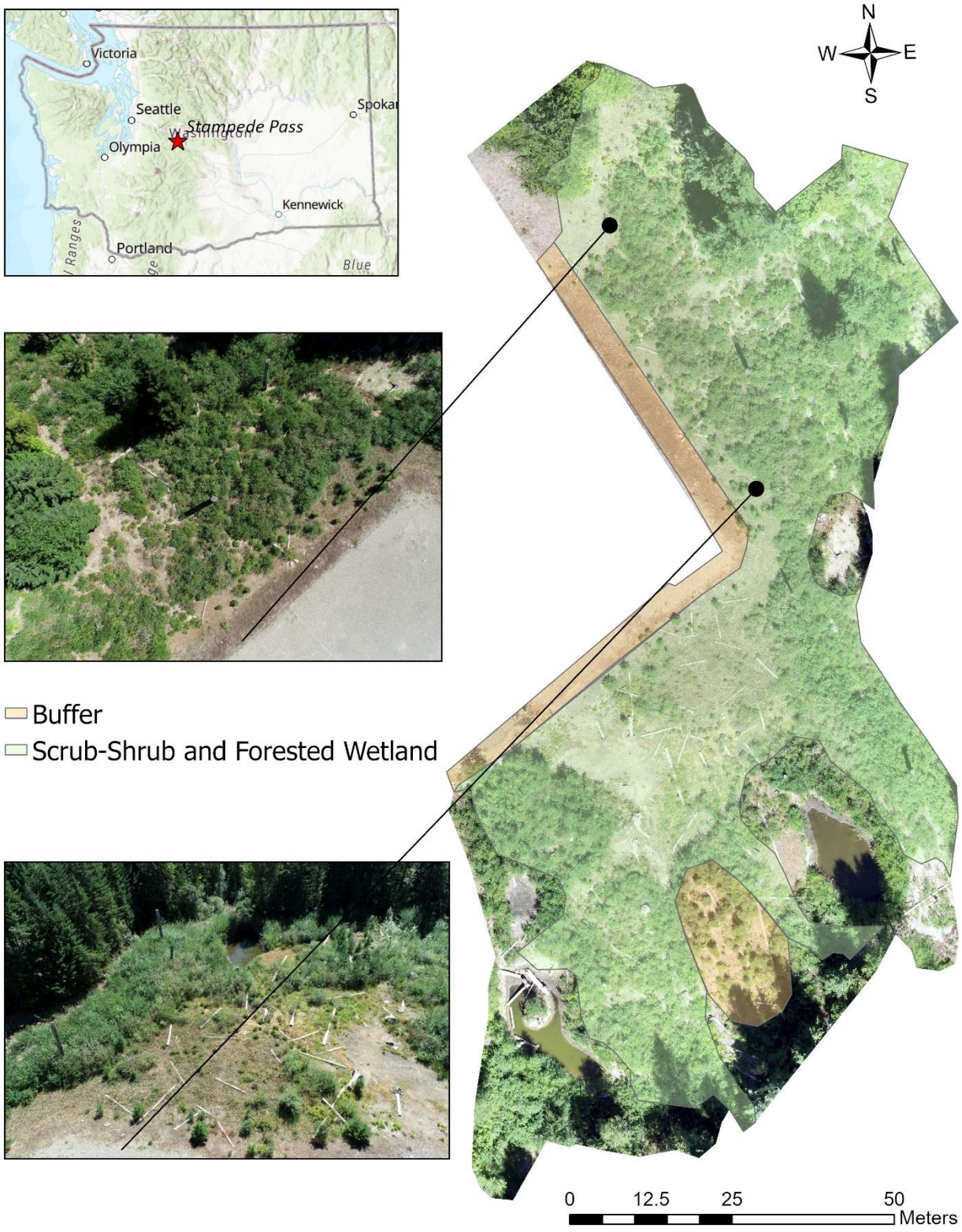

Figure 2.7 Stampede Pass study site, showing scrub shrub and forested wetland and buffer vegetation zones. WSDOT captured drone images for this site on August 16, 2023.

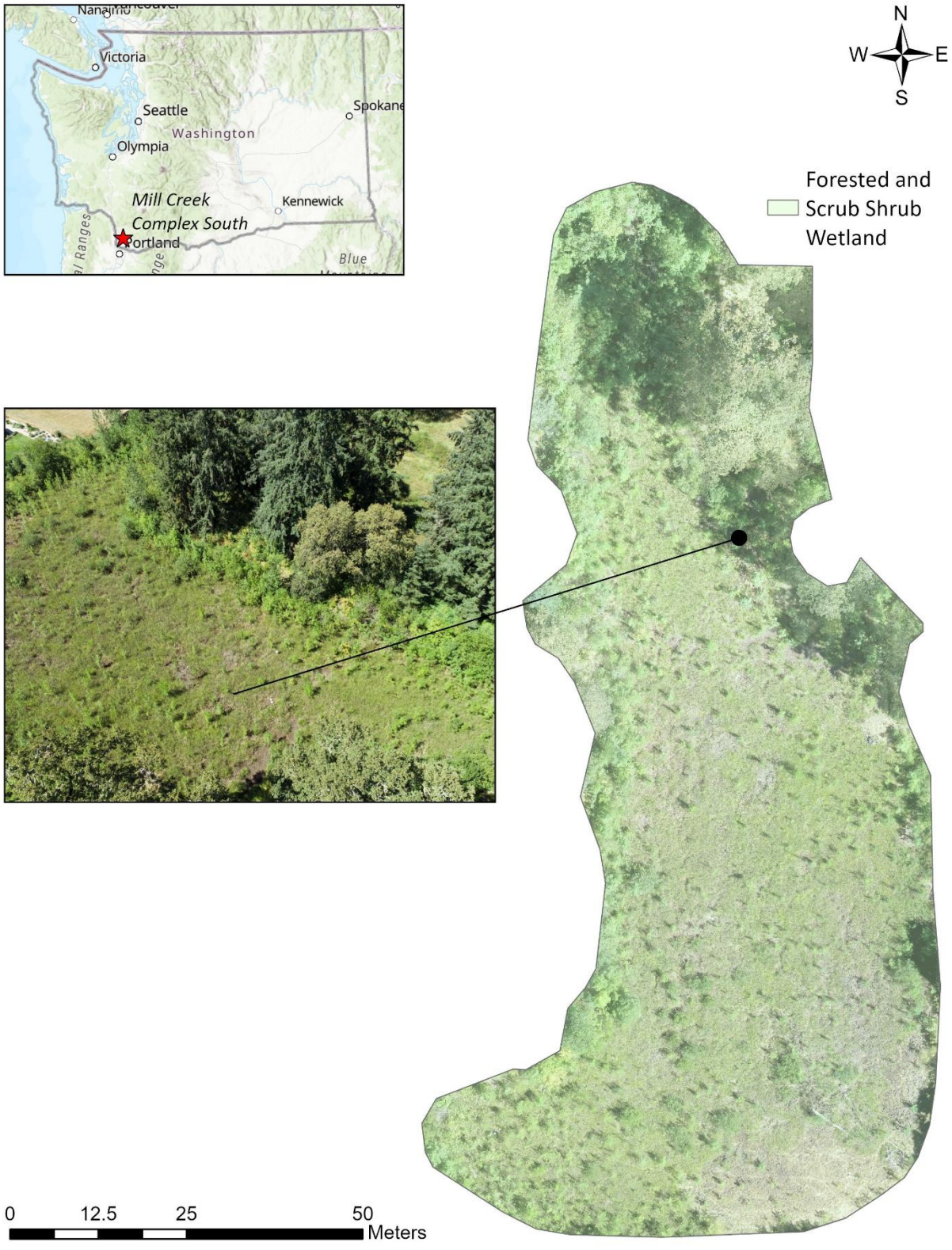

Figure 2.8 Mill Creek Complex South study site, showing the scrub shrub and forested vegetation zone. WSDOT captured drone images for this site on August 2, 2023.

2.2.2 UAV, sensor, and flight parameters

I conducted drone flights using a DJI Phantom 4 quadcopter equipped with an RGB (red-green-blue) sensor, a global positioning system (GPS), and a global navigation satellite system (GLONASS). I selected this drone for its affordability and user-friendliness, which aligns with agency needs (Halpern, 2020; Gomez-Sapiens et al., 2020; Haskins et al., 2021; Evans et al., 2022). I used the DroneDeploy Desktop application to create a flight plan for each mitigation site and operated the drone autonomously through the mobile app. I set the drone images to the WGS 84 coordinate reference system (EPSG: 4326).

I conducted drone surveys at all sites between July 12 and August 23, 2023 (Table 2.1). To minimize shadow effects, I flew the drones between 11:00 AM and 2:00 PM. I set the flight altitude to 150 feet above ground level for all sites, except at Mill Creek Complex South, where I increased it to 160 feet to account for larger vegetation. I configured the front and side image overlap to 80%.

Table 2.1 Drone flight and ground truthing field dates in 2023.

Site Name	Drone Flight	Ground Truthing
Four Mile	August 23	August 25
South Fork Newaukum	July 12	July 14
Lacamas Creek	August 2	August 4
Rock Creek	July 12	July 14
Stampede Pass	August 16	August 18
Mill Creek Complex South	August 2	August 4

At each study site, I deployed high-contrast ground control points (GCPs) to georeference the orthomosaic, digital surface models (DSMs), and point cloud data. Generally, I placed GCPs around the outer corners and the center of the site in areas free of vegetation; however, the number

and placement of GCPs varied based on site size. To assess the final absolute accuracy of the drone-derived products, I used checkpoints (CPs) at each site. GCPs ranged from 4 to 5 per site, while CPs varied between 2 and 3. I surveyed GCPs and CPs using a Trimble R2 GNSS unit to ensure precise geolocation of the GCPs and the final products.

2.2.3 Ground truthing

To train and validate the classification models, I conducted field surveys by measuring vegetation in 1m x 1m ground truthing plots within the same week as the drone flight (Table 2.1). In each plot, I estimated the foliar cover of woody (trees and shrubs) and herbaceous vegetation (Daubenmire, 1959). I also recorded the cover of bare ground and water, when present. I categorized percent cover as 0%; <1%; 1-5%; 5-15%; 15-25%; 25-50%; 50-75%; or >75% for each class. The ground-truthed plots were cross-referenced with the segmentation layer in object-based classifications to create training data for the model. For pixel-based classifications, the ground-truth plots were used to create polygons of the vegetation classes within them. These polygons were used as training data for the model. I surveyed the plots with a Trimble R2 for precise geolocation and integration to the classification process.

2.2.4 Drone data processing

I used *Pix4DMapper Pro* (Pix4D, n.d.) to process and georeference drone images at each site, generating an orthomosaic, point cloud, and DSM. Pix4D uses Structure-from-Motion (SfM) photogrammetry and Multi-View Stereo (MVS) to reconstruct 3D surfaces and create high-resolution geospatial outputs. The Pix4D workflow follows a structured process involving automatic aerial triangulation (AAT), bundle block adjustment (BBA), and orthomosaic generation

(Durgan, 2020). The workflow begins with AAT, which identifies matching key points across overlapping images to generate tie points, refining image positions using GCPs as spatial references (Durgan, 2020). Next, BBA optimizes the internal camera parameters, such as focal length, principal point, radial and tangential distortion, and estimates the absolute camera positions and orientations for each image. The process concludes when BBA converges on a solution, with an evaluation of adjustment error and accuracy metrics (Mora et al., 2019). Rather than relying solely on the drone's GPS coordinates, which may contain errors of a few meters, BBA uses the relationships between overlapping images to refine camera locations and orientations with higher accuracy.

The 2D and 3D drone products achieved sub-meter relative and absolute accuracy, ranging between 0.09-0.3 m for the horizontal component and 0.009-0.8 m for the vertical component (Table 2.2). I calculated relative accuracy using the GCP root-mean-square error (RMSE), which quantifies the precision of measurements relative to each other within the dataset (Durgan, 2020). While GCP RMSE measures relative accuracy, GCPs also enhance absolute accuracy by anchoring the model to real-world coordinates. I assessed absolute accuracy using the CP RMSE, which evaluates how accurately the drone-derived data represents the true positions and features on the Earth's surface. Since CPs serve as independent validation points rather than inputs in processing, they provide an external check on model accuracy. Initially, without GCPs, GPS errors resulted in drone image positions with an accuracy range of 2–10 meters. Incorporating GCPs improved positional accuracy to sub-meter levels.

The spatial resolution of the orthomosaics ranged from 1.24 to 1.52 cm, depending on the site (Table 2.2). Spatial resolution refers to the ground area represented by each pixel in the image.

For example, a resolution of 1.24 cm means that each pixel corresponds to a 1.24 cm x 1.24 cm area on the ground.

At Mill Creek Complex South, Pix4D encountered issues with larger trees in the oak preservation and upland communities. While flying at higher elevations could have resolved these issues, I prioritized mapping woody vegetation in the wetland. By the time I processed the images, it was too late in the season to re-fly the site while capturing the same vegetation characteristics present in the summer. As a result, I excluded the oak preservation and upland buffer areas, focusing the image classification solely on the created and enhanced wetland portions.

Table 2.2 Root Mean Square Error (RMSE) for GCPs and CPs, along with spatial resolution across study sites.

RMSE values for Ground Control Points (GCPs) and Checkpoints (CPs) in the X, Y, and Z directions across six study sites, expressed in meters (m). Image resolution for each site is provided in centimeters (cm). GCPs were used for model calibration to assess relative accuracy, while CPs served as independent validation points for absolute accuracy.

Site	**Type**	**X**	**Y**	**Z**	**Resolution**
Four Mile	GCP (4)	0.01	0.01	0.009	1.24
Four Mile	CP (3)	0.02	0.02	0.01	1.24
Rock Creek	GCP (5)	0.07	0.09	0.3	1.31
Rock Creek	CP (3)	0.2	0.2	0.5	1.31
Stampede Pass	GCP (4)	0.08	0.3	0.2	1.52
Stampede Pass	CP (3)	0.01	0.5	0.3	1.33
Mill Creek	GCP (3)	0.02	0.02	0.003	1.38
Mill Creek	CP (2)	0.03	0.01	0.1	1.38
Lacamas Creek	GCP (4)	0.1	0.3	0.6	1.52
Lacamas Creek	CP (3)	0.3	0.2	0.8	1.52
South Fork	GCP (4)	0.09	0.07	0.08	1.26
South Fork	CP (3)	0.1	0.04	0.3	1.26

2.2.5 Classification workflow

I used supervised classification methods, including both pixel-based and object-based approaches, to perform image classification in *ArcGIS Pro* 3.3.0 (Esri, 2023). I applied the built-in SVM and RT algorithms for classification. At each site, I conducted four classifications: two using a pixel-based approach and two using an object-based approach. At Four Mile and Rock Creek, I performed two additional classifications incorporating fused vertical data from a relativized DSM. Figure 2.9 provides a breakdown of each classification trial, and the following sections outline the workflow methods used for all classifications.

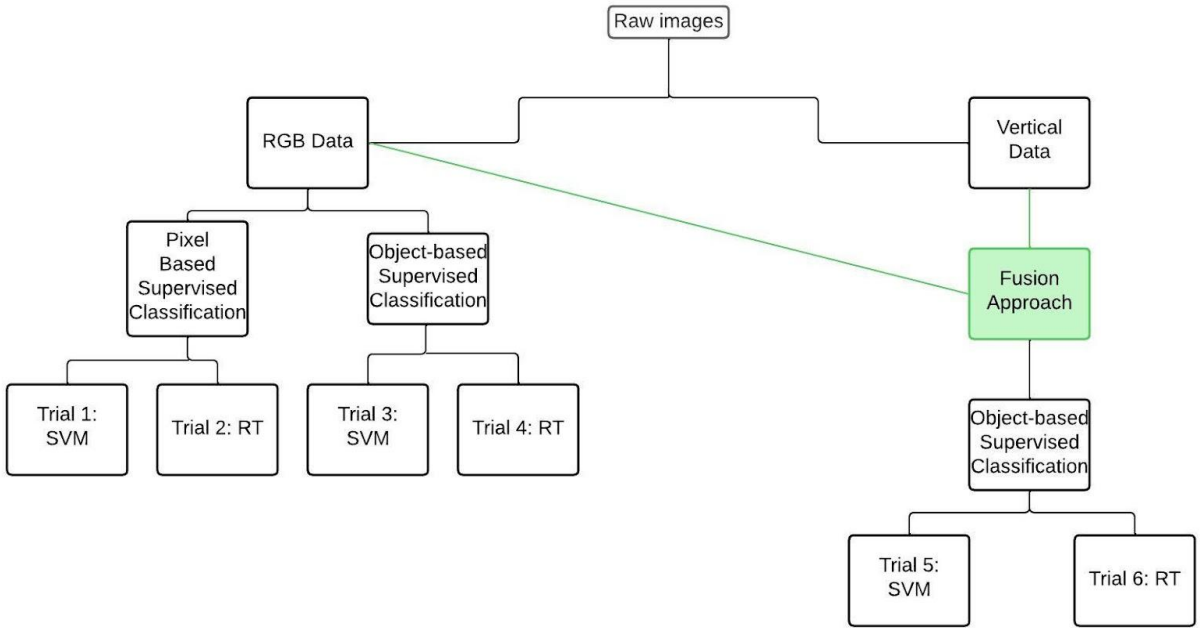

Figure 2.9 Summary of classification trials.

I conducted four classifications (Trials 1–4) at each mitigation site and two additional classifications (Trials 5 and 6) at Four Mile and Rock Creek.

2.2.5.1 Classification scheme

My target class was woody vegetation (Figure 2.10, A). I grouped the remaining pixels into several classes depending on site characteristics: herbaceous (alive-green pixels and dead-brown pixels), bare ground, shadow, and water (Figure 2.10, B-F). Each site classification includes at least woody, herbaceous (green or brown), bare ground, and shadows.

The woody vegetation class included all shrubs and trees, both deciduous and evergreen. I divided the herbaceous class into "green-alive" and "brown-dead" categories due to spectral differences between wetland emergent vegetation (green) and drier upland species (brown). The green herbaceous class typically appeared in shades of light to dark green, while the brown herbaceous class ranged from light brown to light gray.

The bare ground class, which included dirt, mulch, leaf litter, and debris, was spectrally distinct from the brown-dead herbaceous class due to its greater representation of darker brown pixels. While both classes share overlapping brown tones, the bare ground class often included more mixed and darker brown hues resulting from soil and organic debris. In contrast, the brown-dead herbaceous class, dominated by dead upland vegetation, typically had lighter brown and gray spectral signatures. The shadows class included darker gray and black pixels, while the water class included any water bodies, such as creeks, ponds, or areas of wetland inundation, with pixel colors ranging from muddy brown to green blue.

Additionally, some sites included a class for large woody debris to account for habitat structures intentionally added to the site. I classified these features, generally linear in shape and gray in color, separately from vegetation classes. At the Four Mile site, I created a class for stressed vegetation, as a subset of herbaceous vegetation, to account for red pixels, which were too distinct to be included in any existing vegetation classes.

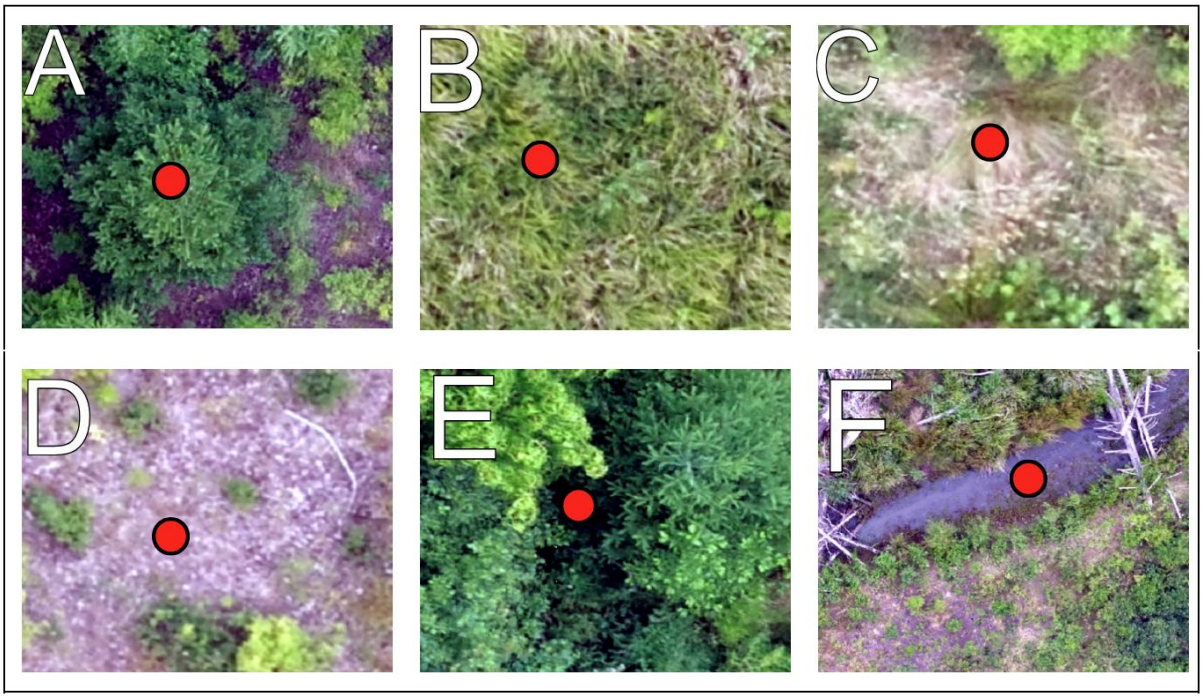

Figure 2.10. Classification Scheme.

(A) woody vegetation, (B) herbaceous-green-alive, (C) herbaceous-brown-dead, (D) bare ground, (E) shadows, (F) water. The red dot indicates an example of the depicted class.

2.2.5.2 Training and validation data

I tailored the training and validation data to the classification method (pixel-based or object-based). For pixel-based (PB) classification, I manually digitized training and validation polygons using ground-truthed data for each class. In contrast, for object-based (OB) classification, I used segments generated by the image segmentation tool for each class. I cross-referenced each segment with ground-truthed data to ensure accuracy for each class. I used approximately 10%-30% of the orthomosaic, or 500-1000 samples per site, for training data, with an additional 10% (300-500 samples) set aside for validation.

The size and number of training polygons or segments varied depending on the feature and its presence in the image. For example, the woody vegetation class in wetland zones required more

training samples and larger polygons due to the spectral similarity between wetland woody vegetation and herbaceous-alive vegetation, as both occupied a similar spectral range of green. In contrast, woody vegetation in upland areas was more distinct from upland herbaceous vegetation, as grasses and forbs in these areas tended to be drier in the summer, appearing light brown or gray compared to the green pixels of shrubs and trees. Similarly, dry herbaceous vegetation in the upland appeared spectrally similar to bare ground, which increased the amount or size of training polygons needed for accurate results. In general, the amount of training and validation data was proportional to the class's presence within an image, with less dominant classes having fewer data.

2.2.5.3 Image segmentation

In *ArcGIS Pro*, segmentation involves dividing an image into distinct objects or pixel clusters based on spectral, spatial, and size characteristics (Esri, n.d.a). For object-based classification, I segmented the orthomosaic before running the classification model using the Image Segmentation/Segment Mean Shift tool. I adjusted the parameters for spectral detail, spatial detail, and segment size iteratively, roughly ten to fifteen times, until the segmentation effectively captured the woody features of interest. These values varied by site, typically ranging from 15 to 20 for both spatial and spectral detail, with a minimum segment size set to 20. Higher values for spectral detail resulted in smaller, more detailed segments, while higher values for spatial detail produced larger segments, enhancing spatial groupings. The minimum segment size determined the fewest number of pixels a segment could contain, with larger values resulting in larger segments (Gómez-Sapiens et al., 2021).

2.2.5.4 Supervised pixel classification (trials 1 and 2)

In a supervised classification, the raster is classified using a set of training samples to guide the process. I chose this method over unsupervised classification due to the low spectral resolution of the orthomosaic (RGB), which made it difficult to distinguish land cover classes without training data. Unsupervised classification, which groups pixels into clusters based on inherent spectral similarities, may work well for land cover classes that are spectrally distinct; however, the woody and herbaceous classes at these mitigation sites were too similar to be reliably separated using this method.

In a pixel-based (PB) classification, the machine learning classifier assigns each pixel to a class based solely on the pixel's individual properties, without considering the characteristics of neighboring pixels. This approach does not incorporate spatial relationships, and segmentation is not applied. For this classification, I selected the classifier, along with training and validation data, and processed them using the Classification Wizard (Esri, n.d.b).

In Trial 1, I set the machine learning classifier parameters for the SVM trials to 500 samples per class, which determined the number of training samples used from each class during the model's training phase (Esri, n.d.c). In Trial 2, I set the RT parameters to 50 trees, a maximum tree depth of 30, and a maximum of 500 samples per class. The number of trees influences the robustness of the model, with more trees improving accuracy but increasing computational time. The tree depth refers to the number of splits each tree can make, with deeper trees capturing more complex patterns in the data (Esri, n.d.d).

