AN ABSTRACT OF THE DISSERTATION OF

Jonathan D. Burnett for the degree of Doctor of Philosophy in

Sustainable Forest Management presented on September 7, 2017.

Title: Environmental Remote Sensing with Unmanned Aircraft Systems

Abstract approved: ____

Michael G. Wing

Small Unmanned Aircraft Systems (sUASs) equipped with optical sensors are capable of remotely sensing landscapes and wildlife at spatial and temporal resolutions that were previously inaccessible due to technical and budgetary limitations. Conventional remote sensing and photogrammetric workflows can be applied to the resulting high resolution imagery to facilitate new types of scientific inquiry. This dissertation explores three novel applications using low-cost consumer grade and commercial grade sensors onboard an sUAS.

The first application uses a quadcopter equipped with a consumer grade camera to detect Swiss needle cast disease (SNC) in Douglas-fir stands in western Oregon. Swiss needle cast is a non-fatal foliar disease in Douglas-fir that reduces annual growth and stumpage value. Conventional detection methods rely on manned aerial detection surveys that are tedious. However, sUAS technologies offer a potential alternative. The presented method fuses sUAS technology with Structure from Motion, automatic stem segmentation and binomial classification with generalized additive models. Four 1.6 ha sites containing more than 3500 Douglas-fir trees were surveyed with a sUAS. Visibly infected trees were distinguished from not visibly infected trees with much greater than random chance (kappa > 0.4) at the four sites surveyed. Near-infrared (NIR) information was not pertinent to successful SNC detection, vastly simplifying the operational complexity of future surveys. The method described in this chapter facilitates mapping of individual Douglas-fir trees infected with SNC in the mountains of western Oregon.

The second application of sUAS technology expands upon the first by adding a narrowband multispectral camera (NMC), additional survey sites, and surveys in different years and months. The effectiveness of narrowband multispectral cameras for assessing vegetation condition has been heavily researched, but the application to Swiss needle cast detection in an industrial forest has not been previously described. Eight 1.6 ha sites encompassing more than 6000 trees were surveyed with a consumer grade camera and a NMC in 2015 and 2016. SNC detection reliability tended to be better with the NMC (kappa difference > 0.10) than the consumer grade camera when surveys were conducted in fully sunny conditions, but the differences were negligible in cloudy conditions. Summer imaging with the NMC yielded highly variable results in comparison to the more stable spring surveys and suggests that summer surveys are not operationally plausible. Detection surveys of the same sites in two different years revealed higher-than-expected levels of disease status change between years. Employing stricter probability thresholds on the classification rules reduced detected change from > 200 trees/site to < 50 trees/site at the cost of creating a third class of trees having an uncertain disease status. There was no evidence that foliage retention related to classified diseased status although additional study is recommended due to the limited inferential power afforded by the small sample size ($n < 28$). Many regulatory, technical, and computational hurdles must be overcome before large scale implementation of the method can be attempted.

The third application uses the integrated camera on a DJI Phantom 3 sUAS to conduct photogrammetric measurements of baleen whale morphology, which is an indication of whale health. UAS photogrammetry has been previously explored and shown to produce accurate measurements, but methods between surveys vary widely, indicating a need for standardization. We imaged 89 gray and six blue whales with a Phantom 3 sUAS. Whales were measured within the images and scaled to metric units using barometric altitude. Linear mixed models with error terms for flight and date were used to to correct scaling error. Post-correction estimates of 1 m calibration object contained 0.17 m less error and 0.25 m less bias than no correction. Total propagated uncertainty analysis was used to examine error contributions from scaling and image measurement (digitization) to determine that digitization accounted for 97% of total variance. Additionally, we present a new body size metric termed Body Area Index (BAI). BAI is scale-invariant and is independent of body length ($R^2 = 0.11$), enabling robust comparisons of body size within and among populations, and over time. Along with this study we present a three-program analysis suite that measures baleen whales and applies scale corrections to produce 11 morphometric attributes from UAS imagery. The photogrammetric method presented and associated software facilitate efficient and standardized analysis of any whales that meet the assumption of a parabolic shape.