2.2.5.5 Supervised object classification

In a supervised object-based classification, pixels are classified based on user-identified objects, or clusters of pixels, generated through image segmentation. Instead of relying solely on individual pixel RGB values, these segments are classified using machine learning algorithms, considering the properties of the entire object or cluster. For this classification, I first segmented the image, then selected the appropriate classifier, input the training and validation data, and used the Classification Wizard to perform the classification. In Trial 3, I set the machine learning classifier parameters for the SVM trials to 500 samples per class. In Trial 4, I set the RT parameters to 50 trees, a maximum tree depth of 30, and a maximum of 500 samples per class.

2.2.5.6 Vertical data fusion

I evaluated two additional classification trials at Four Mile and Rock Creek, which incorporated vertical data from the drone DSM. First, I relativized the DSM raster at each site to a 0-1 range by scaling its values using the raster's minimum and maximum. Next, I created a composite raster by combining the RGB orthomosaic with the relativized DSM (rDSM) resulting in a 4-band raster. I performed an object classification using the Classification Wizard. The composite raster served as the input raster, incorporating both spectral (RGB) and vertical data (rDSM). I applied the same SVM and RT parameters described in Trials 3 and 4 (see Section 2.2.5.5).

2.2.5.7 Accuracy assessment

I aimed to classify drone imagery with at least 85% overall classification accuracy, as well as 85% producer and user accuracies for the woody class. To assess the accuracy of the

classification trial, I employed a stratified random sample with approximately 500 sampling points per site (Esri, n.d.e). The points were randomly distributed within each class, with the number of points proportional to the class area (Gómez-Sapiens et al., 2021). This process generated a confusion matrix and Kappa coefficient (Table 2.3). The confusion matrix compared the classified data (rows) to the validation data (columns). Each cell represents the number of pixels that were classified into a particular class (rows) compared to their true class, determined by the validation data. The diagonal values indicate correct classifications, where the classified data matches the validation data. The off-diagonal values represent misclassifications, where pixels were assigned to the wrong class.

Table 2.3 Confusion matrix for Trial 1 (pixel-based SVM) at Lacamas Creek.

Bold values indicate correctly classified pixels per class. Overall accuracy (OA), producer accuracy (PA), and user accuracy (UA) for the woody class are ***highlighted*** and shown as percentages. Brown Herbs and Green Herbs refer to the two herbaceous classes brown-dead and green-alive.

Classified Data	Woody	Bare	Brown Herbs	Shadows	Green Herbs	Samples	***UA***
Woody	**179**	0	4	7	3	193	***93%***
Bare	5	**81**	1	0	0	87	93%
Brown Herbs	11	9	**73**	4	4	101	72%
Shadows	0	0	0	**22**	0	22	100%
Green Herbs	61	0	1	0	**35**	97	36%
Samples	256	90	79	33	42	500	
PA	***70%***	90%	92%	67%	83%		
OA							***78%***
Kappa							0.69

Producer accuracy, also known as omission, measures the false negatives and indicates how accurately the model meets the expectations of the producer (Esri, n.d.e). An example is when the validation data identifies a pixel as woody, but the classification assigns it to green herbaceous.

The woody class is consequently missing pixels according to the validation data. User accuracy, also referred to as commission, shows false positives, where pixels are incorrectly classified as a known class when they should have been classified as something else (Esri, n.d.e). An example is when the classification assigns a pixel to woody, but the reference data shows it as green herbaceous. The woody class contains extra pixels that should be in the green herbaceous class, containing false positives.

$$PA = \frac{Total\ classified\ points\ from\ validation\ data}{Total\ reference\ samples\ for\ that\ class}$$

$$UA = \frac{Total\ classified\ points\ from\ validation\ data}{Total\ classified\ samples\ for\ that\ class}$$

The Kappa coefficient is a statistical measure of agreement between the classified and validation data. The range of possible values is between 0 and 1, with 1 indicating high classification accuracies (Esri, n.d.e). The formula to calculate the Kappa coefficient is:

$$k = \frac{P_O - P_e}{1 - P_e}$$

Where, P_O represents the overall accuracy. The overall accuracy is calculated as the proportion of correctly classified samples, or the sum of the diagonal values in the confusion matrix divided by the total number of samples.

$$P_O = \frac{sum\ of\ diagonal\ values}{total\ number\ of\ samples}$$

P_e represents the expected agreement by chance and is calculated based on the marginal totals of the confusion matrix.

$$P_e = \Sigma(\frac{row\ total\ x\ column\ total}{total\ number\ of\ samples^2})$$

2.2.6 Statistical analysis

I performed statistical analyses using R software (Version 4.1.2; R Core Team, 2021). I conducted three, two-way fixed effects ANOVA models to test the classification accuracies for each mitigation site under different classification methods (pixel-based vs. object-based) and classifiers (SVM vs. RT). The analysis tested the combined effects of the classification method and classifier on overall accuracy of the model, and producer and user accuracy of the woody class. I treated the data as blocked, with each site contributing four data points, one for each trial (Trials 1-4). The classification method and classifiers were combined into a single factor, while the site was treated as blocking term.

After performing the two-way fixed effects ANOVA, I tested the residuals for normality using a Shapiro-Wilk normality test. Additionally, I performed pairwise comparisons to determine how the different levels of method (pixel-based vs. object-based) and classifier (support vector machine vs. random trees) influenced overall, producer, and user accuracy.

For the vertically fused classification approach at Four Mile and Rock Creek, I summarized the classification results using descriptive statistics. Since this approach was applied to only two mitigation sites, the limited data points made statistical testing infeasible.

2.3 Results

2.3.1 Summary of classification accuracies for pixel-based approach (trials 1-2)

The overall accuracy of the pixel-based classifications (Trials 1 and 2) was low, ranging from 61% at Four Mile to 84% at Mill Creek Complex South (Table 2.4). Across all mitigation sites, Trial 1 achieved an overall accuracy of 74%. In Trial 1 (SVM), producer accuracy for the woody class varied from 19% at Four Mile to 91% at South Fork Newaukum. User accuracy ranged from 58% at Four Mile to 96% at Stampede Pass. User accuracy was higher than producer accuracy by approximately 20%. No site achieved greater than 85% overall accuracy, and South Fork Newaukum was the only site to achieve greater than 85% user and producer accuracy for the woody class.

In Trial 2 (RT), producer accuracy for the woody class ranged from 51% at Four Mile to 88% at South Fork Newaukum, while user accuracy ranged from 78% at Four Mile to 99% at Stampede Pass. Trial 2 achieved an overall accuracy of 76% across all sites. User accuracy was higher than producer accuracy by approximately 20%. Similar to Trial 1, no site surpassed 85% overall accuracy and 85% for user and producer accuracy in the woody class. South Fork Newaukum, again, recorded the highest accuracy, with producer and user accuracy for the woody class reaching 88% and 95% respectively.

Table 2.4 Accuracy Metrics by Trial and Site

This table presents overall accuracy (OA), producer accuracy (PA), and user accuracy (UA) for each trial and site in percentages (%). Averages for each trial are **bolded** and placed at the **top** of the trial group. The averages provide a summary of classification performance across all sites within each trial. Trial 1 = pixel, SVM; Trial 2 = pixel, RT; Trial 3 = object, SVM; Trial 4 = object, RT; Trial 5 = object, SVM, vertical fusion; Trial 6 = object, RT, vertical fusion

Site	Trial	Overall Accuracy (%)	Producer Accuracy (%)	User Accuracy (%)
Average	**1**	**74.3**	**64.1**	**88.3**
Four Mile	1	61.2	18.7	57.5
Lacamas Creek	1	78.0	69.9	92.7
Mill Creek Complex South	1	83.7	76.9	95.0
Rock Creek	1	66.7	64.4	93.5
South Fork Newaukum	1	77.2	91.3	94.6
Stampede Pass	1	79.2	63.4	96.3
Average	**2**	**75.6**	**70.8**	**91.0**
Four Mile	2	69.6	51.2	77.8
Lacamas Creek	2	76.4	73.0	94.9
Mill Creek Complex South	2	82.9	81.9	94.2
Rock Creek	2	69.9	64.0	85.5
South Fork Newaukum	2	75.6	88.2	94.8
Stampede Pass	2	79.0	66.3	98.6
Average	**3**	**84.3**	**86.1**	**90.8**
Four Mile	3	72.5	62.6	81.9
Lacamas Creek	3	84.8	86.3	93.2
Mill Creek Complex South	3	94.8	97.5	97.0
Rock Creek	3	76.0	80.2	89.4
South Fork Newaukum	3	85.2	95.5	98.6
Stampede Pass	3	92.9	94.6	84.6
Average	**4**	**81.8**	**86.0**	**84.5**

Site	Trial	Overall Accuracy (%)	Producer Accuracy (%)	User Accuracy (%)
Four Mile	4	73.9	54.5	72.8
Lacamas Creek	4	79.8	91.0	84.1
Mill Creek Complex South	4	92.3	95.0	91.7
Rock Creek	4	73.4	84.7	84.7
South Fork Newaukum	4	85.6	98.3	90.4
Stampede Pass	4	85.9	92.7	83.3
Average	**5**	**80.5**	**84.5**	**86.0**
Four Mile	5	80.0	78.0	83.0
Rock Creek	5	81.0	91.0	89.0
Average	**6**	**81.5**	**91.0**	**85.0**
Four Mile	6	82.0	92.0	82.0
Rock Creek	6	81.0	90.0	88.0

2.3.2 Summary of classification accuracy for object-based approach (trials 3-4)

The overall accuracy of the object-based classifications (Trials 3 and 4) ranged from 73% at Four Mile to 93% at Mill Creek Complex South (Table 2.4). Trial 3 (SVM) achieved an overall accuracy of 84%, while Trial 4 (RT) achieved an overall accuracy of 82% across all mitigation sites. In Trial 3, producer accuracy was lower than user accuracy by less than 5%, and in Trial 4, producer accuracy was slightly higher than user accuracy. These differences show an improvement compared to the gap exhibited in Trials 1 and 2 between user and producer accuracy. Mill Creek Complex South, South Fork Newaukum, Lacamas Creek, and Stampede Pass achieved greater than 85% user and producer accuracy for the woody class and greater than 85% overall accuracy for the classification.

2.3.3 Summary of classification accuracy for object-based approach with vertical fusion (trial 5 and trial 6 for Four Mile and Rock Creek)

Although I did significantly test any differences in Trials 5 (SVM) and 6 (RT), these trials performed higher than Trials 1-4 at Rock Creek. The overall accuracy in Trial 5 was 5% higher than in Trial 3, and the overall accuracy in Trial 6 was nearly 10% higher than in Trial 4 (Table 2.4). Trial 5 showed an improvement in producer accuracy for the woody class, increasing from 80% to 91%, while the user accuracy remained consistent at 89% (Figure 2.11). Figure 2.11 illustrates the agreement between Trials 3 and 5, with green indicating areas of no change, red representing changes to the woody class, and blue showing changes in all other classes. In Trial 6, the producer accuracy for the woody class improved from 85% to 90%, with similar user accuracy. These improvements indicate that fewer woody class pixels were misclassified into other classes, decreasing the errors of omission and increasing the producer accuracy. As a result, overall accuracy was also higher, but the misclassifications between the other classes prevented the trials from meeting the target accuracy of 85% across the three accuracy metrics.

At Four Mile, the overall accuracies in Trials 5 and 6 were nearly 10% higher than those in Trials 3 and 4, respectively (Table 2.4). Trial 5 (SVM) improved the producer accuracy for the woody class by 15% compared to Trial 3, while the user accuracy remained consistent at approximately 82%. Trial 6 (RT) demonstrated even greater improvements, with the producer accuracy increasing by nearly 40% and the user accuracy improving by 10% compared to Trial 4 (Figure 2.12). Figure 2.12 highlights areas of agreement and disagreement between Trial 4 and Trial 6. Although Trial 6 yielded the highest overall accuracy and the highest producer and user accuracy for the woody class at Four Mile, it still failed to meet the target accuracy goals due to misclassifications between the non-target classes.

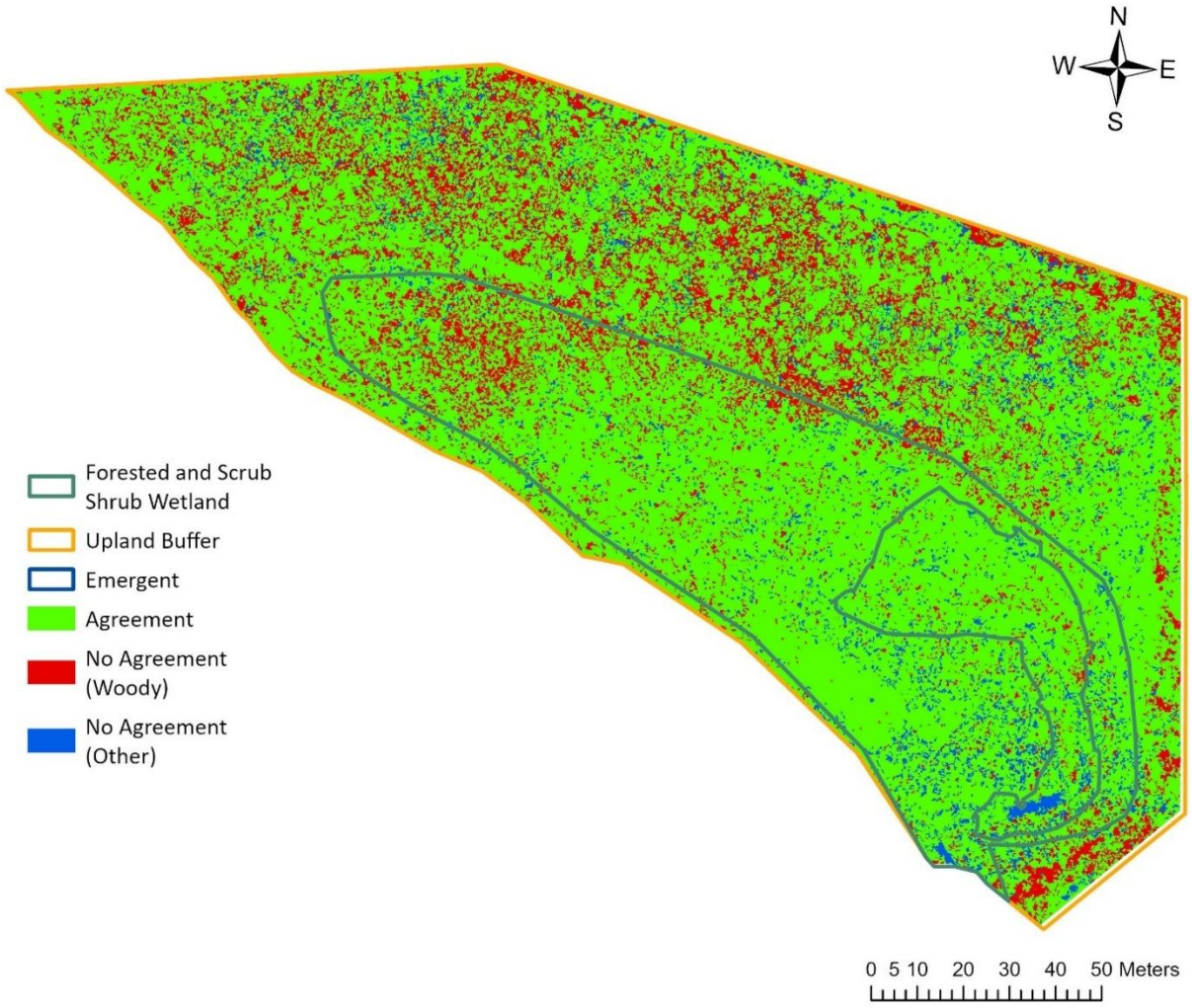

Figure 2.11 Agreement map comparing classifications from Trial 3 and Trial 5 at Rock Creek.

Green areas indicate agreement across all classes between both trials. Red pixels represent classification differences in the woody class, while blue pixels indicate disagreement in all other classes.

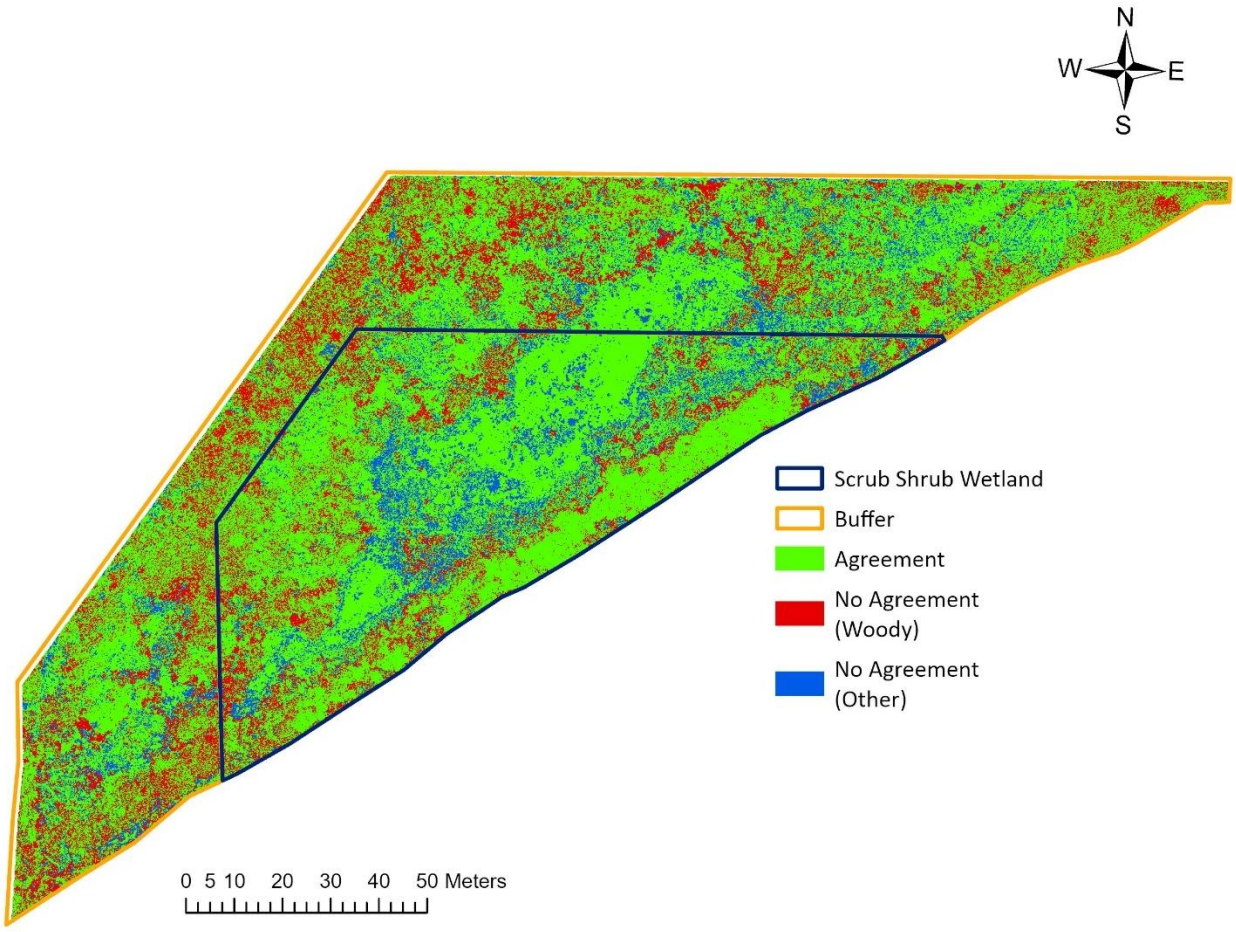

Figure 2.12 Agreement map comparing classifications from Trial 4 and Trial 6 at Four Mile.

Green areas indicate agreement across all classes between both trials. Red pixels represent classification differences in the woody class, while blue pixels indicate disagreement in all other classes.

2.3.4 Fixed-effects ANOVA model

Overall classification accuracy was significantly influenced by site and by the combined method and classifier term (Table 2.5, Figure 2.13). The site effect was highly significant, suggesting that site-specific factors played a significant role in influencing the classification accuracy. The residual variability was relatively small compared to the variability explained by the site and method/classifier terms. This suggests that the model accounted for most of the variation

in the data, demonstrating a good model fit. The Shapiro-Wilk test confirmed that the model's residuals were normally distributed (W=1.0, p = 0.8).

Table 2.5 ANOVA results examining the variance and significance of a combined method/classifier term on overall accuracy.

	Df	Sum.Sq	Mean.Sq	F.value	p.value
Site	5	0.11	0.02	38.00	< .001 ***
Method/Classifier	3	0.04	0.01	24.65	< .001 ***
Residuals	15	0.01	0.00		

The pairwise comparisons of the Method Classifier term revealed two distinct groups (Table 2.6). Confidence intervals do not show overlap between groups, highlighting distinct performance differences between pixel-based and object-based approaches. Pixel-based methods (Group a -Trials 1 and 2) performed significantly lower than object-based methods (Group b - Trials 3 and 4). Within each group, there was no significant difference in performance between methods, highlighting that overall accuracy was strongly influenced by the classification method rather than the machine learning classifier.

Table 2.6 Pairwise comparisons of the combined method/classifier term showing estimated marginal means (emmeans) for overall accuracy. Lower.CL and Upper.CL indicate the lower and upper limits of the 95% confidence interval, respectively, providing a range within which the true accuracy is expected to lie. The Group column indicates the results of pairwise comparisons, where methods sharing the same letter are not significantly different from each other.

Method Class	emmean	SE	df	Lower.CL	Upper.CL	Group
Pixel SVM (Trial 1)	0.74	0.01	15	0.72	0.76	a
Pixel RT (Trial 2)	0.76	0.01	15	0.74	0.78	a
Object RT (Trial 3)	0.84	0.01	15	0.82	0.86	b
Object SVM (Trial 4)	0.82	0.01	15	0.80	0.84	b

Producer accuracy for the woody class was significantly influenced by site and by the combined method and classifier term (Table 2.7, Figure 2.14). The residual variability was small compared to the variability explained by the combined method/classifier term, and the site term, suggesting a good model fit. The Shapiro-Wilk test confirmed that the residuals were normally distributed ($W = 1.0$, $p = 0.54$).

Pairwise comparisons revealed that pixel-based methods (Group a - Trials 1 and 2) had a significantly lower producer accuracy, while object-based methods (Group b - Trials 3 and 4) had a higher producer accuracy (Table 2.8). Within each group, there was no significant difference between classifiers (Trial 1 vs. Trial 2, and Trial 3 vs. Trial 4).

Table 2.7 ANOVA results examining the variance and significance of a combined method/classifier term on producer accuracy.

	Df	**Sum.Sq**	**Mean.Sq**	**F.value**	**p.value**
Site	5	0.53	0.11	21.20	< .001 ***
Method/Class	3	0.22	0.07	14.75	< .001 ***
Residuals	15	0.08	0.00		

Table 2.8 Pairwise comparisons of the combined method/classifier term showing estimated marginal means (emmeans) for producer accuracy. Lower.CL and Upper.CL indicate the lower and upper limits of the 95% confidence interval. The Group column indicates the results of pairwise comparisons, where methods sharing the same letter are not significantly different from each other.

Method Classifier	**emmean**	**SE**	**df**	**Lower CL**	**Upper CL**	**group**
Pixel SVM (Trial 1)	0.64	0.03	15	0.58	0.70	a
Pixel RT (Trial 2)	0.71	0.03	15	0.65	0.77	a
Object SVM (Trial 3)	0.86	0.03	15	0.80	0.92	b
Object RT (Trial 4)	0.86	0.03	15	0.80	0.92	b

User accuracy for the woody class was significantly influenced by site but not by the combined method and classifier term (Table 2.9, Figure 2.15). Pairwise comparisons reinforced the ANOVA findings by grouping all trials together (Table 2.10). The residual error was relatively small compared to the variability explained by both the combined method/classifier and site factor terms, suggesting that the model fit the data well. Although the Shapiro-Wilk test indicated that there was some departure from normality in the residuals (W=0.91, p = 0.04), the small residual error from the ANOVA model supported a good model fit despite this non-normality.

Table 2.9 ANOVA results examining the variance and significance of a combined method/classifier term on user accuracy.

	Df	**Sum.Sq**	**Mean.Sq**	**F.value**	**p.value**
Site	5	0.14	0.03	7.74	< .001 ***
Method/Classifier	3	0.02	0.01	1.54	0.2449
Residuals	15	0.05	0.00		

Table 2.10 Pairwise comparisons of the combined method/classifier term showing estimated marginal means (emmeans) for user accuracy. Lower.CL and Upper.CL indicate the lower and upper limits of the 95% confidence interval. The Group column indicates the results of pairwise comparisons, where methods sharing the same letter are not significantly different from each other.

Method Class	**emmean**	**SE**	**df**	**Lower.CL**	**Upper.CL**	**Group**
Object RT (Trial 4)	0.84	0.02	15	0.79	0.90	a
Pixel SVM (Trial 1)	0.88	0.02	15	0.83	0.93	a
Object SVM (Trial 3)	0.91	0.02	15	0.86	0.96	a
Pixel RT (Trial 2)	0.91	0.02	15	0.86	0.96	a

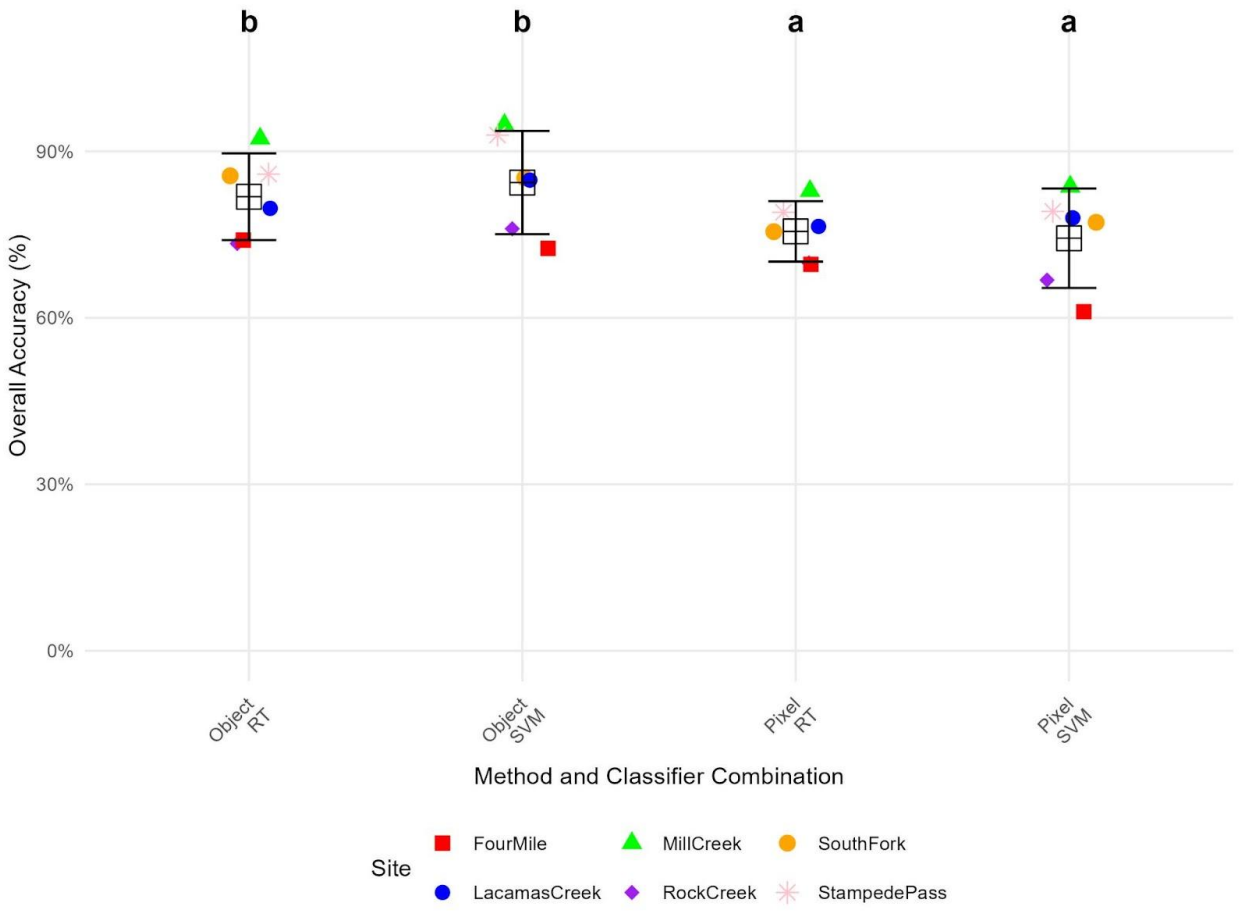

Figure 2.13 Overall accuracy by classification method and classifier across six sites.

Points represent individual site-level accuracies for each combination of method (pixel-based vs. object-based) and classifier (SVM vs. RT). The black boxes denote the mean accuracy for each classification trial (Method/Class combination), with error bars representing the 95% confidence intervals. Object-based methods (Group b - Trials 3 and 4) showed higher accuracies compared to pixel-based methods (Group a - Trials 1 and 2). Four Mile and Rock Creek consistently exhibited the lowest overall accuracies across all trials.

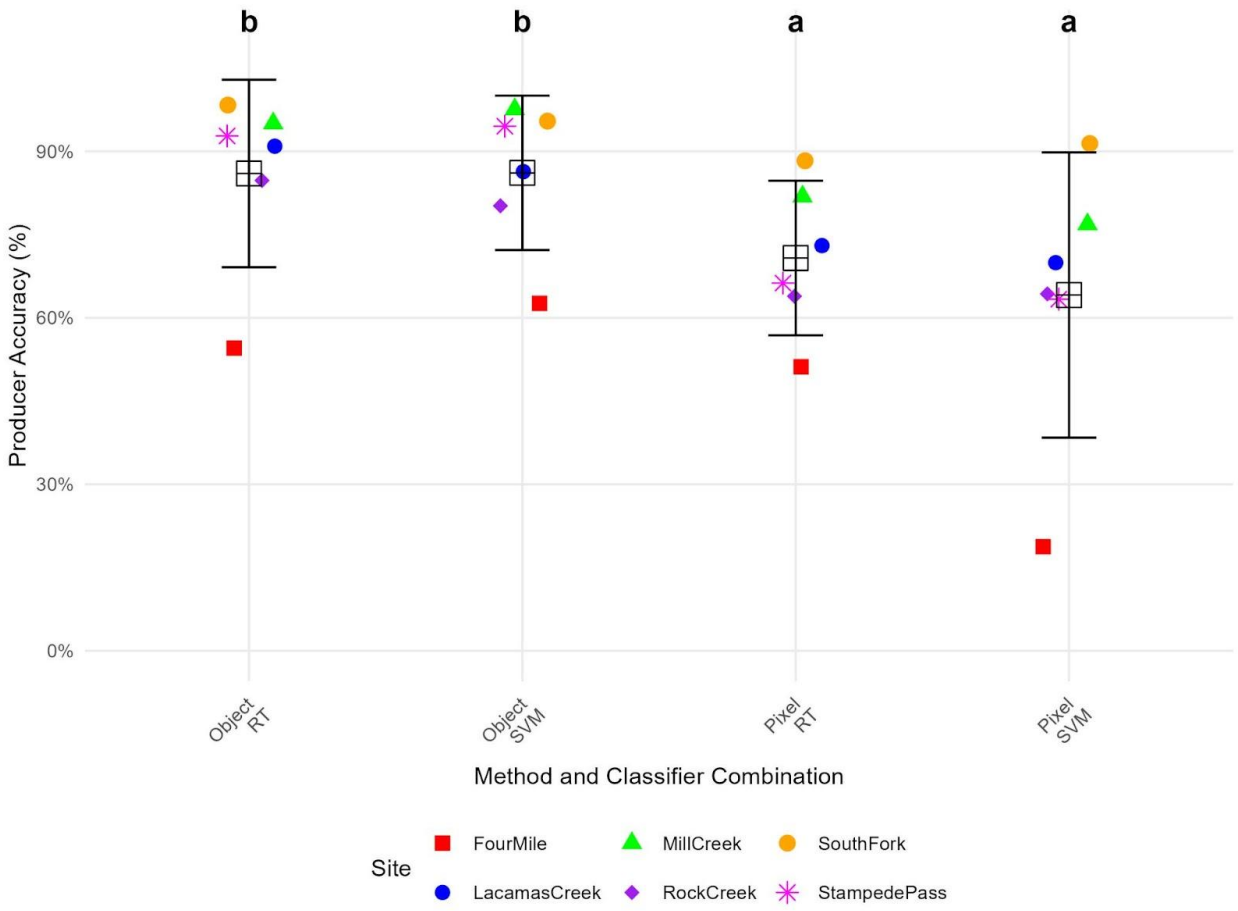

Figure 2.14 Producer accuracy for the woody class by classification method and classifier across six sites.

Points represent individual site-level accuracies for each combination of method (pixel-based vs. object-based) and classifier (SVM vs. RT). The black boxes represent the mean accuracy for each classification trial (Method/Class combination), with error bars representing the 95% confidence intervals. Object-based methods (Group b - Trials 3 and 4) showed higher accuracies compared to pixel-based methods (Group a - Trials 1 and 2). Four Mile consistently exhibited the lowest producer accuracies in the woody class across all trials, suggesting site-specific challenges influenced the classification performance.

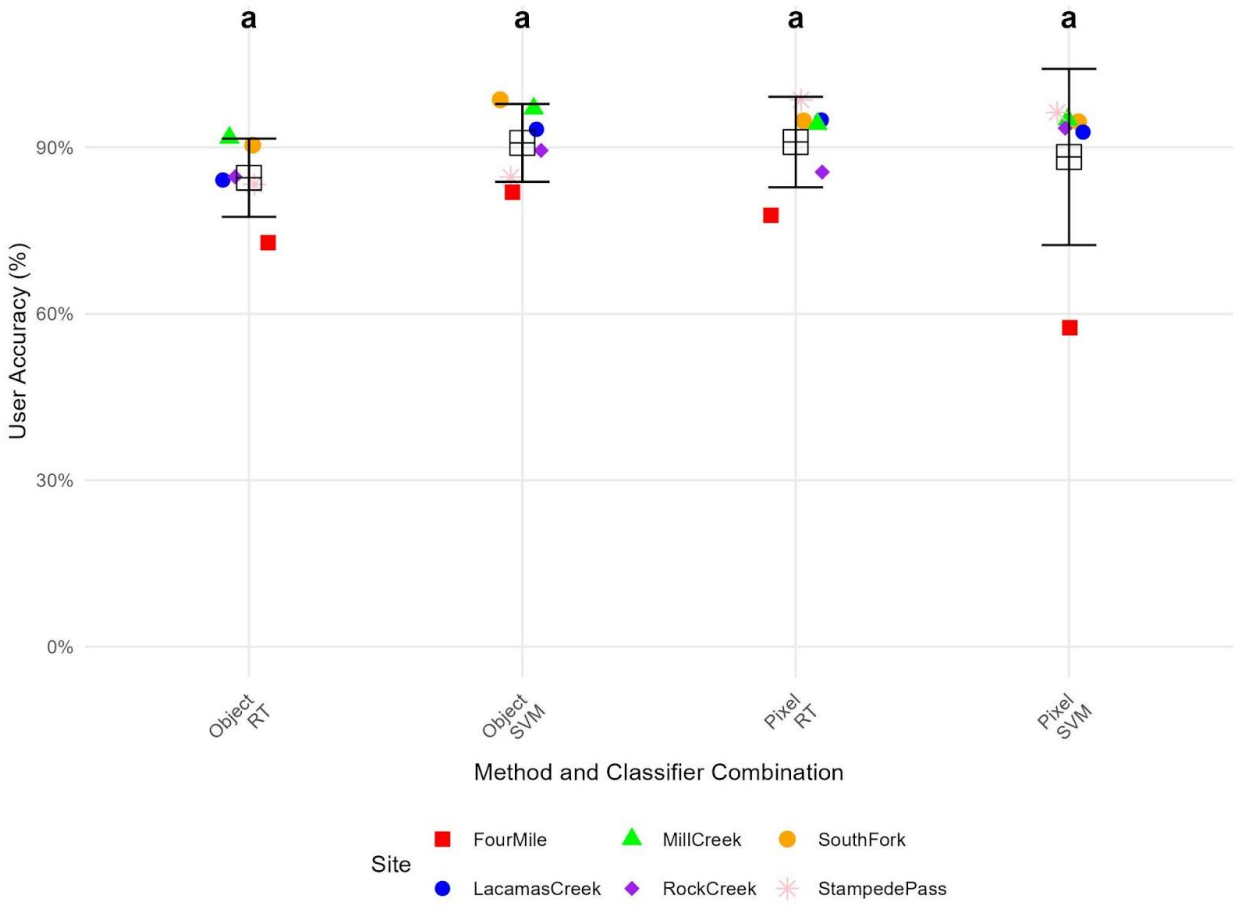

Figure 2.15 User accuracy for the woody class by classification method and classifier across six sites.

Points represent individual site-level accuracies for each combination of method (pixel-based vs. object-based) and classifier (SVM vs. RT). The black boxes denote the mean accuracy for each classification trial (Method/Class combination), with error bars representing the 95% confidence intervals. There were no significant differences in user accuracy between Trials 1-4.

2.3.5 Summary of classification results

Overall, the object-based classification, Trials 3 and 4, significantly outperformed the pixel-based classification, Trials 1 and 2, across all mitigation sites. Additionally, object-based classifications incorporating vertical data, Trials 5 and 6, outperformed those without vertical data, Trials 3 and 4, at Four Mile and Rock Creek. Producer and user accuracy achieved a minimum accuracy of 85% at all sites except Four Mile. Similarly, overall accuracy reached a minimum of 85% at all sites except Four Mile and Rock Creek. The classification method and classifier did not have a significant effect on user accuracy for the woody class. Additionally, the choice of classifier, SVM or RT, did not significantly influence overall accuracy, producer accuracy, or user accuracy within either pixel-based or object-based approaches.

2.4 Discussion

2.4.1 Evaluation of pixel and object-based classification methods (trials 1 - 4)

Across all trials, the woody class consistently achieved a higher user and producer accuracy than the non-target vegetation classes (green and brown herbaceous). While overall classification accuracy was lower than that of the producer and user accuracy for the woody class, this discrepancy was primarily driven by variability in the accuracy of non-target classes, particularly the herbaceous class. A key challenge in classifying the woody class was its spectral and segment similarity to the green herbaceous class, as both appear as green pixels in imagery. This spectral overlap led to misclassification, increasing producer and user errors in both classes. In contrast, the woody class remained sufficiently distinct from brown herbaceous, bare ground, shadow, and large woody debris classes due to differences in color and segment shape. Consequently, the spectral limitations of using only three-band imagery had a greater impact on distinguishing

between green herbaceous and woody than on differentiating other non-target classes. This was supported by the lower accuracy results in the pixel-based trials, which only incorporated RGB values to separate the classes.

For instance, in Trial 1 at Four Mile, 58% of green herbaceous reference samples were correctly classified, but a substantial portion, 84 out of 105 woody samples, were misclassified in the herbaceous class (Table 2.11). This led to high omission errors (false negatives) for the woody class and high commission errors (false positives) in the herbaceous class. These errors reduced the overall reliability of the classification, resulting in a low overall accuracy (OA). Trial 3 improved the classification accuracy by using segmented objects and RGB values, allowing the model to better distinguish vegetation types based on the segment characteristics. The more distinct shrubby growth habit of woody vegetation compared to herbaceous vegetation helped refine the classification, but there was still overlap between the two classes. As a result, the number of omitted woody reference samples decreased from 84 in Trial 1 to 42 in Trial 3 (Table 2.12). This improvement was reflected in the higher producer accuracy (PA) for woody vegetation (63% in Trial 3 vs. 19% in Trial 1) and increased overall accuracy to 73%. However, confusion between the woody and green herbaceous class remained an issue, as many of the omitted woody samples ended up in the herbaceous class, impacting the user accuracy for the green herbaceous class. Despite incorporating segments in the classification, the two classes remained similar because much of the site's woody vegetation was stunted and stressed from prolonged inundation. While the woody vegetation along the site's edges was robust (Appendix B, Figures B.5 and B.7), in the center, it structurally resembled emergent wetland vegetation, which led to similar segment characteristics between the two classes. In summary, Trial 3 showed some improvements, but a substantial overlap between the classes persisted.

Table 2.11 Confusion matrix for Trial 1 (pixel-based SVM) at Four Mile.

Bold values indicate correctly classified samples per class. Overall accuracy (OA), producer accuracy (PA), and user accuracy (UA) for the woody class are ***highlighted*** and shown as percentages.

Classified Data	Woody	Bare	Brown Herbs	Green Herbs	Shadows	Stressed	Samples	***UA***
Woody	**23**	0	8	9	0	0	40	***56%***
Bare	0	**161**	23	0	0	0	184	88%
Brown Herbs	13	13	**88**	6	0	0	120	73%
Green Herbs	84	0	0	**21**	0	0	105	20%
Shadows	2	0	0	0	**9**	0	11	82%
Stressed	1	18	20	0	1	**10**	50	20%
Samples	123	192	139	36	10	10	510	
PA	***19%***	84%	63%	58%	90%	100%		
OA								***61%***
Kappa								0.49

Table 2.12 Confusion matrix for Trial 3 (pixel-based SVM) at Four Mile.

Bold values indicate correctly classified samples per class. Overall accuracy (OA), producer accuracy (PA), and user accuracy (UA) for the woody class are ***highlighted*** and shown as percentages.

Classified Data	Woody	Bare	Brown Herbs	Green Herbs	Shadows	Stressed	Samples	***UA***
Woody	77	0	3	14	0	0	94	***82%***
Bare	1	154	25	0	0	0	180	86%
Brown Herbs	3	38	103	3	0	3	150	69%
Green Herbs	42	0	0	19	0	0	61	31%
Shadows	0	0	0	0	10	0	10	100%
Stressed	0	0	8	0	0	7	15	46%
Samples	123	192	139	36	10	10	510	
PA	***63%***	80%	74%	53%	100%	70%		
OA								***73%***
Kappa								0.63

Appendix A contains all confusion matrices for all trials across all sites, providing a detailed breakdown of classification performance for all.

In contrast, at all other sites, Trials 3 and 4 improved classification results enough to meet the target accuracy of 85%. For example, at Stampede Pass (Figure 2.16), the presence of well-established woody plantings, with the highest cover of any class, reduced overall confusion with the herbaceous classes. This allowed the object-based classification approach to perform well, highlighting the benefits of incorporating spatial features. Misclassifications still occurred between green herbaceous and woody classes, but at a lower amount than at sites with struggling woody vegetation, such as Four Mile and Rock Creek.

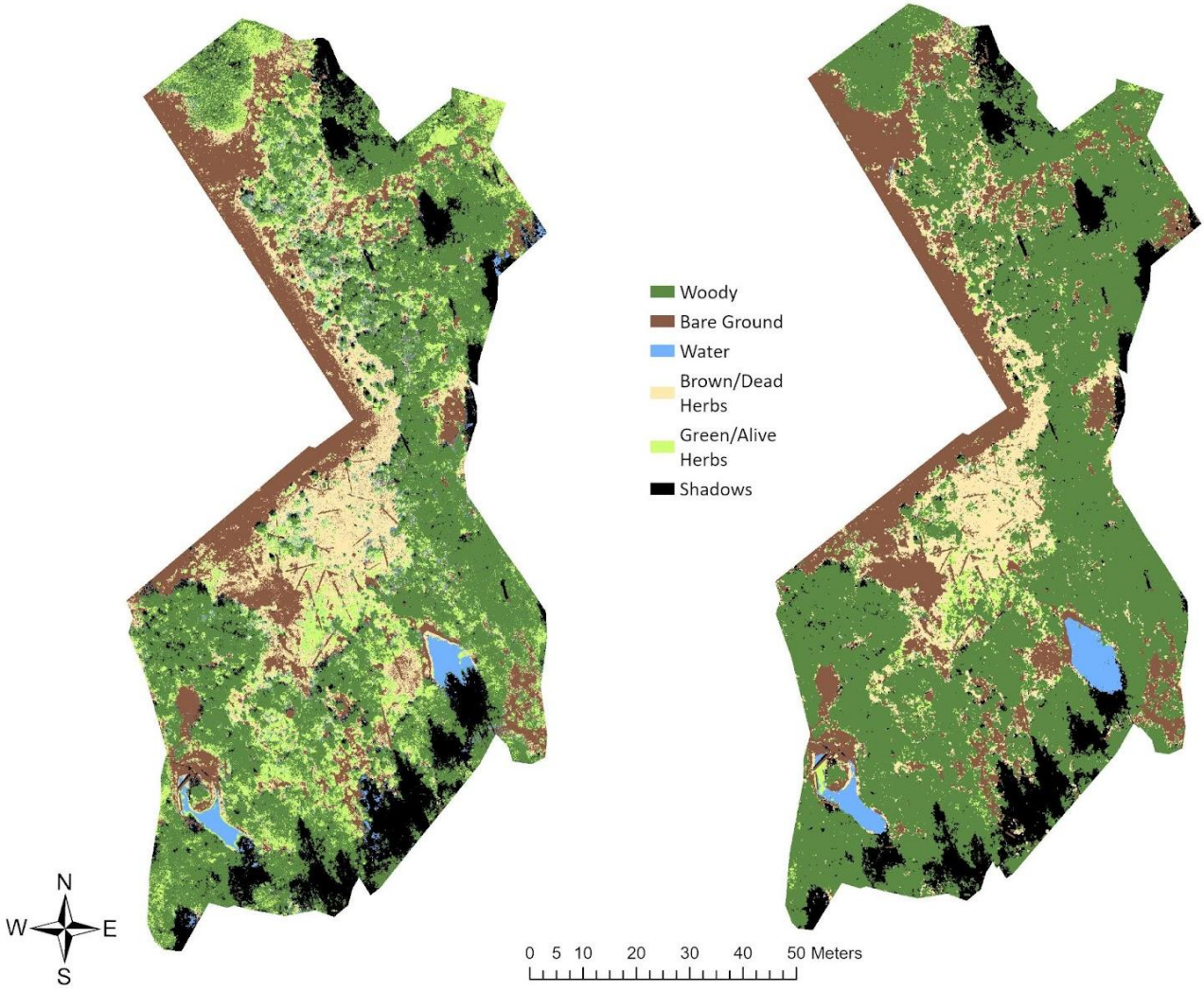

Figure 2.16 Drone-derived land cover classification maps of the Stampede Pass mitigation site using pixel-based (left) and object-based (right) methods with an SVM classifier.

2.4.2 Evaluation of DSM (trials 5 and 6) at Rock Creek and Four Mile

Integrating rDSM values at Rock Creek and Four Mile improved overall accuracy and producer accuracy for the woody class (Appendix B, Figures B.6 and B.7). At Rock Creek, the agreement map between Trial 3 and Trial 5 (Figure 2.11) showed strong agreement in the wetland portion of the site, while the upland buffer exhibited greater disagreement, highlighted by red pixels. This suggests that areas in the buffer with taller, more robust woody vegetation were easier to classify with rDSM integration, as the height difference between woody and herbaceous vegetation was more distinct. As a result, Trial 5 showed more classification changes in the buffer zone. However, a key limitation of drone-derived DSMs compared to lidar-based DSMs is that the rDSM relied only on minimum elevation values, which did not accurately capture terrain differences on site. In contrast, lidar-based nDSMs use digital terrain models (DTMs) to normalize vegetation height more precisely, as lidar can penetrate the canopy and create a DTM. However, lidar sensors for drones typically exceed agency budgets and were not feasible for this study. While the rDSM improved woody classification in flat areas with clear height differences, it introduced errors in sloped areas, such as the transition between wetland and upland zones at Rock Creek, where terrain variations were not effectively accounted for.

2.4.3 Fixed-effects ANOVA model

The ANOVA models highlighted the critical role of the classification method in achieving accurate results. Pixel-based methods were significantly less accurate than object-based methods, possibly because they relied solely on spectral values. With only three spectral bands, pixel-based classification struggled to differentiate between classes. These limitations led to significantly lower overall and producer accuracies in Trials 1 and 2. Object-based methods outperformed pixel-based

methods for classifying woody vegetation, primarily due to their ability to capture spatial patterns of mature woody plants that extended across multiple pixels. This advantage was particularly evident at South Fork Newaukum and Lacamas Creek, where dense shrub plantings created a clear separation between woody and herbaceous vegetation. Since these sites lacked active emergent wetland vegetation planting, the spectral and structural distinction between classes was more pronounced. Similarly, object-based methods performed better at Mill Creek Complex South and Stampede Pass, two mature sites with wetland components and emergent vegetation. At these sites, segmentation effectively captured the distinct growth habits of the two classes, leading to higher classification accuracy.

Within each method, the choice of classifier (SVM vs. RT) did not significantly impact overall or producer accuracy, suggesting that classifier selection was less critical than the classification approach itself. However, there were performance differences across sites. At the riparian sites (Lacamas Creek and South Fork Newaukum), user accuracy was approximately 10% lower in Trial 4 (RT) compared to Trial 3 (SVM). Although this difference was not statistically significant, it suggested that the RT classifier exhibited higher commission errors, with more pixels incorrectly assigned to the woody class from other vegetation types. At these sites, the distinction between woody and herbaceous vegetation appeared relatively clear, making the lower user accuracy in Trial 4 unexpected. One possible explanation is that woody vegetation was the largest class, and RT, as a frequency-based classifier, may have overclassified uncertain pixels as woody vegetation, leading to more false positives. Additionally, the imbalanced training data, which was proportional to class representation in an image, may have caused RT to overrepresent the larger class. Since RT builds decision trees using majority-rule voting, it likely assigned ambiguous pixels to the dominant class, whereas SVM, which optimizes a decision boundary, may have been

better suited for distinguishing classes where some spectral and structural separation already existed.