Environmental remote sensing with sUAS can produce survey data at very high detail (*i.e.*, tree-level) and provide high measurement precision without the use of high-cost sensors. However, regulatory limitations within the United States National Airspace combined with the low-endurance of most multirotor sUASs limits efficient use to small areas, or one or two whale sightings. sUAS survey data is of such high resolution that data storage and management because burdensome even when survey areas are small. Furthermore, low-cost sUAS systems suffer from reliability challenges and steep learning curves that can heavily limit technology accessibility. In spite of the tradeoffs relative to manned surveys, sUAS remote sensing provides researchers with unprecedented access to data of high temporal and spatial resolution at low costs without putting human lives into the air.

Environmental Remote Sensing with Unmanned Aircraft Systems

by

Jonathan D. Burnett

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 7, 2017
Commencement June 2018

Doctor of Philosophy dissertation of Jonathan D. Burnett presented on September 7, 2017.

APPROVED:

Major Professor, representing Sustainable Forest Management

Head of the Department of Forest Engineering, Resources and Management

Dean of the Graduate School

Jonathan D. Burnett, Author

ACKNOWLEDGEMENTS

I want to take this opportunity to acknowledge those who made this possible. Michael Wing and Paul Anderson, your thoughtful and continual mentorship always gave me what I needed precisely when I needed it. Committee members, your patient support saw me through to the end. Seth Johnson and Ryan Benz, your engineering and flight expertise put my ideas into the air. Mom and Dad, your insistence on education started me on this path. Jen, my lovely wife, you carried the lion's share of this burden. Thanks for getting me to the other side. Finally, to my son Rowan and my daughter Ellie, you ignited the fire that motivated me to finish.

CONTRIBUTION OF AUTHORS

Dr. Dave Shaw assisted with the grant writing, study design, scope of inference and writing of Chapter 2 and 3. Ben Smith assisted with the grant writing and image interpretation for Chapters 2 and 3. Dr. Leigh Torres assisted with the study design and writing of Chapter 4. Leila Lemos and Dawn Barlow assisted with the measurement of whales and writing of chapter 4. Finally, Todd Chandler managed boat and drone operations, and provided logistical support for the study described in Chapter 4.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF APPENDIX FIGURES

Figure Page

LIST OF APPENDIX TABLES

Table | Page

1 Introduction

Small Unmanned Aircraft Systems (sUAS) technology has veritably exploded (Watts et al., 2012) since budget-strapped ecologists first fitted low-cost autopilot systems onto remote-control airframe chasses (Koh and Wich, 2012; Wing et al., 2013, 2014). Do-it-yourself (DIY) systems like those employed by Koh and Wich (2012) are surprisingly sophisticated, and are capable of producing georeferenced orthomosaics of a landscape at very high resolutions (< 3 cm) when equipped with a low-cost camera (Wing et al., 2014). DIY solutions are especially well-suited for carrying non-standard sensors that are common place in remote sensing applications research (Zarco-Tejada et al., 2012; Candiago et al., 2015; Patrick et al., 2017). Paralleling the evolution of low-cost DIY systems is the rapidly evolving low-cost commercial (*i.e.*, "prosumer") sUAS segment. The low-cost commercial segment is characterized by turn-key systems that are capable of pre-planned autonomous flights or fly-by-wire manual control (Wing et al., 2014), integrated gimbal stabilized camera systems, and have endurances (*i.e.*, flight times) of approximately 20 minutes (González-Jorge et al., 2017). DJI is presently the dominant player in this industry with their Phantom series sUASs (González-Jorge et al., 2017), although 3DR solo has been employed to similar effect (Bedell et al., 2017). Small UASs from both manufacturers typically cost less than $2,000 USD, although more sophisticated (and expensive) systems are available (González-Jorge et al., 2017).