2.4.4 Future research and implications for monitoring

These findings indicated that complex site conditions, such as those at Four Mile and Rock Creek, posed significant mapping challenges when using limited spectral resolution and vertical data from a drone. It is important to note that most natural resource agencies rely on affordable equipment, similar to the tools I used in this study. As drone-based lidar and multispectral sensors become more affordable and are adopted by natural resource agencies, we can expect improvements in both accuracy and the classification workflow, allowing for better mapping at more variable sites. The growth of artificial intelligence also allows user-friendly approaches and workflows to handle complicated data, especially through ArcGIS Pro. While all classification methods have limitations, understanding these constraints is crucial for assessing their suitability for site monitoring and performance standard evaluation. In this chapter, the results were averaged by site rather than analyzed at the vegetation zone level, though performance standards are typically assessed at that finer scale. In the next chapter, I evaluated whether the classification raster provided woody cover estimates comparable to ground-based methods at two scales, further assessing their suitability for regulatory monitoring.

References

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). *A land use and land cover classification system for use with remote sensor data* (U.S. Geological Survey Professional Paper 964). U.S. Government Printing Office. https://pubs.usgs.gov/pp/0964/report.pdf

Berhane, T. M., Lane, C. R., Wu, Q., Anenkhonov, O. A., Chepinoga, V. V., Autrey, B. C., & Liu, H. (2018). Comparing pixel- and object-based approaches in effectively classifying wetland-dominated landscapes. *Remote Sensing, 10*(1), 46. https://doi.org/10.3390/rs10010046

Bottou, L., & Vapnik, V. (1992). Local learning algorithms. *Neural Computation*, 4(6), 888–900.

Contreras, J., Sickert, S., & Denzler, J. (2020). Region-based edge convolutions with geometric attributes for the semantic segmentation of large-scale 3-D point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 2598–2609. https://doi.org/10.1109/JSTARS.2020.2998037

Daubenmire, R. F. (1959). A canopy-coverage method of vegetational analysis. *Northwest Science, 33*, 43–64.

Dronova, I., Kislik, C., Dinh, Z., & Kelly, M. (2021). A review of unoccupied aerial vehicle use in wetland applications: Emerging opportunities in approach, technology, and data. *Drones, 5*(2), 45. https://doi.org/10.3390/drones5020045

Durgan, S. D. (2020). *Evaluating unmanned aircraft system photogrammetry for coastal Florida Everglades restoration and management* (Unpublished dissertation). Florida Atlantic University.

Esri. (2023). *ArcGIS Pro* (Version 3.3.0) [Software]. Environmental Systems Research Institute. https://www.esri.com/en-us/arcgis/products/arcgis-pro

Esri. (n.d.a). *Segmentation* (ArcGIS Pro 3.3 documentation). Environmental Systems Research Institute. Retrieved February 28, 2025, from https://pro.arcgis.com/en/pro-app/3.3/help/analysis/image-analyst/segmentation.htm

Esri. (n.d.b). *The Image Classification Wizard* (ArcGIS Pro 3.3 documentation). Environmental Systems Research Institute. Retrieved February 28, 2025, from https://pro.arcgis.com/en/pro-app/3.3/help/analysis/image-analyst/the-image-classification-wizard.htm

Esri. (n.d.c). *Train support vector machine classifier* (ArcGIS Pro documentation). Environmental Systems Research Institute. Retrieved February 28, 2025, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/train-support-vector-machine-classifier.htm

Esri. (n.d.d). *Train Random Trees classifier* (ArcGIS Pro documentation). Environmental Systems Research Institute. Retrieved February 28, 2025, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/train-random-trees-classifier.htm

Esri. (n.d.e). *Accuracy assessment* (ArcGIS Pro 3.3 documentation). Environmental Systems Research Institute. Retrieved February 28, 2025, from https://pro.arcgis.com/en/pro-app/3.3/help/analysis/image-analyst/accuracy-assessment.htm

Evans, A. D., Gardner, K. H., Greenwood, S., & Still, B. (2022). UAV and structure-from-motion photogrammetry enhance river restoration monitoring: A dam removal study. *Drones, 6*(5), 100. https://doi.org/10.3390/drones6050100

Gómez-Sapiens, M., Schlatter, K. J., Meléndez, Á., Hernández-López, D., Salazar, H., Kendy, E., & Flessa, K. W. (2021). Improving the efficiency and accuracy of evaluating aridland riparian habitat restoration using unmanned aerial vehicles. *Remote Sensing in Ecology and Conservation, 7*(3), 488–503. https://doi.org/10.1002/rse2.204

Halabisky, M., Moskal, L. M., & Hall, S. A. (2011). Object-based classification of semi-arid wetlands. *Journal of Applied Remote Sensing, 5*(1), 053511-053511. https://doi.org/10.1117/1.3563569

Halabisky, M., Miller, D., Stewart, A. J., Yahnke, A., Lorigan, D., Brasel, T., & Moskal, L. M. (2023). The wetland intrinsic potential tool: Mapping wetland intrinsic potential through machine learning of multi-scale remote sensing proxies of wetland indicators. *Hydrology and Earth System Sciences, 27*(20), 3687–3699. https://doi.org/10.5194/hess-27-3687-2023

Halpern, S. (2020). *Wetland mapping and restoration decision making using remote sensing and spatial analysis: A case study at the Kawainui Marsh* (Doctoral dissertation, University of Southern California). https://spatial.usc.edu/wp-content/uploads/formidable/12/Sarah-Halpern-1.pdf

Haskins, J., Endris, C., Thomsen, A. S., Gerbl, F., Fountain, M. C., & Wasson, K. (2021). UAV to inform restoration: A case study from a California tidal marsh. *Frontiers in Environmental Science, 9*, 642906. https://doi.org/10.3389/fenvs.2021.642906

Husson, E., Reese, H., & Ecke, F. (2017). Combining spectral data and a DSM from UAS-images for improved classification of non-submerged aquatic vegetation. *Remote Sensing, 9*(3), 247. https://doi.org/10.3390/rs9030247

Intergovernmental Panel on Climate Change. (2019). *Chapter 3: Consistent Representation of Lands*. In *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Retrieved from https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch03_Land%20Representation.pdf

Jones, H. G., & Vaughan, R. A. (2010). *Remote sensing of vegetation: Principles, techniques, and applications.* Oxford University Press, USA.

Mora, O. E., Suleiman, A., Chen, J., Pluta, D., Okubo, M. H., & Josenhans, R. (2019). Comparing sUAS photogrammetrically-derived point clouds with GNSS measurements and terrestrial laser scanning for topographic mapping. *Drones, 3*(3), 64. https://doi.org/10.3390/drones3030064

Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., & Wilhelm, V. L. (2017). Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery. *European Journal of Remote Sensing, 50*(1), 564–576. https://doi.org/10.1080/22797254.2017.137360

Pix4D. (n.d.). *Pix4DMapper Pro* (Version 4.1) [Software]. Pix4D SA. Retrieved February 28, 2025, from https://www.pix4d.com/product/pix4dmapper

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 4.1.2) [Software]. R Foundation for Statistical Computing. https://www.r-project.org

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected*

Topics in Applied Earth Observations and Remote Sensing, 13, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724

Sibaruddin, H. I., Shafri, H. Z. M., Pradhan, B., & Haron, N. A. (2018, June). Comparison of pixel-based and object-based image classification techniques in extracting information from UAV imagery data. In *IOP Conference Series: Earth and Environmental Science* (Vol. 169, No. 1, p. 012098). IOP Publishing. https://doi.org/10.1088/1755-1315/169/1/012098

U.S. Geological Survey. (2023). *Annual National Land Cover Database*. U.S. Department of the Interior. Retrieved from https://www.usgs.gov/centers/eros/science/annual-national-land-cover-database

Washington Department of Transportation, Northwest Region. (2020). *I-5/Northbound on-ramp at Bakerview Road, milepost 257.65 to 258.50: Four Mile Creek mitigation site* (Wetland Mitigation Report).

Washington Department of Transportation, South Central Region. (2016). *I-90 Snoqualmie Pass East Stampede Pass Stockpile Site: Advance mitigation plan*.

Washington Department of Transportation, Southwest Region. (2012). *SR 502 corridor widening: Final critical areas mitigation plan*.

Washington Department of Transportation, Southwest Region. (2013). *SR 6 Rock Creek Bridge replacements (Bridge 6/102, 6/103), MP 25.9-26.9: Final critical areas mitigation plan*.

Washington Department of Transportation, Southwest Region. (2016). *SR-508 South Fork Newaukum River Bridge replacement: Wetland delineation and critical area mitigation memorandum*.

Washington Department of Transportation, Southwest Region. (2021). *SR-506 Lacamas Creek Bridge replacement: Wetland delineation and critical area mitigation memorandum*.

Whiteside, T., & Ahmad, W. (2005, September). A comparison of object-oriented and pixel-based classification methods for mapping land cover in northern Australia. In *Proceedings of SSC2005 Spatial intelligence, innovation and praxis: The national biennial conference of the Spatial Sciences Institute* (pp. 1225-1231).

Yousefi, S., Mirzaee, S., Tazeh, M., Pourghasemi, H., & Karimi, H. (2015). Comparison of different algorithms for land use mapping in dry climate using satellite images: A case study of the central regions of Iran. *Desert, 20*(1), 1-10. 10.22059/jdesert.2015.54077

Zheng, J.-Y., Hao, Y.-Y., Wang, Y.-C., Zhou, S.-Q., Wu, W.-B., Yuan, Q., Gao, Y., Guo, H.-Q., Cai, X.-X., & Zhao, B. (2022). Coastal wetland vegetation classification using pixel-based, object-based, and deep learning methods based on RGB-UAV. *Land, 11*(11), 2039. https://doi.org/10.3390/land11112039

Zhou R, Yang C, Li E, Cai X, Yang J, Xia Y. Object-Based Wetland Vegetation Classification Using Multi-Feature Selection of Unoccupied Aerial Vehicle RGB Imagery. *Remote Sensing*. 2021; 13(23):4910. https://doi.org/10.3390/rs13234910

Chapter 3: Comparison of drone-derived and field-based woody vegetation cover estimates

3.1 Introduction

In wetlands, vegetation plays a crucial role by acting as a filter to trap sediments, slowing the distribution of floodwaters, and reducing the velocity of surface runoff (Washington State Department of Ecology [Ecology], 2005). Vegetation cover, defined as the percentage of ground area covered by green vegetation (Barnas et al., 2019), is commonly used as a proxy to assess the success of mitigation efforts. At compensatory mitigation sites, vegetation cover is commonly used as a key performance standard metric to ensure successful mitigation and regulatory compliance. These performance standards are typically categorized by community types, such as woody vegetation for tree and shrub habitats or emergent vegetation for wetlands dominated by herbaceous species such as grasses, sedges, and rushes.

Traditional methods to assess vegetative performance standards rely on field surveys, using quadrats, line-intercept, point-line, or point-frame techniques (WSDOT, 2021). Among these, the line-intercept technique is commonly used to estimate the cover of woody species (Bonham, 1989; Coulloudon et al., 1999; WSDOT, 2021). This method involves randomly positioning sample units, represented by line segments used for sampling, along transects (Figure 1) and recording where woody vegetation intercepts the line segment (WSDOT, 2021). This process requires navigating the site to establish transects using monitoring tapes and systematically sampling along each transect. While the technique is straightforward, conducting surveys over large areas can be resource-intensive, time-consuming, and sometimes impractical due to site hazards commonly found in aquatic environments, such as flooding or deep channels. When resources such as personnel are limited, these challenges are further compounded by not having enough staff to adequately sample the entire site or deploy enough sample units to draw meaningful inferences

about vegetation across the site. In some cases, while monitoring may still be completed, it often requires substantially more time and staff, making monitoring less efficient.

Drones offer an alternative way to monitor mitigation sites, particularly when traditional field methods face constraints. Using drone technology can be beneficial when the demand for monitoring exceeds available staff—an ongoing challenge for agencies that often operate with limited capacity. Drones can capture images of large areas in a single automated flight, and these images can be used to classify vegetation groups. This information can then be used to estimate cover, supporting applications in regulatory monitoring. Moreover, their ability to capture imagery in areas that are difficult to access on the ground highlights their versatility and value in monitoring efforts.

Some studies have demonstrated the efficiency of drones compared to ground-based methods in monitoring applications. Gomez-Sapiens et al. (2021) reported that drone surveys were more cost and time efficient than field surveys, after accounting for the initial drone cost and the time required to learn new drone methods. They observed moderate but significant positive correlations between drone and ground-derived data for grouped species of riparian vegetation. Similarly, Barnas et al. (2019) found large time savings with drones during data collection but larger time investments during the imagery processing steps. In their study on habitat destruction by snow geese, they compared drone imagery with ground-based methods to estimate cover in barren, shrub, and non-shrub categories. Drone estimates showed higher agreement with ground-based data for barren and shrub categories but poor agreement for non-shrub vegetation.

3.1.1 Objectives, questions, and hypotheses

Building on the proven efficiency of drones in monitoring applications, this study explored their potential for regulatory monitoring, an area where established guidelines for these tools are absent. This study aimed to evaluate whether drone surveying is comparable to field surveys by comparing woody cover measurements between drone-derived and field-derived data from sample units (hereafter referred to as plots), sampled by the line-intercept method. To ensure generality of the results, I made comparisons at six mitigation sites, each containing one or two zones.

This chapter addressed the following questions:

How do woody cover estimated from drone imagery compare to field estimates at the zone and plot scales?

a. Do different classification trials (Trial 3 or Trial 4) yield woody cover estimates that are more comparable to field estimates?
b. How do zone and age class influence woody cover estimates derived from drone imagery? In other words, do drone and field methods yield different cover estimates based on the zone or age class they're in?

I hypothesized that drone imagery would produce woody cover estimates comparable to those from the line-intercept method at both the zone and plot scales, indicating that both methods similarly accounted for cover. At the zone scale, I represented drone-derived vegetation cover as the percentage of woody cover across the entire zone and field-derived data as the average woody cover of all sample units. At the plot scale, I expressed drone-derived vegetation cover as the percentage of woody cover within each plot and field-derived data as the raw data collected in each plot. Further, I hypothesized that there would be a strong correlation with traditional field

measurements at the zone and plot scale, particularly for sites that achieved classification accuracies greater than 85% in Chapter 2: Lacamas Creek, South Fork Newaukum, Mill Creek Complex South, and Stampede Pass. While the first hypothesis on comparability suggests that the two methods yield similar results on average, the second hypothesis on correlation examines whether variations in one method consistently align with variations in the other.

3.2 Methods

3.2.1 Study Sites

I conducted this study at six mitigation sites managed by the Washington State Department of Transportation (WSDOT): Four Mile, South Fork Newaukum, Lacamas Creek, Rock Creek, Stampede Pass, and Mill Creek Complex South. Three sites—Four Mile, Rock Creek, and Stampede Pass—were each divided into wetland and buffer zones. Mill Creek Complex South consisted of a wetland zone, while Lacamas Creek and South Fork Newaukum each consisted of a riparian zone. In total, this study encompassed nine unique site-zone combinations.

These sites were selected based on three major criteria: Cowardin/vegetation zones; age; and performance standards. The sites consisted of upland, wetland, and riparian vegetation zones. Three sites were in year 3 of mitigation at the time of data collection, while the other sites ranged in age from year 7 to year 10. Performance standards varied by site and zone and are listed in Table 3.1.

3.2.2 Vegetation monitoring

The WSDOT Wetland Program conducted field surveys in July and August 2023 using the line-intercept method on temporary plots to collect cover data for woody species (Table 3.1). WSDOT developed the sample design in ArcGIS Pro to determine the placement of the baselines

and transects in the field. WSDOT randomized the initial position of each transect along the baseline, after which transects were systematically spaced within each zone. Along each transect, WSDOT also randomized the starting point of the plot(s). In some cases, WSDOT positioned plots at both the start and end of a transect due to their *om method* of sampling. This approach allows a plot's randomly determined starting point to be near the beginning or end of a transect. When the plot extended beyond the boundary of the zone or site, it wrapped around to the beginning of the transect, continuing until its predetermined length was achieved.

Each site had a unique sample design: the initial transect position and spacing between transects varied by site but were consistent within a site. Along each transect, the position of the plot was randomized to determine where monitoring would occur. For sites with two zones (Stampede Pass, Rock Creek, and Four Mile), the same transects were used for both zones; however, plot locations along the transects were independently randomized for each zone. The length of the plot and spacing between plots were consistent within each zone.

Table 3.1 Summary of site and sample design characteristics.

The plot length, plot interval, and distance between transects are expressed in meters.

Site Name	Date of fieldwork (in 2023)	Age (in *2023*)	Zone	Number of plots	Plot Length	Plot Interval	Distance between transect
Four Mile	July 31	Young (3)	Wetland	14	10	10	10
Four Mile	July 31	Young (3)	Buffer	17	10	10	10
South Fork Newaukum	July 14	Young (3)	Riparian	10	20	25	25
Lacamas Creek	July 3 - 5	Young (3)	Riparian	19	20	20	10
Rock Creek	July 10 - 12	Mature (7)	Wetland	12	15	15	15
Rock Creek	July 10 - 12	Mature (7)	Buffer	16	15	15	15
Stampede Pass	August 18	Mature (7)	Wetland	15	10	20	10
Stampede Pass	August 16	Mature (7)	Buffer	10	5	10	10
Mill Creek Complex South	July 24	Mature (10)	Wetland	23	10	10	10

In the field, WSDOT used the sample design map in Collector and GPS to navigate to the start of each transect along the baseline. The GPS unit integrated into the Mesa tablets provided an accuracy range of approximately 2 to 5 meters (Juniper Systems, n.d.). Once field staff reached the transect's starting point, they randomized the plot's location and walked to the designated position along the transect tape. Within each plot, WSDOT applied the line-intercept method by

recording the start and stop locations where woody vegetation intersected the sampling tape. To meet performance standard requirements, WSDOT treated overlapping vegetation as a single segment to measure aerial cover rather than cumulative cover.

3.2.3 Vegetation cover from classification rasters

To calculate woody cover at the zone scale, I clipped the classification maps produced in Chapter 2 to the zone boundaries provided by WSDOT. To determine the area of woody vegetation in the zone, I multiplied the pixel count of each land cover class (i.e. woody, green/alive herbaceous, brown/dead herbaceous, bare, shadows, water, large woody debris (LWD), and stressed) by the pixel size. To calculate the cover of woody vegetation, I divided the area of woody vegetation by the total area (the sum of all land cover classes) and multiplied the result by 100 to express the cover as a percentage.

At the plot scale, I created a 0.5-meter buffer around each plot sampled in the field for woody vegetation. This buffer size, larger than the pixel size of 0.01–0.02 meters for most classification maps, accounted for minor GPS inaccuracies during field data collection and errors in reeling out the tape. Although the transects were assumed to be perpendicular to the baseline, sampling errors in the field occasionally led to slight deviations. The buffer around the plot also provided a defined area to measure woody cover. Measuring woody cover directly along a line feature in ArcGIS Pro was not feasible, making the buffered area a practical alternative. Although GPS accuracy in the field ranged from 2 to 5 meters, I kept the buffer width small to ensure the plot size remained comparable to the length of the field-sampled plot. This limitation was necessary because the results are sensitive to buffer width, with larger buffers introducing additional vegetation not sampled by the line-intercept method, potentially skewing the results.

To calculate the area of woody vegetation within the buffered plots, I used the Tabulate Area geoprocessing tool in ArcGIS Pro (Esri, n.d.). This tool computed the area of the woody vegetation class derived from the classification within the buffered plots. I calculated the percentage of woody vegetation by dividing the amount of woody vegetation in each plot by the total area in the plot. The total area of each plot was consistent across all plots within a given site-zone.

3.2.4 Statistical analysis

All statistical analyses were conducted using R statistical software (Version 4.1.2; R Core Team, 2021).

3.2.4.1 Zone scale analysis

I compared woody cover estimates from two drone-derived methods, Trial 3 (object-based SVM) and Trial 4 (object-based RT), to field-collected data across nine site-zone combinations. The analysis aimed to evaluate which trial aligned more closely with field data at the zone-scale, based on statistical differences in mean woody cover and correlation.

I tested the normality of the drone and field data using the Shapiro-Wilk test. Then, I conducted two paired *t*-tests to determine whether there were significant differences in mean woody cover between each drone method and the field method. Next, I calculated Spearman correlation coefficients for each drone trial to assess the strength and direction of the relationship between the drone and field data across all sites and zone combinations.

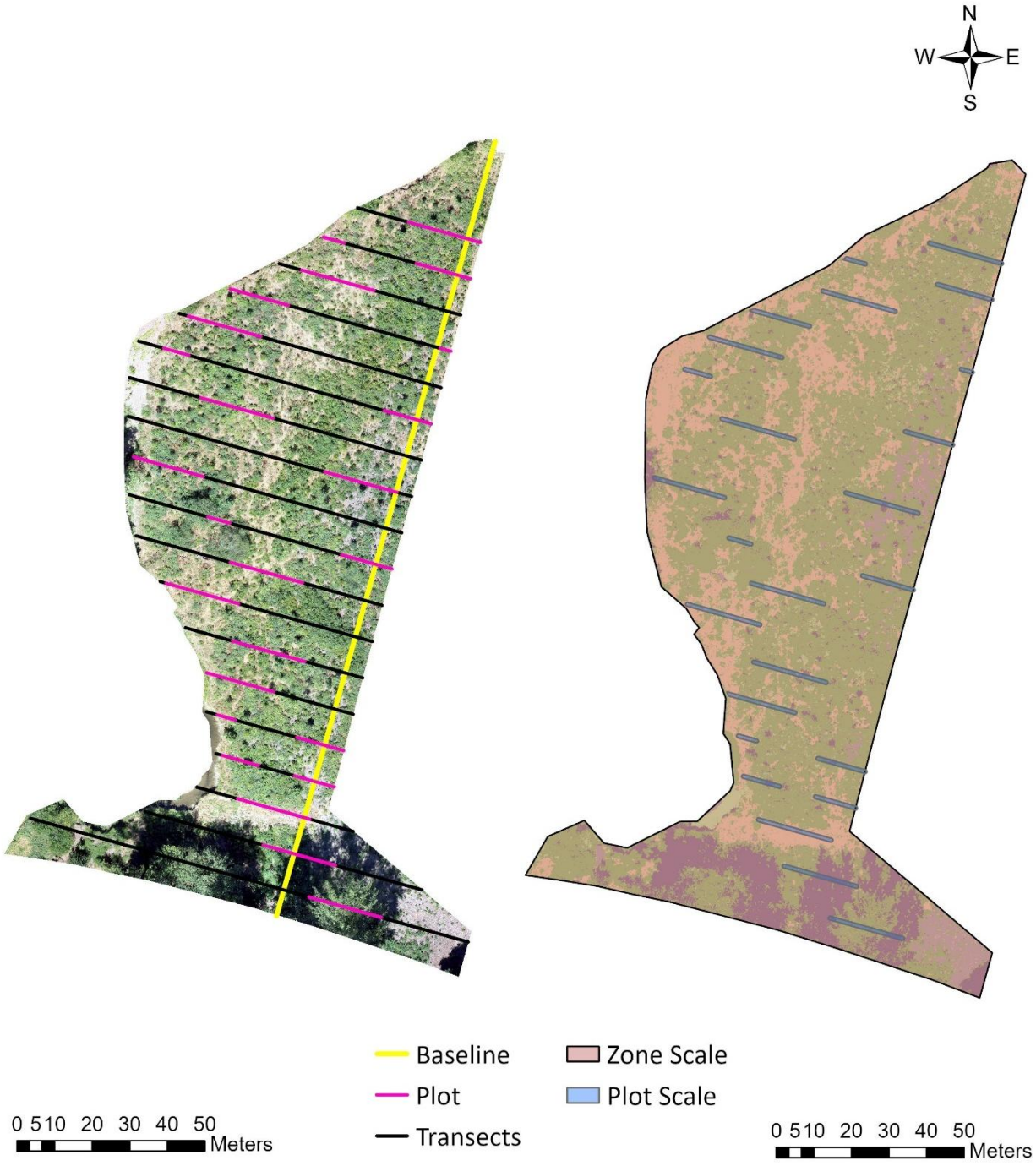

Figure 3.1 Left: Sample design at Lacamas Creek. Right: Zone and plot scales from a drone-derived classification raster.

Left: WSDOT's sample design includes a baseline (yellow), transects (black), and plots (pink). The average woody cover across all plots represents the zone scale, while raw data from individual plots represents the plot scale. **Right**: For drone-derived data, woody cover across the entire zone (red) represents the zone scale, while woody cover within buffered plots (blue) represents the plot scale.

3.2.4.2 Plot scale analysis

I applied a Generalized Linear Mixed Model (GLMM) using the glmmTMB package in R (Brooks et al., 2017) to evaluate the effects of zone (riparian, buffer, and wetland), age (young and mature), method (drone Trial 3, drone Trial 4, and field), and select interactions (zone × method and age × method) on vegetation cover. I fit a Beta GLMM with a logit link function, where the proportional cover data was modeled as a function of zone, age, and method, while accounting for random variation among sites and plots (site: plot_id):

$$cover\_proportion \sim zone * method + age * method + (1 \mid site : plot\_id)$$

I tested for model assumptions using DHARMa residual diagnostics, including dispersion tests and zero-inflation checks, given that the dataset contained zero values in some plots. To test the significance of the fixed effects, I conducted an Analysis of Deviance (Type III Wald Chi-Square tests) using the ANOVA function. When a term or interaction was statistically significant, I conducted post hoc pairwise comparisons using Tukey's Honest Significant Difference (HSD) test with the emmeans package (Lenth, 2024) to determine which methods differed within each zone and age group, as my primary interest was whether the methods differed. To assess the relationship between drone-based and field-based cover estimates, I performed Spearman's Rank Order correlation analyses to measure the strength and direction of the association between the two methods at each site-zone.

3.3 Results

3.3.1 Zone scale analysis: Comparison of woody cover estimates from drone-trial 3 and field methods

The results from the paired *t*-test indicated that the mean cover difference between Trial 3 and the line-intercept (field) method is marginally significant across all paired observations ($t(8)$

= -2.31, $p = 0.05$). On average, Trial 3 underestimated woody cover by 7.5% compared to the line-intercept method (Table 3.2). The largest underestimation occurred at Rock Creek in the buffer zone and Four Mile in the wetland zone, with differences of 29.8% and 11.8%, respectively.

The drone and field cover estimates were strongly and positively correlated ($r_s = 0.883$, $p = 0.003$, $n = 9$), suggesting that the two methods measured woody cover similarly, despite differences in their absolute values. Across all site-zone observations, field estimates of woody cover were consistently higher than drone estimates, except for the buffer zone at Stampede Pass (Figure 3.2). This pattern supports the underestimation observed in the paired *t*-test.

Table 3.2 Comparison of mean woody cover (%) in each site and zone as derived from drone Trial 3 and Field cover values.

The last column shows the difference in mean value. The **last row** represents the average for each method and the overall average difference between methods.

Site	**Zone**	**Trial 3**	**Field**	**Difference**
Rock Creek	Wetland	18.4	26.5	-8.1
Rock Creek	Buffer	40.0	69.8	-29.8
Four Mile	Buffer	33.5	38.0	-4.5
Four Mile	Wetland	31.9	43.7	-11.8
Stampede Pass	Wetland	64.3	75.0	-10.7
Stampede Pass	Buffer	30.1	24.8	5.3
South Fork Newaukum	Riparian	63.9	64.4	-0.6
Mill Creek Complex South	Wetland	25.9	29.8	-3.9
Lacamas Creek	Riparian	61.3	64.3	-3.0
Averages		**41.0**	**48.5**	**-7.5**

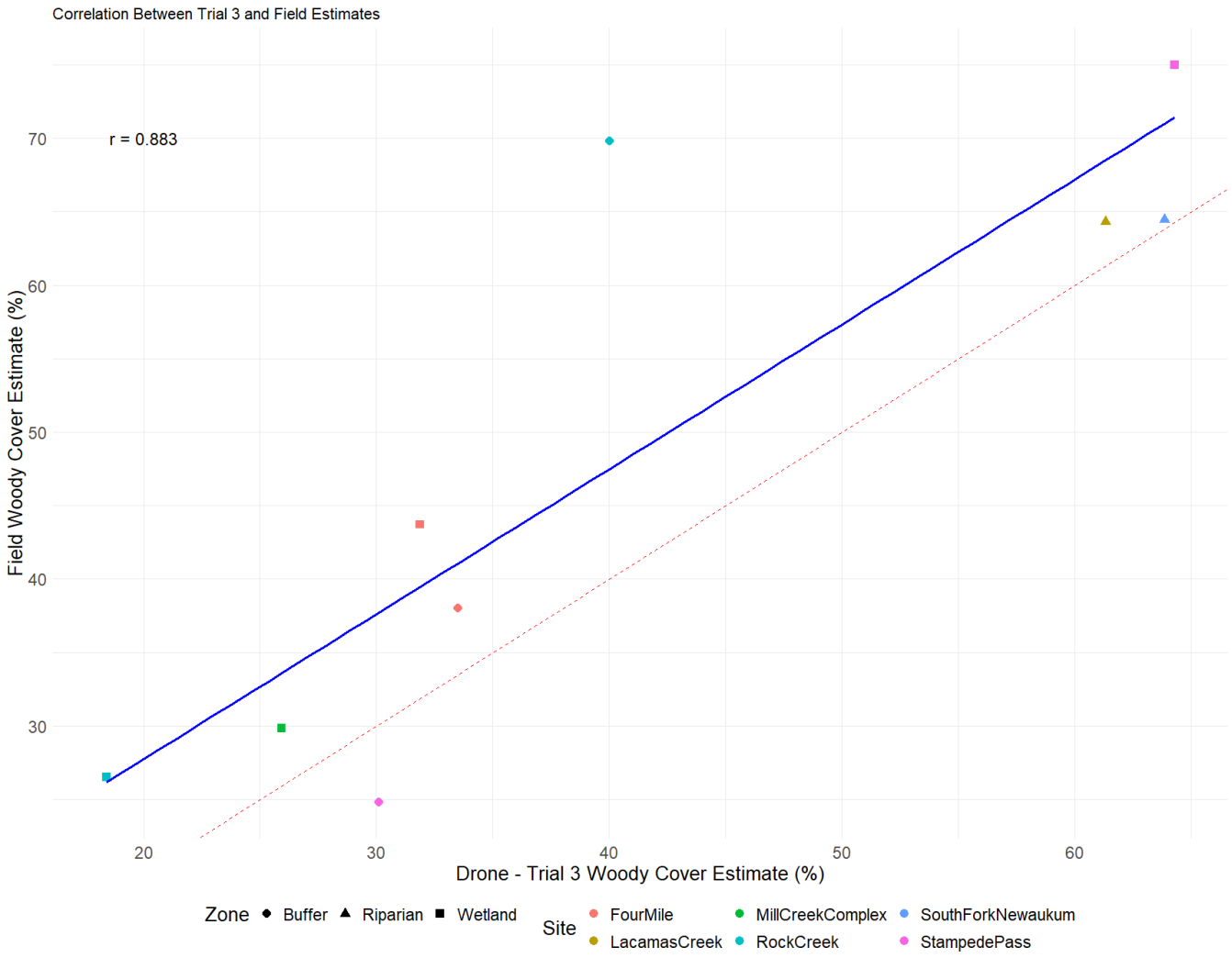

Figure 3.2 Comparison of woody cover estimates between the drone (Trial 3) and field (line-intercept) methods across all site-zone observations.

Each point represents a paired observation, differentiated by color for site and shape for zone. The blue line represents the line of best fit (correlation: $r_s = 0.883$, $p = 0.003$, $n = 9$), indicating a strong positive relationship between the two methods. The red dashed line represents the 1:1 relationship. Points above the 1:1 line were underestimated by the drone method, while points below the 1:1 line were overestimated.

3.3.2 Comparison of woody cover estimates from drone-trial 4 and field methods

The results from the paired *t*-test indicated that the difference in woody cover between Trial 4 and field methods was marginally significant across all paired observations ($t(8) = -2.06$, $p = 0.07$). On average, the drone method underestimated woody cover by 7.1% compared to the line-intercept method (Table 3.3). The largest underestimations occurred at Rock Creek in the buffer zone and Lacamas Creek in the riparian zone, with differences of 27.4% and 14.3%, respectively. The largest overestimation occurred at South Fork Newaukum, with a difference of 11.1%. The drone and field cover estimates were strongly positively correlated ($r_s = 0.9$, $p = 0.002$, $n = 9$), suggesting that the two methods measured woody cover similarly (Figure 3.3).

Table 3.3 Comparison of mean woody cover (%) in each site and zone as derived from drone Trial 4 and Field cover values.

The last column shows the difference in mean value. The **last row** represents the average for each method and the overall average difference between methods.

Site	**Zone**	**Trial 4**	**Field**	**Difference**
Rock Creek	Wetland	23.8	26.5	-2.7
Rock Creek	Buffer	42.5	69.8	-27.4
Four Mile	Buffer	33.4	38.0	-4.6
Four Mile	Wetland	32.4	43.7	-11.3
Stampede Pass	Wetland	66.0	75.0	-9.0
Stampede Pass	Buffer	21.3	24.8	-3.5
South Fork Newaukum	Riparian	75.6	64.4	11.1
Mill Creek Complex South	Wetland	27.1	29.8	-2.7
Lacamas Creek	Riparian	50.1	64.3	-14.3
Averages		**41.3**	**48.5**	**-7.1**

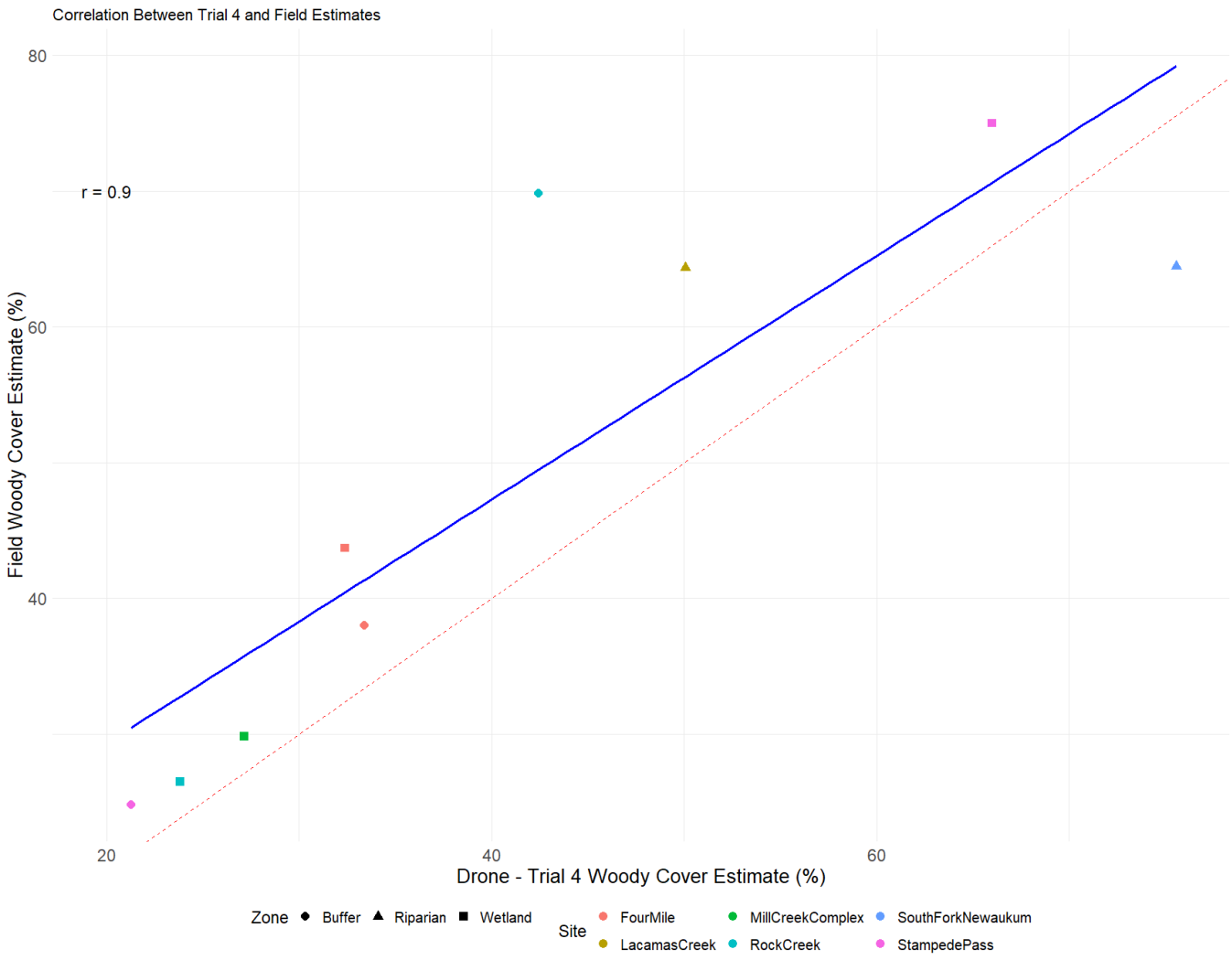

Figure 3.3 Comparison of woody cover estimates between drone (Trial 4) and field methods across all site-zone observations.

Each point represents a paired observation, differentiated by color for site and shape for zone. The blue line represents the line of best fit (correlation: $r_s = 0.9$, $p = 0.002$, $n = 9$), indicating a strong positive relationship. The red dashed line represents the 1:1 relationship, highlighting consistent underestimation by the drone method compared to field estimates.

3.3.3 Comparison of trial 3 and trial 4 with field data

Both Trial 3 and Trial 4 have strong, positive, and statistically significant correlations with the field data for woody cover estimates across all site-zone observations. Trial 4 had a slightly higher correlation coefficient than Trial 3 ($r_s = 0.9$ vs. 0.88). Despite this, results from the paired

t-test indicated that Trial 3's woody cover estimates were marginally different from the field data, with a mean difference of 7.5%. Similarly, Trial 4 had a mean difference of 7.1%, which was also marginally different. Despite Trial 3's slightly lower correlation coefficient and marginally higher mean difference, its trendline (Figure 3.4) is more parallel to the 1:1 line, suggesting that underestimation was more consistent across the range of values calculated for site-zones.

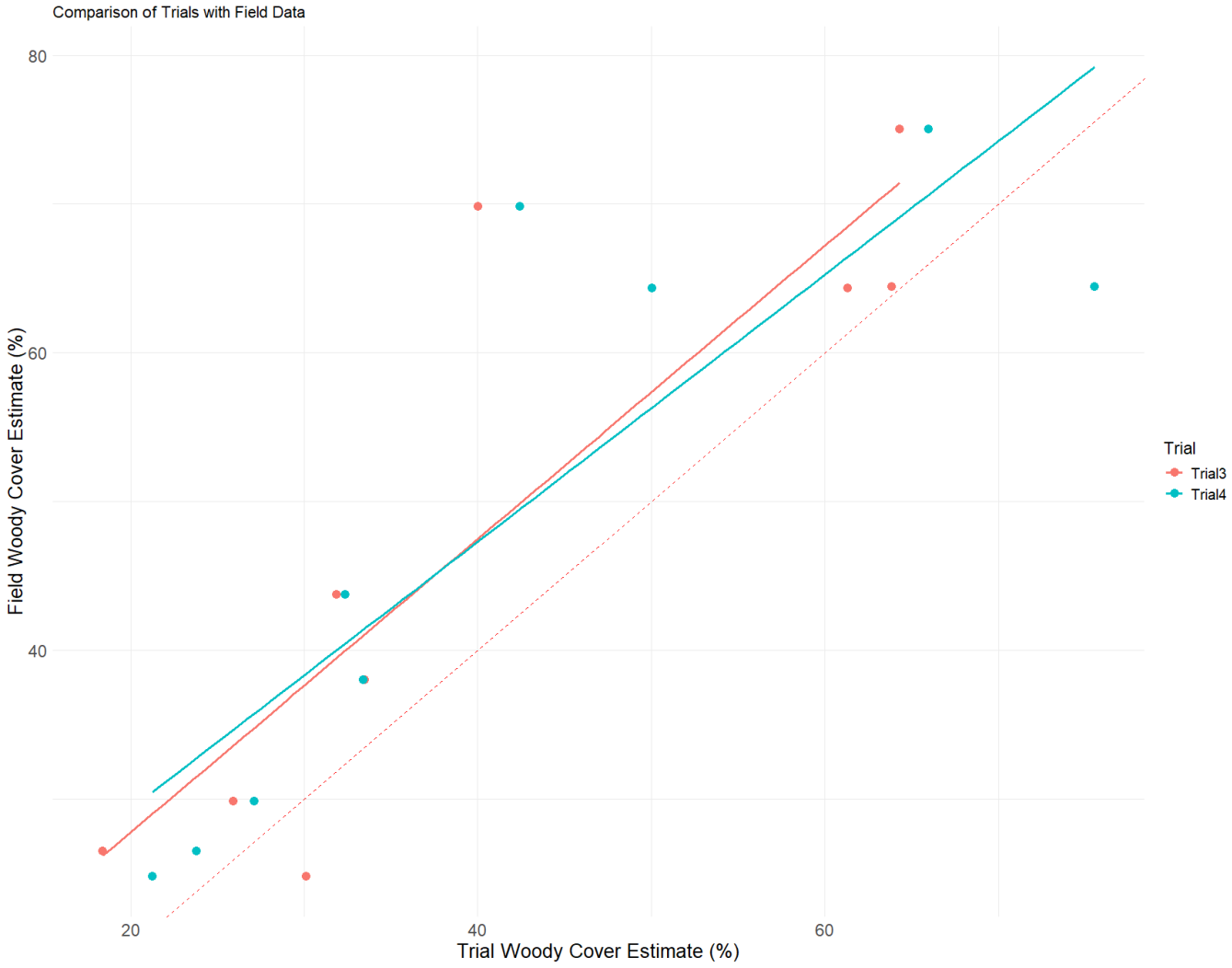

Figure 3.4 Scatterplot of woody cover estimates for Trial 3 and Trial 4 compared to Field data. The dashed red 1:1 line represents perfect agreement between the methods and Field data. The trendlines for both trials closely align with the 1:1 line, with Trial 3 appearing more parallel.

3.3.4 Plot scale analysis - Beta GLMM and Tukey's post hoc

Cover estimates were significantly influenced by method (Table 3.4), indicating that cover differed between drone and field methods. Additionally, there were significant interactions between zone × method and method × age, suggesting that the effect of method varied across zones and was influenced by age.

Table 3.4 Type III ANOVA results for the GLMM

Method and zone had significant main effects, and both interaction terms were also significant. Significance levels are denoted as follows: $***p < 0.001$, $**p < 0.01$, and $*p<0.05$.

Effect	χ^2 (Chi-square)	**df**	***p*-value**
(Intercept)	8.31	1	0.004**
zone	7.96	2	0.019*
method	15.78	2	<0.001***
age	0.005	1	0.941
Zone:method	14.42	4	0.006**
Method:age	9.80	2	0.007**

Pairwise comparisons using Tukey's HSD found no statistically significant differences between the three methods overall (Table 3.5). This suggests that while the ANOVA detected a significant effect of method, post hoc comparisons did not reveal significant differences between individual methods. However, pairwise comparisons of the zone × method interaction identified a significant difference between methods in the riparian zone (Table 3.6): field methods produced higher cover estimates than Drone Trial 3 and Drone Trial 4. Similarly, pairwise comparisons revealed a significant difference in mature sites with field methods producing higher cover

estimates than the drone-based methods (Table 3.7). These results indicate that method differences are stronger in the riparian zone and mature sites.

Table 3.5 Tukey's HSD pairwise comparisons of methods

Pairwise comparisons of methods using Tukey's HSD test adjusted for multiple comparisons. Odds ratios indicate the relative likelihood of differences between methods.

Comparison	Odds Ratio	SE	*z*-ratio	*p*-value
Drone Trial 3 vs. Drone Trial 4	1.01	0.09	0.07	0.997
Drone Trial 3 vs. Field	0.82	0.08	-1.97	0.120
Drone Trial 4 vs. Field	0.82	0.08	-2.05	0.101

Table 3.6 Tukey's HSD pairwise comparisons of methods by zone (i.e., zone × method interaction)

Pairwise comparisons of methods within each zone using Tukey's HSD test adjusted for multiple comparisons. Odds ratios indicate the relative likelihood of differences between methods. Significance levels are denoted as follows: $***p < 0.001$, $**p < 0.01$, and $*p<0.05$.

Comparison	Odds Ratio	SE	*z*-ratio	*p*-value
Riparian				
Drone Trial 3 vs. Drone Trial 4	1.08	0.25	0.34	0.94
Drone Trial 3 vs. Field	0.61	0.14	-2.15	0.08
Drone Trial 4 vs. Field	0.56	0.13	-2.50	0.03*
Buffer				
Drone Trial 3 vs. Drone Trial 4	1.02	0.18	0.09	1.00
Drone Trial 3 vs. Field	0.72	0.13	-1.78	0.18
Drone Trial 4 vs. Field	0.71	0.13	-1.88	0.14
Wetland				
Drone Trial 3 vs. Drone Trial 4	0.93	0.15	-0.47	0.89

Comparison	Odds Ratio	SE	*z*-ratio	*p*-value
Drone Trial 3 vs. Field	1.27	0.20	1.51	0.29
Drone Trial 4 vs. Field	1.36	0.21	1.99	0.12

Table 3.7 Tukey's HSD pairwise comparisons of methods by age (i.e., age × method interactions)

Pairwise comparisons of methods within each age using Tukey's HSD test adjusted for multiple comparisons. Odds ratios indicate the relative likelihood of differences between methods. Significance levels are denoted as follows: $***p < 0.001$, $**p < 0.01$, and $*p<0.05$.

Comparison	Odds Ratio	SE	*z*-ratio	*p*-value
Mature				
Drone Trial 3 vs. Drone Trial 4	0.94	0.15	-0.37	0.93
Drone Trial 3 vs. Field	0.57	0.09	-3.47	0.002**
Drone Trial 4 vs. Field	0.60	0.10	-3.11	0.005**
Young				
Drone Trial 3 vs. Drone Trial 4	1.08	0.17	0.46	0.89
Drone Trial 3 vs. Field	1.19	0.19	1.08	0.52
Drone Trial 4 vs. Field	1.10	0.17	0.62	0.81

3.3.5 Spearman correlation analyses for trials 3 and 4

Trial 3 was significantly and strongly correlated with field data in four site-zones (Four Mile – Wetland, Mill Creek Complex South – Wetland, South Fork Newaukum – Riparian, and Stampede Pass – Buffer) indicating a strong agreement between drone and field plots (Table 3.8, Figure 3.5). Low to moderate but non-significant correlations were observed at Four Mile – Buffer, Lacamas – Riparian, and Stampede Pass – Wetland. Rock Creek–Wetland showed no correlation ($r_s = -0.02$, $p = 0.94$), suggesting that drone-derived cover estimates did not match field-based estimates.

Table 3.8 Spearman's rank-order correlation between drone-derived (Trial 3 and Trial 4) and field woody cover estimates.

This table presents Spearman's rank-order correlation coefficients (r_s) assessing the relationship between drone-derived (Trial 3) and field-based woody cover estimates across different site-zone combinations. Significance levels are denoted as follows: $***p < 0.001$, $**p < 0.01$, and $*p<0.05$.

Site	Zone	Correlation (r_s)	*p*-value	Correlation (r_s)	*p*-value
Trial 3 vs. Field			**Trial 4 vs. Field**		
Four Mile	Buffer	0.45	0.07	0.47	0.06
Four Mile	Wetland	0.86	<0.001***	0.85	**1.33e-04**
Lacamas	Riparian	0.36	0.13	0.55	**0.02**
Mill Creek Complex South	Wetland	0.76	<0.001***	0.74	**4.90e-05**
Rock Creek	Buffer	0.35	0.19	0.51	**0.04**
Rock Creek	Wetland	-0.02	0.94	0.20	0.53
South Fork Newaukum	Riparian	0.70	0.02*	0.54	0.11
Stampede Pass	Buffer	0.73	0.02*	0.80	**0.01**
Stampede Pass	Wetland	0.35	0.21	0.32	0.24

For Trial 4, significant and strong correlations were found at Four Mile – Wetland, Mill Creek Complex South – Wetland, Lacamas – Riparian, Rock Creek – Buffer, and Stampede Pass – Buffer, indicating strong agreement between drone and field estimates at these sites (Table 3.8, Figure 3.5). Moderate but non-significant correlations were observed at Four Mile – Buffer and South Fork Newaukum – Riparian. Rock Creek – Wetland and Stampede Pass – Wetland showed weak and non-significant correlations, suggesting that drone-derived cover estimates are less reliable at the plot level, despite higher agreement at the zone level.

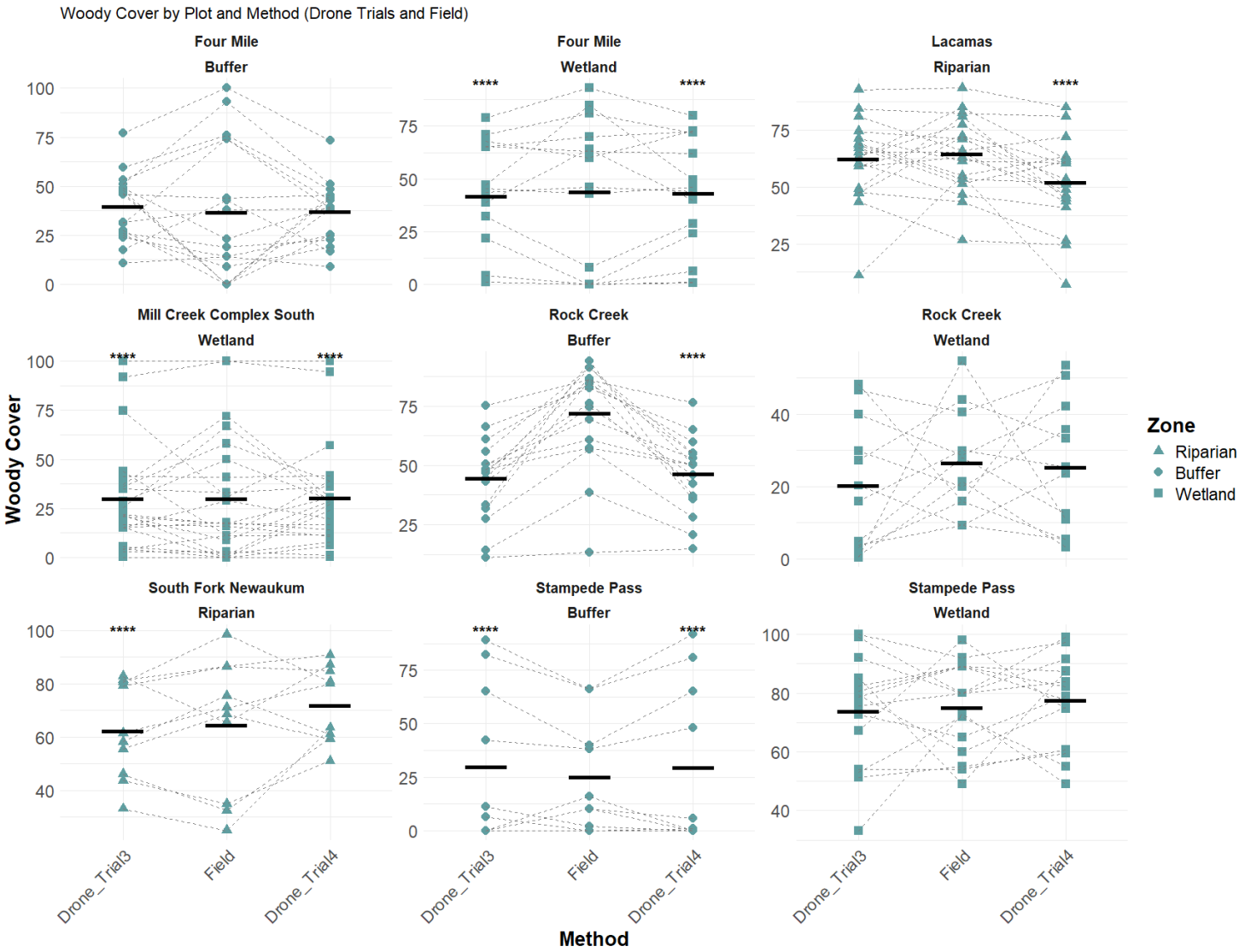

Figure 3.5 Scatterplots of woody cover estimates for Trial 3, field data, and Trial 4, grouped by site and zone.

Dashed lines connect paired plots within the same site, illustrating relationships among the three monitoring methods. Black lines indicate the mean cover value for each method. Significant correlations with field data are denoted by *** above the Trial 3 and Trial 4 column

3.3.6 Comparison of drone trials at each scale

Despite the variability in field and drone correlations at the plot scale, the drone trials exhibited a strong correlation within their respective scales (Figure 3.6). The correlation between Trial 3 estimates at the zone and plot scales was statistically significant ($r_s = 0.98$, $p < 0.001$), as was the correlation between Trial 4 estimates at the two scales ($r_s = 0.96$, $p < 0.001$). Therefore, within each trial, cover estimates remained consistent across scales. At the zone scale, the total woody cover estimate for the entire zone closely aligned with the average woody cover within each drone plot. However, the cover of each drone plot did not align with the cover of each field plot, suggesting that while drone estimates are internally consistent, they do not necessarily correspond with field-based measurements at the plot level.

Even though field and drone measurements did not always align at the plot scale, the mean difference between plot-level drone estimates and field data was notably smaller in both Trial 3 and Trial 4 compared to the mean difference observed at the zone level (Table 3.9). Trial 4 exhibited the smallest mean difference, at just 2.8%. This suggests that the weak correlations between field and drone methods at the plot scale may be influenced by factors beyond methodological differences between drone and field estimates. The smaller mean differences at the plot level compared to zone-level indicated that the average cover across drone plots was more closely aligned with the average cover across field plots, even if the raw data were not correlated.

Table 3.9 Differences between drone and field estimates of woody cover at zone and plot scales for Trials 3 and 4.

The differences are presented as $(T3\text{-}F)_Z$ and $(T4\text{-}F)_Z$ for zone-level comparisons, with T3 representing Trial 3 and T4 representing Trial 4. The subscript "z" represents zone-level comparisons, and the subscript "p" represents plot-level comparisons. In the site column, Mill Creek C.S. refers to Mill Creek Complex South. Negative values indicate that drone estimates were lower than field measurements. The average differences for each category are provided in the final row in bold. Trial 4 exhibits the smallest mean difference from field data at both the zone and plot scales.

Site	Zone	Difference $(T3\text{-}F)_Z$	Difference $(T4\text{-}F)_Z$	Difference $(T3\text{-}F)_p$	Difference $(T4\text{-}F)_p$
Rock Creek	Wetland	-8.1	-2.7	-6.4	-1.3
Rock Creek	Buffer	-29.8	-27.3	-25.5	-23.5
Four Mile	Buffer	-4.5	-4.6	1.4	-1.3
Four Mile	Wetland	-11.8	-11.3	-2.0	-0.9
Stampede Pass	Wetland	-10.7	-9.0	-1.4	2.3
Stampede Pass	Buffer	5.3	-3.5	4.8	4.4
South Fork Newaukum	Riparian	-0.6	11.1	-2.3	7.3
Mill Creek C.S.	Wetland	-3.9	-2.7	-0.2	0.4
Lacamas Creek	Riparian	-3.0	-14.2	-2.3	-12.5
Averages		**-7.5**	**-7.1**	**-3.8**	**-2.8**

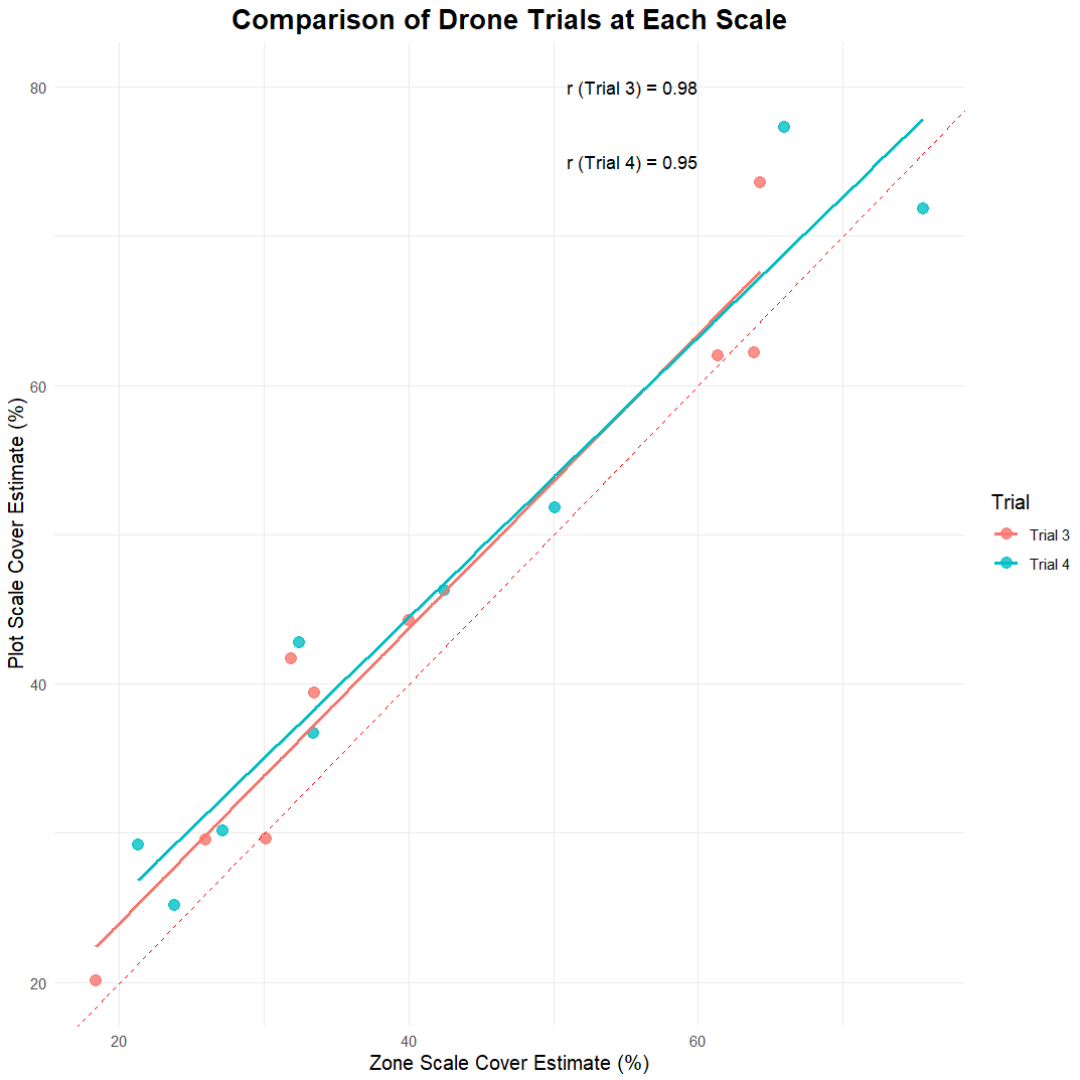

Figure 3.6 Comparison of drone trials at each scale

Scatterplot comparing woody cover estimates between drone trials at the zone and plot scales. Points represent individual site-zone combinations, with red points and regression line corresponding to Trial 3 and blue points and regression line corresponding to Trial 4. The dashed red line represents the 1:1 line, showing the perfect agreement between the zone and plot measurements.

3.4 Discussion

3.4.1 Zone scale analysis

Trial 3 and 4 showed marginally significant differences from field data. On average, both drone trials consistently underestimated woody cover compared to field methods by 7.5% in Trial

3 and 7.1% in Trial 4. The largest underestimations were observed at Rock Creek in the buffer zone, suggesting that site-zone factors, rather than differences between the drone trials, may be influencing the estimates. At this site, underestimation was greatest in the buffer zone, likely because both machine learning classifiers used in each trial, SVM for Trial 3 and random trees for Trial 4, struggled to differentiate woody vegetation from herbaceous vegetation. Rock Creek also failed to meet target classification accuracies in Chapter 2 due to misclassification between woody and green herbaceous vegetation, which contributed to woody cover underestimation. Within the buffer zone, both trials misclassified green herbaceous pixels at the zone and plot scales, reducing the number of woody pixels and, consequently, the estimated woody cover.

Both trials demonstrated similar patterns of underestimation across most sites, except the buffer zone in Stampede Pass and the riparian zone in South Fork Newaukum. At Stampede Pass, Trial 3 overestimated cover by 5.3%, whereas Trial 4 underestimated it by 3.5%. Although Stampede Pass met target classification accuracies in both trials, its complexity, due to overlapping woody and herbaceous vegetation, as well as non-vegetation classes (e.g., water and shadows), may have influenced these differences. Trial 4, which used a random trees classifier, appeared to handle this complexity better, reducing noise between overlapping classes but slightly overestimating woody cover. In contrast, at South Fork Newaukum, Trial 3 underestimated cover by less than 1%, while Trial 4 overestimated it by 11%. This discrepancy may be due to SVM's ability to classify linear and clearly separable data more effectively. At South Fork Newaukum, where woody vegetation was more distinct and classification accuracies were higher (Chapter 2), SVM produced estimates closer to field data. Although Chapter 2 showed that classifier choice was not statistically significant, these results suggest that classifiers can still produce different woody cover estimates. In zones where woody vegetation was clearly distinguishable, SVM

produced more consistent estimates aligned with field data, while Trial 4 had more variability. Conversely, in more complex sites with overlapping classes, random trees performed better, yielding woody cover estimates closer to field data.

The differences between classifiers suggested that site complexity and vegetation distinctiveness should be carefully considered when selecting a classification method. On sites where woody vegetation was clearly distinguishable from other classes, SVM provided more consistent and accurate estimates. However, in more complex sites with overlapping vegetation and non-vegetation classes, random trees performed better due to their ability to handle noisy data. These findings emphasized the importance of adapting classification approaches to site-specific characteristics rather than relying on a one-size-fits-all method. With the advancement of AI and built-in classifiers in GIS software like ArcGIS Pro, running both classifiers initially helped determine which method performed best for a given site. Since built-in classifiers in ArcGIS Pro were faster than custom applications, running multiple models was a feasible approach.

One potential explanation for the overall pattern of underestimation in both trials was the difference in sampling intensity between the drone and field methods. The drone's zonal estimates encompassed the entire zone of interest (e.g., riparian, wetland, or buffer), whereas the field estimates were based on a sample *within* the zone of interest. As a result, the drone's assessment represented a broader study area, whereas the field data reflected only a subset of that area. This difference in scale was meaningful because the drone assessment included areas within the zone that were not sampled in the field due to constriction points or site inaccessibility. Constriction points, which are smaller areas along the edges of a zone that are narrower than the sample unit length, were typically excluded from field sampling. Although these areas represented only a small portion of the site, their exclusion could contribute to differences in cover estimates between the

two methods, as edges often contain distinct spatial patterns of vegetation. Another contributing factor to the underestimation was the exclusion of inaccessible areas from the field data. During ground-based sampling, it was common for sample units to be discarded if they could not be accessed on foot. This issue was more common in larger or more mature sites, as well as sites with safety concerns such as elevated water levels from natural stream features or unexpected inundation caused by beaver activity. These challenges made it difficult to systematically account for the entire site, leading to potential gaps in the field data. As such, cover differences between the drone and field methods may have resulted from the omission of areas that could have contained either sparser or denser woody vegetation not represented in the field samples. Since the sample design for field data did not consistently record the number or locations of discarded sample units, the specific areas that remained unsampled were often unclear. In some cases, however, the sample design did document which areas were excluded from sampling.

For instance, at Lacamas Creek, the north end of the site was excluded from field sampling because it was not large enough to meet the sample unit length, resulting in approximately 30 meters of the site being excluded from the field-based sample design. This gap occurred due to a constriction point. Additionally, the western edge of the site was unstable due to erosion, leaving approximately 1 to 3 meters along the creek edge inaccessible. This limitation highlights the importance of considering scale and accessibility differences when comparing drone-based and field-based woody cover estimates, especially at the zone scale.

Conversely, a disadvantage of the zonal estimates from the drone trials was the lack of flexibility to exclude areas that were not traditionally part of the woody cover assessment. The drone zonal estimate may have included sections that were intentionally left unplanted. For example, aquatic features and habitat structures were extracted from field data, because they were

not planted, but these features were included in the drone's zonal assessment. The inclusion of these features lowered the overall cover estimate in the drone data, while the exclusion of these features increased the cover estimate in the field data. A simple solution would be to map and exclude these areas from the drone raster, particularly if common practices dictate the exclusion of features like habitat structures. Implementing this solution could reduce discrepancies in cover estimates between the two methods.

While there are no agency guidelines on what constitutes being extracted from or included in a zone's sample design, field methods offer the advantage of allowing on-the-ground decisions about how to monitor a zone. However, this flexibility comes with the tradeoff of potentially introducing biases by excluding areas that should have been included in the sample. In contrast, drone-based assessments offer a standardized approach but lack the ability to make site-specific exclusions, if needed. Ultimately, the differences between these two methods underscore the need for a monitoring approach that balances consistency with adaptability. In this analysis, drones offer a more consistent and standardized approach for assessing woody cover across large and inaccessible areas, reducing subjectivity in monitoring efforts.

3.4.2 Plot scale analysis

Field-based cover estimates were significantly higher than those from Trial 3 and Trial 4, however post hoc tests revealed that these differences were strongest in the riparian zone and on mature aged sites. The slightly higher field estimates, at the plot scale, were likely associated with ground-based observers capturing understory vegetation and smaller woody plantings that were not easily detected in drone imagery. Additionally, shadows present in drone imagery may have contributed to the underestimation pattern, as shadowed pixels reduced the apparent cover of

woody vegetation. This issue was particularly pronounced when larger trees cast extensive shadows over surrounding woody plants, effectively masking vegetation that would have been recorded during ground-based surveys. Shadows can significantly alter the spectral characteristics of affected pixels, making it difficult to distinguish between shaded vegetation and bare ground or non-woody cover. As a result, even areas densely populated with smaller woody plants could appear as having lower cover in the drone-derived estimates.

Despite the underestimation, plot-scale woody cover estimates had smaller mean differences from field estimates compared to zonal estimates. This continued to support that differences in woody cover estimates observed at the zone scale were explained by scale differences, as field data typically excluded areas that were incorporated into the broader zonal assessment. These results reinforce the importance of considering both spatial scale and imagery characteristics when interpreting differences between drone-derived and field-based woody cover estimates.

The correlation between drone and field estimates at the plot scale was variable. While some sites exhibited significant positive correlations, others showed low to moderate correlations that were not statistically significant. The most likely explanation for this discrepancy was the accuracy of field-sampled areas compared to their corresponding drone-assessed plots in drone imagery. The GPS accuracy of the field equipment used to navigate transects ranged from 2 to 5 meters, meaning plots could be offset by up to 5 meters from their drone locations. Additionally, any field errors, such as misalignment of transects that were not set exactly perpendicular to the baseline, further compounded the margin of error. These errors likely weakened correlations at the plot level and contributed to differences between methods, as GPS inaccuracies alone could have resulted in a field plot sampling vegetation up to five meters away from its paired drone plot.

Even though there were weak correlations and significant differences between methods, the mean difference between drone and field plot estimates was 3.8% for Trial 3 and 2.8% for Trial 4. By zone, the mean differences for Trials 3 and 4 were 6.6% and 7% in the buffer zone, 2.3% and 2.6% in the riparian zone, and 2.6% and 0% in the wetland zone, respectively. By age, the mean differences were 6.2% and 4% in mature plots and 1% and 1.5% in young-aged plots, respectively. Although post hoc tests revealed statistical differences in the riparian zone and mature plots, these results may have been influenced by a small and unbalanced sample and were based on logit-transformed values in the GLMM rather than the mean of raw values. However, mean differences in the raw values supported drones as a viable alternative for estimating woody cover. The following section evaluated whether these differences were meaningful in assessing performance standards.

3.4.3 Evaluating performance standards with different methods

Drone-based methods demonstrated consistency in performance standard assessment with field estimates at seven of the nine sites (Table 3.10). At five sites, all methods and scales consistently indicated that the performance standard was met. At two sites, all methods at both scales consistently showed that the standard was not met. This consistency across the majority of sites highlights the potential for integrating drone-based methods to efficiently and accurately assess performance standards, regardless of whether sites are successfully establishing or underperforming.

Differences in performance standard assessments by method occurred at Four Mile – Wetland and Rock Creek – Buffer, despite overall consistency across methods. At Four Mile – Wetland, zone-scale assessments from Trials 3 and 4 indicated that the standard was not met,

whereas plot-scale assessments showed compliance. Since the plot-scale assessment had a lower mean cover difference from field data than the zone-scale assessment, this suggested that scale influenced cover estimates. The zone-scale assessment likely captured areas that were excluded from the field sampling design, highlighting a scale-related issue rather than a limitation of drone-based methods. However, classification accuracy may have also impacted the evaluation of performance standards, as Four Mile did not meet the target accuracy levels in either trial (Chapter 2).

In contrast, At Rock Creek – Buffer, none of the trials at either scale reported compliance, despite field estimates showing the standard was met. If drone-derived results were submitted in a monitoring report, this site would be incorrectly labeled as non-compliant. This outcome was expected, as Rock Creek had the lowest classification accuracies and the largest mean cover underestimation at both scales. However, such misclassification should be avoided when using drones for monitoring. A key challenge was distinguishing woody vegetation from herbaceous vegetation within the buffer zone. This was not possible given the minimum spectral resolution and simplified classification approach used in this study. This highlights an important consideration: while drones can be highly effective for monitoring, as demonstrated by their success across most sites (Table 3.10), users must be aware of the limitations of their data. Rock Creek and Four Mile were more complex sites, where the tools and methods used in this study may not have been sufficient to accurately capture woody cover. Future refinements could incorporate additional spectral bands, explore alternative classification algorithms with more robust object-based methods, or utilize the vertical data trials to determine if they improve differences between methods at the plot level.

Another important assumption of this study was that all woody vegetation on the sites was native, since performance standards specifically address “native woody vegetation”. Since these sites were typically cleared, mulched, and planted, most non-native noxious weeds were expected to be present in the herbaceous layer rather than the woody layer. While non-native woody species in the shrub and tree strata were uncommon, exceptions may have occurred. For example, field data, which is collected at the species level, revealed that South Fork Newaukum and Lacamas Creek had less than one percent cover of Himalayan blackberry. Even if the drone data included an additional percentage of woody cover for Himalayan blackberry, it would not have made a difference in how the drone methods evaluated the performance standards for these sites.

Overall, Table 3.10 reinforces that drone surveys were effective for assessing native woody vegetation performance standards across seven of the nine site-zones. Cautious managers, biologists, and regulators considering the use of drones for monitoring could observe a general pattern of underestimation in woody cover. While drone methods consistently underestimated woody cover compared to field methods, this was preferable to overestimation, as the greater concern would have been falsely meeting performance standards due to overestimated cover. When spectral overlap within the woody class was minimal, the methods explored in this study performed adequately. However, when spectral overlap was greater, the analysis became more complex, requiring either more advanced classification techniques or reliance on field-based monitoring for more accurate assessments. Differences between classifiers were also evident in Table 3.11, where one classifier tended to overestimate woody cover while the other underestimated it.

Table 3.10 Performance standard assessment.

The table shows woody cover performance standards and percentage cover values. The table compares field-derived data with drone-derived estimates from Trial 3 and Trial 4 at both zone and plot scales. Green cells indicate that the performance standard is met, while red cells indicate it is not met. Notable discrepancies in methods are highlighted in **bold**, reflecting differences in how the methods determine whether the performance standard is achieved.

Site Name/ Zone	Performance Standard	Field	T3 (zone)	T4 (zone)	T3 (plot)	T4 (plot)
Four Mile/ Wetland	Minimum 35 percent cover of native woody vegetation in the wetland areas	**43.7**	**31.9**	**32.4**	41.7	42.8
Four Mile/ Buffer	Minimum 25 percent cover of native woody vegetation in the buffer areas	38	33.5	33.4	39.4	36.7
South Fork Newaukum/ Riparian	Minimum 35 percent cover of native woody vegetation in the restoration areas	64.5	63.9	75.6	62.2	71.8
Lacamas Creek/ Riparian	Minimum 35 percent cover of native woody vegetation in the restoration areas	64.3	61.3	50.1	62	51.8
Rock Creek/ Wetland	Minimum 50 percent cover of native woody vegetation in the wetland	26.5	18.4	23.8	20.1	25.2
Rock Creek/ Buffer	Minimum 50 percent cover of native woody vegetation in the buffer	**69.8**	**40**	**42.5**	**44.3**	**46.3**
Stampede Pass/ Wetland	Minimum 20 percent cover of native woody vegetation in the scrub-shrub wetland areas	75	64.3	66	73.6	77.3
Stampede Pass/ Buffer	Minimum 12.5 percent cover of native woody vegetation within the upland areas	24.8	30.1	21.3	29.6	29.2
Mill Creek Complex South/ Wetland	Minimum 70 percent cover of native woody vegetation in the wetland	29.8	25.9	27.1	29.6	30.2

3.5 Future research and implications for monitoring

These findings highlight the potential of drone-based monitoring for assessing woody vegetation performance standards but also emphasize the importance of site-specific considerations when interpreting results. While drones provide a more standardized and scalable method compared to traditional field-based monitoring, agencies and land managers should establish accuracy thresholds and site-selection criteria before integrating drones into monitoring programs. Given the small sample sizes in some site-zones, these results may lack the statistical power needed to fully capture method variability across different zones and ages (i.e. in riparian zones and mature sites). Expanding validation efforts across a larger number of vegetation zones and age classes would strengthen confidence in drone-based monitoring. The results also suggest that machine learning classifiers can influence woody cover estimates, which underscores the need for further validation studies across a wider range of sites.

For managers and biologists, a strong understanding of imagery analysis and classification remains critical when working with drone data. Equally important is an awareness of input imagery and its limitations. Just as understanding sample design and field monitoring protocols has always been essential for traditional methods, understanding classification methods is key to successfully applying drone-based monitoring. Recognizing these limitations and strengths will help optimize drone-based assessments and ensure reliable performance evaluations.

References

Barnas, A. F., Darby, B. J., Vandeberg, G. S., Rockwell, R. F., & Ellis-Felege, S. N. (2019). A comparison of drone imagery and ground-based methods for estimating the extent of habitat destruction by lesser snow geese (Anser caerulescens caerulescens) in La Pérouse Bay. *PLOS ONE, 14*(8), e0217049. https://doi.org/10.1371/journal.pone.0217049

Bonham, C. D. (1989). Monitoring and evaluation. In *Measurements for terrestrial vegetation* (pp. 265–311). John Wiley & Sons.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). *glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling*. *The R Journal, 9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Coulloudon, B., Eshelman, K., Gianola, J., Habich, N., Hughes, L., Johnson, C., ... & Willoughby, J. W. (1999). *Sampling vegetation attributes: Interagency technical reference.* U.S. Department of the Interior, Bureau of Land Management.

Esri. (n.d.). Tabulate area (spatial analyst). *ArcGIS Pro Documentation*. Retrieved March 2, 2025, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/tabulate-area.htm

Gómez-Sapiens, M., Schlatter, K. J., Meléndez, Á., Hernández-López, D., Salazar, H., Kendy, E., & Flessa, K. W. (2021). Improving the efficiency and accuracy of evaluating aridland riparian habitat restoration using unmanned aerial vehicles. *Remote Sensing in Ecology and Conservation, 7*(3), 488–503. https://doi.org/10.1002/rse2.204

Juniper Systems. (n.d.). *Mesa rugged tablet*. Retrieved January 25, 2025, from https://junipersys.com/products/mesa-rugged-

tablet#:~:text=The%20Mesa%203%20and%20Mesa,to%205%2Dmeter%20typical%20accuracy

Lenth, R. V. (2024). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.9.0) [R package]. The Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=emmeans

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 4.1.2) [Software]. R Foundation for Statistical Computing. https://www.r-project.org

Washington State Department of Ecology. (2005). *Wetlands in Washington State: Volume 1: A synthesis of the science* (Publication No. 05-06-006). Washington State Department of Ecology. https://apps.ecology.wa.gov/publications/documents/0506006.pdf

Washington State Department of Transportation. (2021). *Wetland monitoring methods* (Publication No. ENV-WET). Washington State Department of Transportation. https://wsdot.wa.gov/sites/default/files/2021-10/Env-Wet-MonitoringMethods.pdf

Chapter 4: Summary and conclusions

This chapter synthesizes the key findings and insights from this research, which focused on evaluating the potential of drone-based methods for regulatory monitoring of woody vegetation on compensatory mitigation sites. The primary objective of this thesis was to determine whether drone surveys could offer reliable, accurate, and efficient alternatives to traditional field-based methods, particularly for assessing woody cover performance standards. Throughout this study, I investigated how various image classification approaches influenced accuracy and how drone-derived woody cover estimates compared with conventional field measurements. Additionally, I identified practical challenges and limitations associated with drone monitoring, providing a comprehensive assessment of both the benefits and constraints of this emerging technology.

The results presented in this research demonstrate that drone-based methods hold significant promise for enhancing monitoring efforts, especially in terms of capturing woody vegetation cover at mitigation sites. These findings indicate that drones can be a valuable addition to traditional monitoring practices, offering increased spatial coverage and detail with less intensive fieldwork. However, integrating these technologies into regulatory frameworks requires addressing practical challenges, including image processing demands, data storage, and the need for specialized training. This chapter concludes by discussing the broader implications of these findings, the limitations encountered, and potential future applications. By emphasizing the strengths and weaknesses of drone-based monitoring, this chapter aims to provide guidance for natural resource agencies and practitioners as they consider adopting emerging remote sensing technologies.

4.1 Chapter 2 summary

This thesis primarily aimed to assess the potential of drone surveys as a method for regulatory monitoring of woody vegetation on compensatory mitigation sites. First, I explored how different steps in the image classification process influence overall accuracy and the accuracy of the woody land cover class. The findings in Chapter 2 show that object-based classifications consistently outperformed pixel-based classifications across all sites. For two sites with lower accuracies, integrating vertical data improved overall accuracy and producer and user accuracies for the woody class. Additionally, while the choice of machine learning classifiers, whether support vector machines or random trees, did not significantly impact the accuracy metrics, they did perform differently in their delineation of woody vegetation. These findings highlight the importance of selecting an appropriate classification approach and classifier, as object-based methods and the integration of vertical data can improve accuracy. This suggests that future regulatory monitoring efforts using drone-based imagery should prioritize methodologies in segmentation and feature selection to improve woody vegetation classification.

4.2 Chapter 3 summary

In Chapter 3, I compared woody cover estimates derived from Trials 3 and 4 in Chapter 2 to field-based estimates obtained through the line-intercept method at both the zone and plot scales. The analysis revealed marginally significant differences between drone trials (Trial 3 and 4) and field-derived estimates at the zone scale. Cover estimates from both trials were highly correlated with field estimates. These findings suggest that drone-based methods, particularly Trial 4, provide comparable woody cover estimates, supporting their use in regulatory monitoring. At the plot scale, there were significant differences across monitoring methods (Trial 3, Trial 4, and

field) in riparian zones and mature aged sites. Estimates from both trials were highly correlated with field estimates for some site-zone combinations, while most site-zone combinations exhibited low to moderate, non-significant correlations. These findings indicated that drone-derived cover estimates align well with field measurements under certain conditions but were more variable at plot scales, particularly in some site-zone combinations, such as Rock Creek – Buffer and Four Mile - Wetland. This underscores the need for further investigation into how site-specific factors, such as vegetation structure, spatial homogeneity, and the proportion of herbaceous and woody vegetation, influence classification accuracy and woody cover estimates at the plot scale. These larger underestimations of woody cover also highlight the importance of adapting classification approaches to the unique characteristics of each site rather than relying on a one-size-fits-all method. Implementing more advanced classification techniques at sites like Rock Creek and Four Mile could improve accuracy and better align drone-derived and field-based estimates. Despite differences in methods, drone-based and field-based methods arrived at identical conclusions for seven of nine woody cover performance standards.

4.3 Limitations

This study established a foundation for integrating drone surveys into regulatory monitoring and demonstrated that drone-based methods can effectively assess woody cover performance standards in a manner comparable to traditional field techniques. While these results are promising, several limitations must be addressed in future research. One key challenge was that performance standards were written for native vegetation, yet the standard RGB imagery used in this study was not well-suited for species-level classification. This limitation was relevant for herbaceous vegetation where non-native vegetation was more prevalent and difficult to discern

from native species. As a result, the drone-based approach created in this study may not be directly applicable to regulatory assessments of herbaceous cover or species composition performance standards.

To address this, agencies must evaluate whether investing in more advanced equipment, such as sensors that incorporate additional spectral bands, including near-infrared, would enhance drone monitoring capabilities. However, this decision involves a tradeoff between cost and complexity, as more sophisticated equipment requires not only a larger budget but also expertise in advanced remote sensing techniques. These are important considerations, as all remote sensing tools involve tradeoffs, and while higher-quality data can be obtained with better equipment, it comes at the expense of increased technical complexity and training requirements.

Additionally, while drone surveys can estimate woody cover, field surveys remain necessary for ecological assessments of compensatory mitigation sites. Remote sensing cannot fully capture wetland functions and values, which are critical components of determining mitigation success. Therefore, while drones provide a valuable tool for monitoring certain performance standards, field-based assessments will continue to play a crucial role in evaluating overall site success.

Another important consideration for the use of drone surveys in monitoring is the time investment required for image processing and the cost of data storage. In this study, I used 25 gigabytes (GB) of storage for all GIS products. This included four classification trials for four sites and six classification trials for two sites. I also used an additional 25 GB for all image products and processing outputs, ranging from individual drone-captured photos to orthomosaics and DSMs. For agencies considering drone monitoring, data storage needs will scale significantly with the number of sites, resolution of imagery, and frequency of surveys. Additionally, image

processing can be time-intensive, requiring both computational power, which can typically exceed the capabilities of a standard laptop. Agencies must assess whether they have the infrastructure, personnel, and processing capabilities to manage large datasets efficiently. Cloud-based storage and automated processing workflows, such as those available through DroneDeploy, may help streamline these challenges but introduce additional costs. As drone technology continues to evolve, agencies should weigh the benefits of high-resolution imagery against the logistical demands of data management to ensure sustainable and efficient monitoring programs.

4.4 Future applications

Given that mitigation sites are monitored over a 10-year period, future research should focus on comparing the cost and time efficiency of drone-based surveys versus traditional field methods for monitoring. While some studies have explored these efficiencies (Barnas et al. 2019; Gómez-Sapiens et al. 2021), they have only examined the first two years of monitoring, leaving a gap in understanding the advantages and disadvantages of drones over extended monitoring periods. Although drone-based methods involve longer processing times and a steeper learning curve, particularly when becoming familiar with classification techniques, these challenges would diminish over time, especially when conducting repeat surveys at the same site. As workflows for imagery processing and classification become more standardized due to advances in AI, drones have the potential to match or even surpass the efficiency of traditional monitoring methods. For example, the ability to conduct large-scale assessments in areas that are difficult or impossible to access using field methods not only reduces field bias but also improves monitoring coverage and consistency. Additionally, repeat drone surveys enable spatial change analysis over time, providing insights into vegetation dynamics and site development that are not feasible with conventional

field data alone. Future studies should further investigate how automation and streamlined workflows can enhance the long-term feasibility of drone-based regulatory monitoring.

Future research should explore broader agency applications for standardized drone monitoring. One key area of focus is assessing the effectiveness of adaptive management strategies, such as replanting efforts and invasive species removal, by tracking how vegetation responds to restoration activities over time. Drones could also be used to collect baseline vegetation and cover data prior to planting, providing a reference for evaluating site development. Another research direction is the role of drones in long-term site management after the monitoring period has ended. Compensatory mitigation sites are intended to be maintained in perpetuity with the responsibility falling on the permittee, yet there are no clear regulatory guidelines on what that entails. Drone monitoring can help fill this gap by providing cost-effective and repeatable assessments to document major changes to site conditions.

4.5 Conclusion

Integrating drones into standard monitoring practices allows agencies to diversify their monitoring toolkits, ensuring that monitoring continues smoothly, particularly when traditional field methods face constraints. As modern biologists increasingly rely on technology-based tools, there remains a disconnect between wetland practitioners and remote sensing applications. While platforms such as ArcGIS and QGIS provide accessible tools for image classification and analysis, their integration requires specialized training.

As a complex discipline, usability is the greatest potential barrier to remote sensing. This research aimed to bridge that gap by making drone applications more accessible to biologists, ensuring that new tools remain practical and user-friendly. Thus, prioritizing tangible and simple

applications over unnecessarily complex workflows was essential. Despite the increased use of drones in environmental studies, they have not yet been extensively explored for regulatory monitoring. This work represented a foundational step toward developing standardized guidelines for their integration into regulatory frameworks. By reducing technological barriers and demonstrating feasible approaches, this research paves the way for improvements to wetland management.

Appendix A: Chapter 2 Confusion Matrices

Overall accuracy (OA), producer accuracy (PA), and user accuracy (UA) for the woody class are ***highlighted*** and shown as percentages.

Table A.1 Confusion matrix for Trial 1 (pixel-based SVM) at Lacamas Creek

Classified Data	Woody	Bare	Brown Herbs	Shadows	Green Herbs	Samples	***UA***
Woody	**179**	0	4	7	3	193	***93%***
Bare	5	**81**	1	0	0	87	93%
Brown Herbs	11	9	**73**	4	4	101	72%
Shadows	0	0	0	**22**	0	22	100%
Green Herbs	61	0	1	0	**35**	97	36%
Samples	256	90	79	33	42	500	
PA	***70%***	90%	92%	67%	83%		
OA							***78%***
Kappa							0.69

Table A.2 Confusion matrix for Trial 1 (pixel-based SVM) at Mill Creek Complex South

Classified Data	Woody	Brown Herbs	Green Herbs	Bare	Shadows	Samples	***UA***
Woody	153	2	5	1	0	161	***95%***
Brown Herbs	0	169	4	0	0	173	98%
Green Herbs	40	5	36	0	0	81	44%
Bare	1	16	0	9	0	26	35%
Shadows	5	3	0	0	55	63	87%
Samples	199	195	45	10	55	504	
PA	***77%***	87%	80%	90%	100%		
OA							***84%***
Kappa							0.77

Table A.3 Confusion matrix for Trial 1 (pixel-based SVM) at Rock Creek

Classified Data	Woody	Green Herbs	Bare	Brown Herbs	Shadows	Water	Samples	***UA***
Woody	143	10	0	0	0	0	153	***93%***
Green Herbs	54	67	3	7	0	0	131	51%
Bare	0	2	49	18	0	0	69	71%
Brown Herbs	9	8	12	60	0	0	89	67%
Shadows	16	4	5	0	10	0	35	29%
Water	0	1	20	0	0	10	31	32%
Samples	222	92	89	85	10	10	508	
PA	***64%***	73%	55%	71%	100%	100%		
OA								***67%***
Kappa								0.57

Table A.4 Confusion matrix for Trial 1 (pixel-based SVM) at Stampede Pass

Classified Data	Woody	Bare	Water	Brown Herbs	Green Herbs	Shadows	Samples	***UA***
Woody	130	0	1	1	2	1	135	***96%***
Bare	4	141	0	3	0	0	148	95%
Water	1	0	64	0	0	1	66	97%
Brown Herbs	1	12	0	42	0	0	55	76%
Green Herbs	69	0	2	6	8	0	85	9%
Shadows	0	1	0	0	0	15	16	94%
Samples	205	154	67	52	10	17	505	
PA	***63%***	91%	96%	81%	80%	88%		
OA								***79%***
Kappa								0.73

Table A.5 Confusion matrix for Trial 1 (pixel-based SVM) at South Fork Newaukum

Classified Data	Woody	Brown Herbs	Bare	LWD	Shadows	Samples	***UA***
Woody	263	1	8	5	1	278	***95%***
Dead Herbaceous	7	19	28	0	0	54	35%
Bare	6	27	65	14	0	112	58%
LWD	0	2	3	26	0	31	84%
Shadows	12	0	0	0	13	25	52%
Total Samples	288	49	104	45	14	500	
PA	***91%***	39%	63%	58%	93%		
OA							***78%***
Kappa							0.63

Table A.6 Confusion matrix for Trial 2 (pixel-based RT) at Lacamas Creek

Classified Data	Woody	Bare	Brown Herbs	Shadows	Green Herbs	Samples	***UA***
Woody	187	0	3	2	5	197	***95%***
Bare	8	70	1	0	0	79	87%
Brown Herbs	8	20	67	4	6	105	64%
Shadows	1	0	0	27	0	28	96%
Green Herbs	52	0	8	0	31	91	34%
Samples	256	90	79	33	42	500	
PA	***73%***	78%	85%	81%	74%		
OA							***76%***
Kappa							0.67

Table A.7 Confusion matrix for Trial 2 (pixel-based RT) at Mill Creek Complex South

Classified Data	Woody	Brown Herbs	Green Herbs	Bare	Shadows	Samples	***UA***
Woody	163	2	8	0	0	173	***94%***
Brown Herbs	0	154	1	0	0	155	99%
Green Herbs	34	5	36	0	0	75	48%
Bare	0	29	0	10	0	39	26%
Shadows	2	5	0	0	55	62	89%
Samples	199	195	45	10	55	504	
PA	***82%***	79%	80%	100%	100%		
OA							***83%***
Kappa							0.76

Table A.8 Confusion matrix for Trial 2 (pixel-based RT) at Rock Creek

Classified Data	Woody	Green Herbs	Bare	Brown Herbs	Shadows	Water	Samples	***UA***
Woody	142	23	1	0	0	0	166	***86%***
Green Herbs	63	59	2	8	0	0	132	45%
Bare	2	2	69	11	0	1	85	81%
Brown Herbs	5	4	7	66	0	0	82	80%
Shadows	10	4	5	0	10	0	29	34%
Water	0	0	5	0	0	9	14	64%
Samples	222	92	89	85	10	10	508	
PA	***64%***	64%	78%	78%	100%	90%		
OA								***70%***
Kappa								0.60

Table A.9 Confusion matrix for Trial 2 (pixel-based RT) at Four Mile

Classified Data	Woody	Bare	Brown Herbs	Green Herbs	Shadows	Stressed	Samples	***UA***
Woody	63	2	3	13	0	0	81	***78%***
Bare	2	152	19	0	0	0	173	88%
Brown Herbs	7	23	105	7	0	1	143	73%
Green Herbs	48	0	0	16	0	0	64	25%
Shadows	2	0	0	0	10	0	12	83%
Stressed	1	15	12	0	0	9	37	24%
Samples	123	192	139	36	10	10	510	
PA	***51%***	79%	76%	44%	100%	90%		
OA								***70%***
Kappa								0.59

Table A.10 Confusion matrix for Trial 2 (pixel-based RT) at Stampede Pass

Classified Data	Woody	Bare	Water	Brown Herbs	Green Herbs	Shadows	Samples	***UA***
Woody	136	0	0	0	1	1	138	***99%***
Bare	1	134	0	1	0	0	136	99%
Water	3	0	58	0	1	0	62	94%
Brown Herbs	1	20	0	47	0	0	68	69%
Green Herbs	64	0	9	4	8	0	85	9%
Shadows	0	0	0	0	0	16	16	100%
Samples	205	154	67	52	10	17	505	
PA	***66%***	87%	87%	90%	80%	94%		
OA								***79%***
Kappa								0.73

Table A.11 Confusion matrix for Trial 2 (pixel-based RT) at South Fork Newaukum

Classified Data	Woody	Brown Herbs	Bare	LWD	Shadows	Samples	***UA***
Woody	254	3	8	2	1	268	***95%***
Brown Herbs	10	17	25	0	0	52	33%
Bare	8	27	66	15	0	116	57%
LWD	0	2	5	28	0	35	80%
Shadows	16	0	0	0	13	29	45%
Samples	288	49	104	45	14	500	
PA	***88%***	35%	63%	62%	93%		
OA							***76%***
Kappa							0.61

Table A.12 Confusion matrix for Trial 3 (pixel-based SVM) at Lacamas Creek

Classified Data	Woody	Bare	Brown Herbs	Shadows	Green Herbs	Samples	***UA***
Woody	221	0	0	6	10	237	***93%***
Bare	2	66	0	0	0	68	97%
Brown Herbs	15	24	79	0	1	119	66%
Shadows	0	0	0	27	0	27	100%
Green Herbs	18	0	0	0	31	49	63%
Samples	256	90	79	33	42	500	
PA	***86%***	73%	100%	81%	74%		
OA							***85%***
Kappa							0.78

Table A.13 Confusion matrix for Trial 3 (pixel-based SVM) at Mill Creek Complex South

Classified Data	Woody	Brown Herbs	Green Herbs	Bare	Shadows	Samples	***UA***
Woody	194	0	6	0	0	200	***97%***
Brown Herbs	0	188	4	0	0	192	98%
Green Herbs	4	2	35	3	1	45	78%
Bare	0	5	0	7	0	12	58%
Shadows	1	0	0	0	54	55	98%
Samples	199	195	45	10	55	504	
PA	***97%***	96%	78%	70%	98%		
OA							***95%***
Kappa							0.92

Table A.14 Confusion matrix for Trial 3 (pixel-based SVM) at Rock Creek

Classified Data	Woody	Green Herbs	Bare	Brown Herbs	Shadows	Water	Samples	***UA***
Woody	178	19	0	2	0	0	199	***89%***
Green Herbs	37	72	6	9	0	0	124	58%
Bare	0	0	55	3	0	6	64	86%
Brown Herbs	4	1	26	71	0	4	106	67%
Shadows	3	0	2	0	10	0	15	67%
Water	0	0	0	0	0	0	0	0%
Samples	222	92	89	85	10	10	508	
PA	***80%***	78%	62%	84%	100%	0%		
OA								***76%***
Kappa								0.67

Table A.15 Confusion matrix for Trial 3 (pixel-based SVM) at Four Mile

Classified Data	Woody	Bare	Brown Herbs	Green Herbs	Shadows	Stressed	Samples	***UA***
Woody	77	0	3	14	0	0	94	***82%***
Bare	1	154	25	0	0	0	180	86%
Brown Herbs	3	38	103	3	0	3	150	69%
Green Herbs	42	0	0	19	0	0	61	31%
Shadows	0	0	0	0	10	0	10	100%
Stressed	0	0	8	0	0	7	15	46%
Total Samples	123	192	139	36	10	10	510	
PA	***63%***	80%	74%	53%	100%	70%		
OA								***73%***
Kappa								0.63

Table A.16 Confusion matrix for Trial 3 (pixel-based SVM) at Stampede Pass

Classified Data	Woody	Bare	Water	Brown Herbs	Green Herbs	Shadows	Samples	***UA***
Woody	194	0	0	5	5	1	205	***95%***
Bare	2	145	0	0	0	0	147	99%
Water	0	1	67	0	0	1	69	97%
Brown Herbs	2	8	0	43	0	0	53	81%
Green Herbs	7	0	0	4	5	0	16	31%
Shadows	0	0	0	0	0	15	15	100%
Samples	205	154	67	52	10	17	505	
PA	***95%***	94%	100%	83%	50%	88%		
OA								***93%***
Kappa								0.90

Table A.17 Confusion matrix for Trial 4 (pixel-based RT) at Lacamas Creek

Classified Data	Woody	Bare	Brown Herbs	Shadows	Green Herbs	Samples	***UA***
Woody	233	0	24	0	20	277	***84%***
Bare	6	60	0	0	0	66	91%
Brown Herbs	8	30	55	0	4	97	57%
Shadows	0	0	0	33	0	33	100%
Green Herbs	9	0	0	0	18	27	67%
Samples	256	90	79	33	42	500	
PA	91%	67%	70%	100%	43%		
OA							***80%***
Kappa							0.69

Table A.18 Confusion matrix for Trial 4 (pixel-based RT) at Mill Creek Complex South

Classified Data	Woody	Brown Herbs	Green Herbs	Bare	Shadows	Samples	***UA***
Woody	189	2	13	2	0	206	***92%***
Brown Herbs	0	187	1	3	0	191	98%
Green Herbs	9	3	31	1	1	45	69%
Bare	0	3	0	4	0	7	57%
Shadows	1	0	0	0	54	55	98%
Samples	199	195	45	10	55	504	
PA	***95%***	96%	69%	40%	98%		
OA							***92%***
Kappa							0.88

Table A.19 Confusion matrix for Trial 4 (pixel-based RT) at Rock Creek

Classified Data	Woody	Green Herbs	Bare	Brown Herbs	Shadows	Water	Samples	***UA***
Woody	188	28	1	5	0	0	222	***85%***
Green Herbs	32	64	22	11	0	0	129	50%
Bare	1	0	45	3	0	6	55	81%
Brown Herbs	0	0	18	66	0	4	88	75%
Shadows	1	0	2	0	10	0	13	77%
Water	0	0	1	0	0	0	1	0%
Samples	222	92	89	85	10	10	508	
PA	***85%***	70%	51%	78%	100%	0%		
OA								***73%***
Kappa								0.63

Table A.20 Confusion matrix for Trial 4 (pixel-based RT) at Four Mile

Classified Data	Woody	Bare	Brown Herbs	Green Herbs	Shadows	Stressed	Samples	***UA***
Woody	67	0	8	15	1	1	92	***73%***
Bare	1	163	19	0	0	0	183	89%
Brown Herbs	9	29	109	1	0	0	148	74%
Green Herbs	46	0	0	20	0	0	66	30%
Shadows	0	0	0	0	9	0	9	100%
Stressed	0	0	3	0	0	9	12	75%
Samples	123	192	139	36	10	10	510	
PA	***54%***	85%	78%	56%	90%	90%		
OA								***74%***
Kappa								0.64

Table A.21 Confusion matrix for Trial 4 (pixel-based RT) at Stampede Pass

Classified Data	Woody	Bare	Water	Brown Herbs	Green Herbs	Shadows	Samples	***UA***
Woody	190	2	33	0	1	2	228	***83%***
Bare	3	140	0	4	0	0	147	95%
Water	0	0	34	0	0	0	34	100%
Brown Herbs	1	12	0	47	1	0	61	77%
Green Herbs	11	0	0	1	8	0	20	40%
Shadows	0	0	0	0	0	16	16	100%
Samples	205	154	67	52	10	17	505	
PA	***93%***	91%	50%	90%	80%	88%		
OA								***86%***
Kappa								0.80

Table A.22 Confusion matrix for Trial 4 (pixel-based RT) at South Fork Newaukum

Classified Data	Woody	Brown Herbs	Bare	LWD	Shadows	Samples	***UA***
Woody	283	11	0	19	0	313	***90%***
Brown Herbs	4	33	32	0	0	69	48%
Bare	1	5	72	0	0	78	92%
LWD	0	0	0	26	0	26	100%
Shadows	0	0	0	0	14	14	100%
Samples	288	49	104	45	14	500	
PA	***98%***	67%	70%	58%	100%		
OA							***86%***
Kappa							0.76

Table A.23 Confusion matrix for Trial 5 (object based SVM with rDSM) at Rock Creek

Classified Data	Woody	Green Herbs	Bare	Brown Herbs	Shadows	Water	Samples	***UA***
Woody	203	23	1	0	1	0	228	***89%***
Green Herbs	16	66	6	7	0	0	95	65%
Bare	0	1	58	1	0	10	70	83%
Brown Herbs	2	2	22	77	0	0	103	75%
Shadows	1	0	2	0	9	0	12	75%
Water	0	0	0	0	0	0	0	0%
Samples	222	92	89	85	10	10	508	
PA	***91%***	72%	65%	91%	90%	0%		
OA								***81%***
Kappa								0.74

Table A.24 Confusion matrix for Trial 6 (object based RT with rDSM) at Rock Creek

Classified Data	Woody	Green Herbs	Bare	Brown Herbs	Shadows	Water	Samples	***UA***
Woody	200	21	3	4	0	0	228	***88%***
Green Herbs	18	70	5	9	0	0	102	69%
Bare	0	0	59	2	0	5	66	89%
Brown Herbs	2	1	19	70	0	0	92	76%
Shadows	2	0	3	0	10	0	15	67%
Water	0	0	0	0	0	5	5	100%
Samples	222	92	89	85	10	10	508	
PA	***90%***	76%	66%	82%	100%	50%		
OA								***81%***
Kappa								0.74

Table A.25 Confusion matrix for Trial 5 (object based SVM with rDSM) at Four Mile

Classified Data	Woody	Bare	Brown Herbs	Green Herbs	Shadows	Stressed	Samples	***UA***
Woody	96	0	1	14	5	0	116	***83%***
Bare	1	173	21	0	0	0	195	89%
Brown Herbs	6	16	117	13	0	2	154	76%
Green Herbs	20	0	0	9	0	0	29	31%
Shadows	0	0	0	0	5	0	5	100%
Stressed	0	3	0	0	0	8	11	73%
Samples	123	192	139	36	10	10	510	
PA	***78%***	90%	84%	25%	50%	80%		
OA								***80%***
Kappa								0.72

Table A.26 Confusion matrix for Trial 6 (object based RT with rDSM) at Four Mile

Classified Data	Woody	Bare	Brown Herbs	Green Herbs	Shadows	Stressed	Samples	***UA***
Woody	113	0	3	14	8	0	138	***82%***
Bare	0	175	31	1	0	0	207	85%
Brown Herbs	3	17	103	5	0	3	131	79%
Green Herbs	7	0	0	16	0	0	23	70%
Shadows	0	0	0	0	2	0	2	100%
Stressed	0	0	2	0	0	7	9	78%
Samples	123	192	139	36	10	10	510	
PA	***92%***	92%	74%	44%	20%	70%		
OA								***82%***
Kappa								0.74

Appendix B: Classification Figures

Figure B.1 Drone-derived land cover classification maps of the Lacamas Creek mitigation site using pixel-based (left) and object-based (right) methods with an SVM classifier.

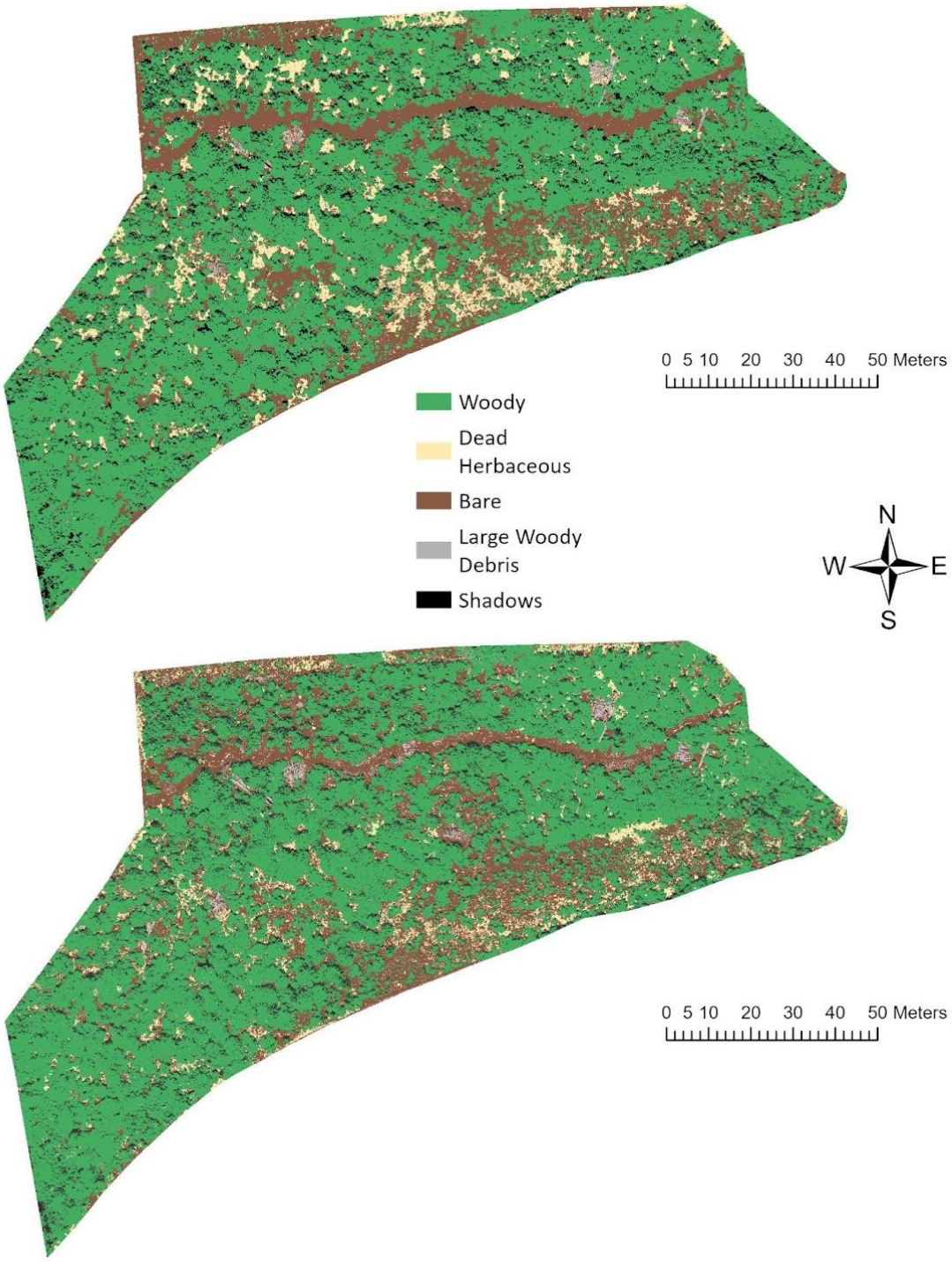

Figure B.2 Drone-derived land cover classification maps of the South Fork Newaukum mitigation site using object-based (top) and pixel-based (bottom) methods with an SVM classifier.

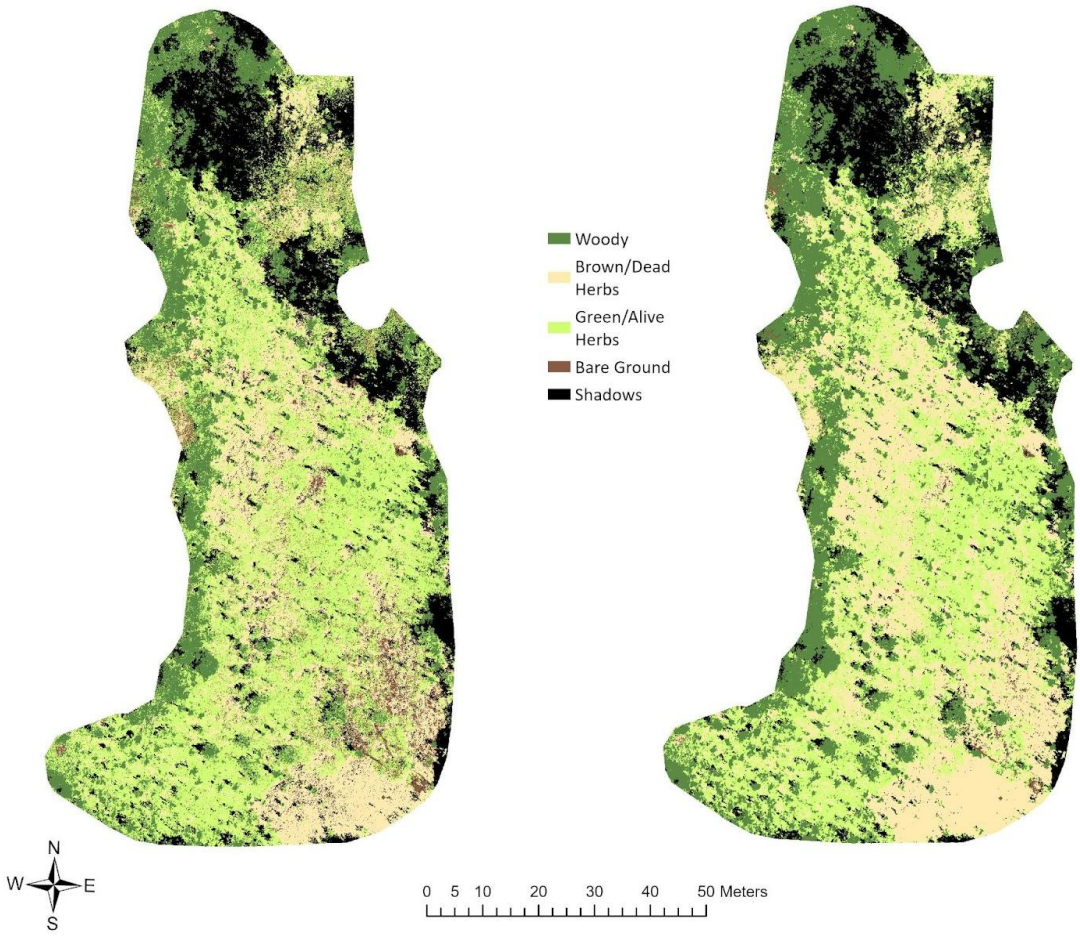

Figure B.3 Drone-derived land cover classification maps of the Mill Creek Complex South mitigation site using pixel-based (left) and object-based (right) methods with an SVM classifier.

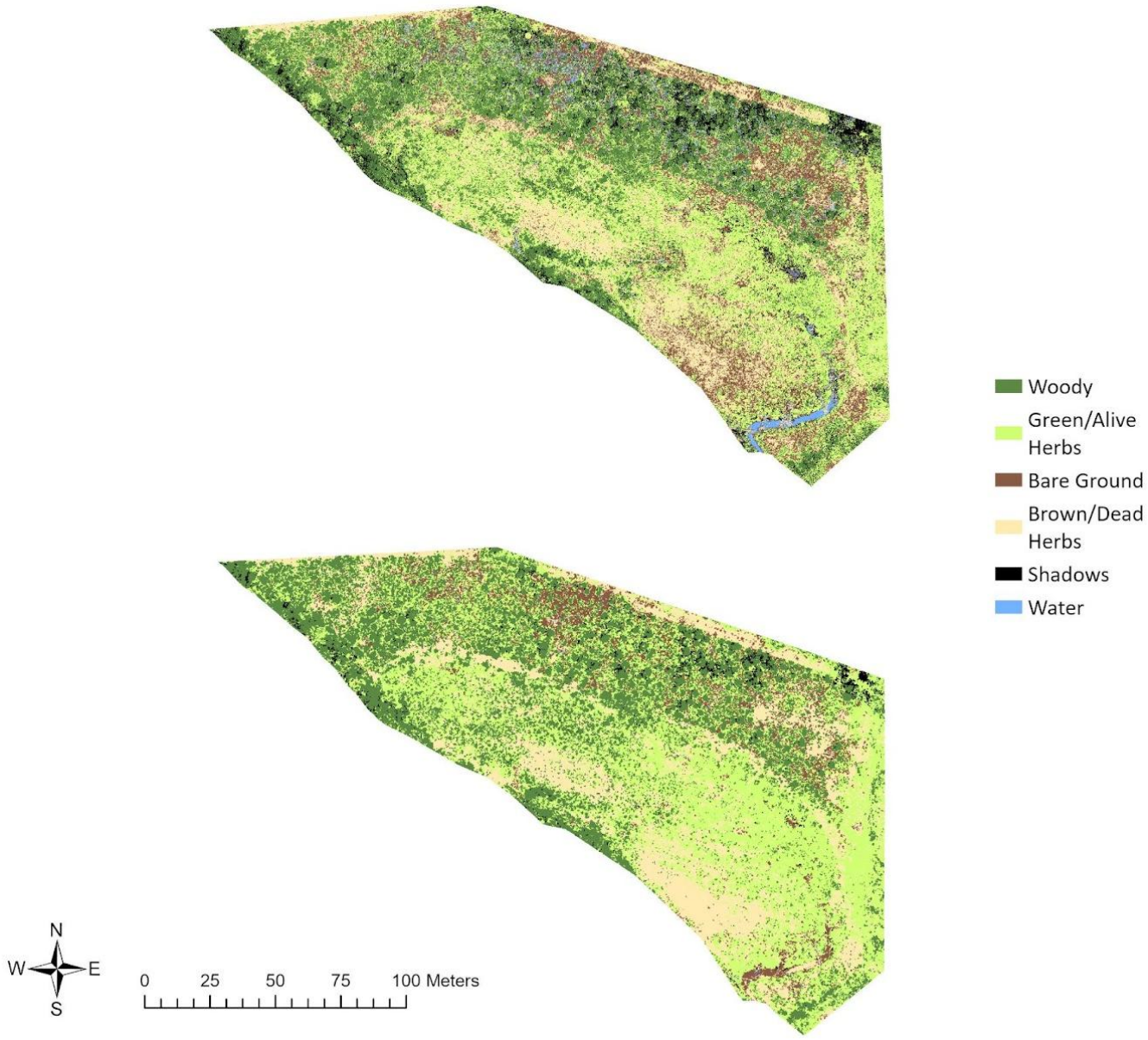

Figure B.4 Drone-derived land cover classification maps of the Rock Creek mitigation site using pixel-based (top) and object-based (bottom) methods with an SVM classifier.

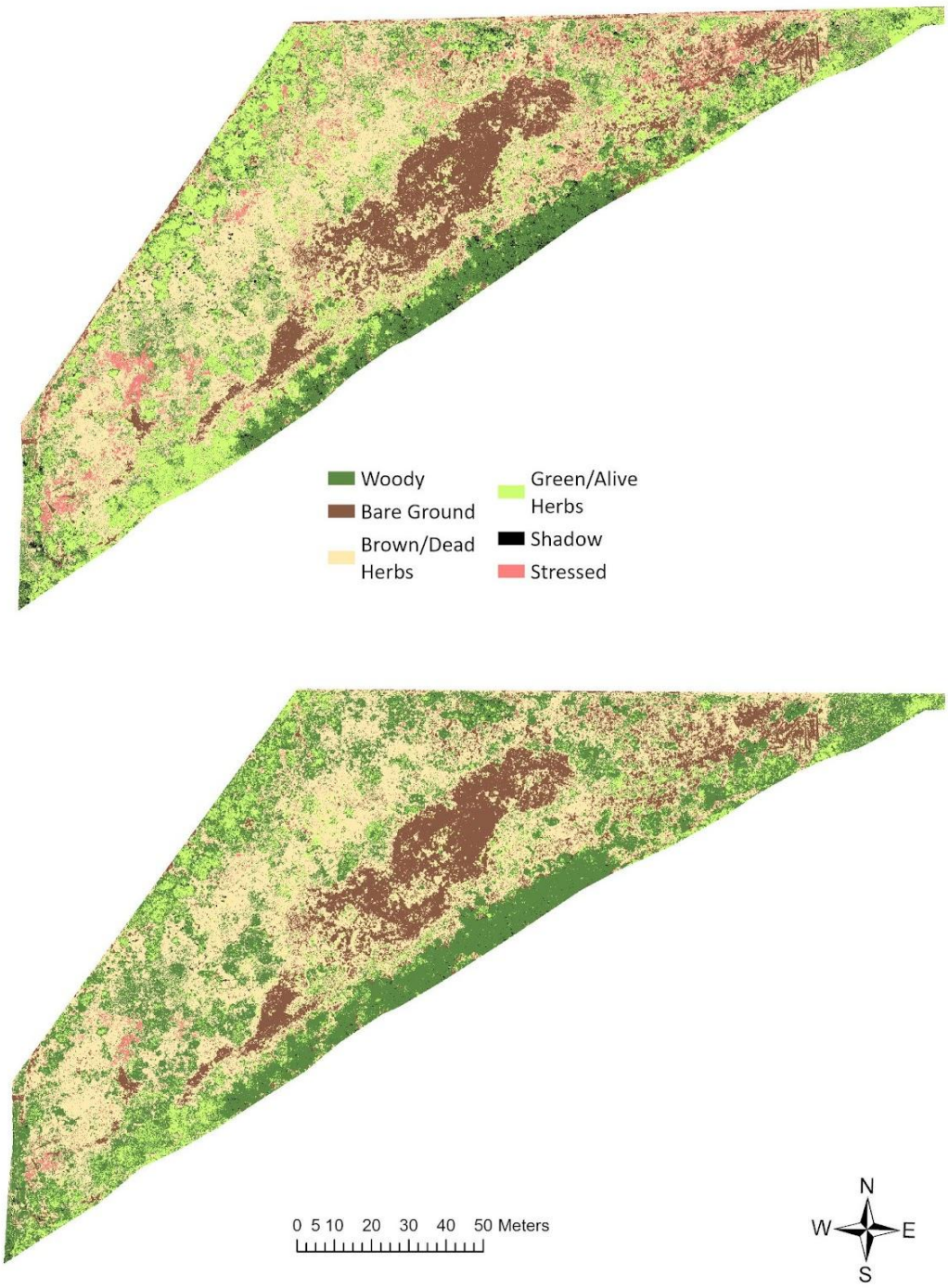

Figure B.5 Drone-derived land cover classification maps of the Four Mile site using pixel-based (top) and object-based (bottom) methods with an SVM classifier.

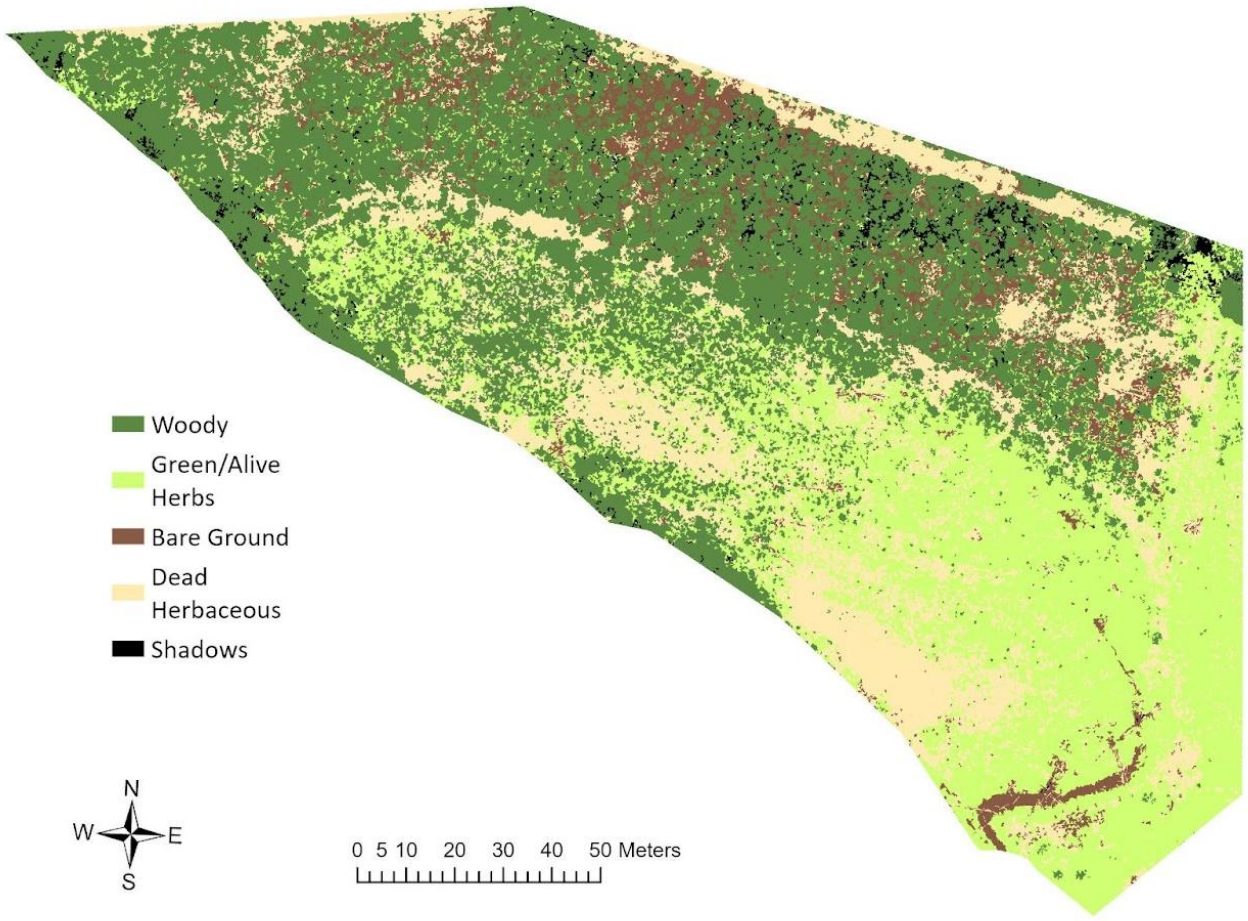

Figure B.6 Drone-derived land cover classification map of the Rock Creek mitigation site using an object-based classification method with an integrated relativized DSM raster and an SVM classifier.

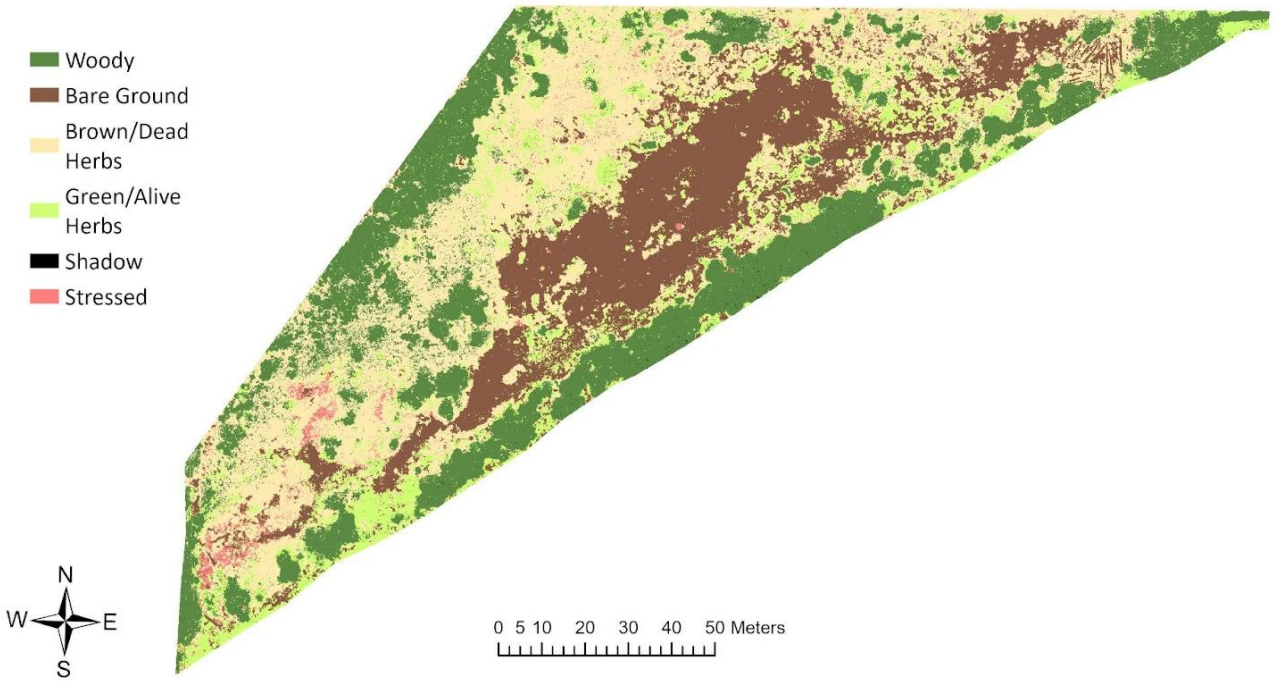

Figure B.7 Drone-derived land cover classification map of the Four Mile site using an object-based classification method with an integrated relativized DSM raster and a random tree classifier.