Small UASs are increasingly accessible to the research community, especially in the United States where recent changes in regulations have opened class "G" airspace to licensed sUAS operators (Pomfret, 2016; FAA, 2016). Small UASs are uniquely suited to the research community because the relatively small scope of efficient utilization (Wing et al., 2014; Martinelli et al., 2015; Tang and Shao, 2015) is typically well within the spatial extent of a research plot. More importantly, sUASs give the researcher direct control of the timing of flights and specific parameters related to flight, specifically velocity and altitude (Wing et al., 2014).

Two of the environmental remote sensing fields that stand to benefit sUAS technology are agricultural remote sensing and, by extension forestry and wildlife remote sensing. The commonality between these two fields are that subjects of interest tend to be in remote locations with study areas that are described in hectares or even square kilometers (*e.g.* extent of Swiss Needle Cast in Oregon (Kanaskie et al., 2007) or humpback whale populations in the Pacific (Christiansen et al., 2016)), and the subject of interest trends to have both a spatial and a temporal component that limits static viewing (*e.g.* elk survey (Otten et al., 1993)) that ultimately makes pre-planned surveys (*e.g.* bridge surveys (Gillins et al., 2016)) difficult to implement.

However, utilizing sUAS technology for remote sensing is a complex endeavor (Wing et al., 2014; Hugenholtz et al., 2016; Carbonneau and Dietrich, 2017) that requires skills in mission planning, diagnosing electrical anomalies, and equipment configuration. Furthermore, there are additional components such as pilot training, maintenance, record keeping, and numerous safety considerations that contribute to complexity. As such, sUAS technology is not appropriate for every situation. Large scale employment requires careful consideration of not just the science questions, but also the limitations imposed by the equipment and the operational environment (Hayhurst et al., 2016). This dissertation examines the efficacy of sUAS remote sensing methods in the context of Swiss Needle cast detection, a foliar disease in Douglas-fir, and accurate measurement of baleen whale body condition.

Agricultural and forest remote sensing for disease detection is becoming an increasingly studied area of interest (Everitt et al., 1999; Franke and Menz, 2007; Martinelli et al., 2015). The interest stems from the economic ramifications of diseases on cash crops (*e.g.* hazelnut blight (Olsen, 2002)) and forests (*e.g.* emerald ash borer (Poland and McCullough, 2006)). Furthermore, understanding of the timing and extent of diseases that affect vegetation may inform future management practices in the face of global climate change (Millar et al., 2007; Sturrock et al., 2011). Historically, disease detection surveys have used manned aircraft (Johnson and Wittwer, 2008). However, these types of surveys tend to be 'dull, dirty, and dangerous' and thus present a potential niche for sUAS remote sensing (Watts et al., 2012). Small UAS technology is already being adopted by researchers investigating foliar diseases in plants and trees.

Calderón et al. (2013) used a hyperspectral scanner and thermal sensor onboard a sUAS to assess *Verticillium* wilt severity levels on olives. They found a positive relationship ($R^2 = 0.83$) between disease severity and the Crop Water Stress Index derived from sensor data. Albetis et al. (2017) employed a sUAS equipped with a multispectral camera to detect *Flavescence dorée* in grapes and found that successful detection was possible when using a combination of spectral and biophysical predictors. Patrick et al. (2017) investigated the feasibility of detecting tomato spot wilt on peanuts using a multispectral sensor onboard a sUAS. They found strong correlations between field observations and vegetation indices, and further concluded that a normalized differenced vegetation index using the red and red edge bands resulted in the most successful detection. However, the aforementioned studies were executed in easily accessed agricultural fields and orchards. Expanding these types of surveys to remote and mountainous forest settings like those found in the western United States adds additional challenges, especially in the context of the requirements of sUASs to be in line of sight (between aircraft and pilot) and fly no higher than 120 m AGL (FAA, 2016). Few studies have employed sUASs in the context of more natural forest settings (Koh and Wich, 2012; Wallace et al., 2016), and none of those have investigated foliar disease detection. Equally understudied is the application of sUAS remote sensing, and by extension sUAS photogrammetry, to studying baleen whale morphometrics.

Morphometric comparison of baleen whale body condition, across individuals and over time can reveal reproductive state, offspring growth rates, energetic capacity, body size demographic structure, and incidents of compromised health due to injury (Lockyer, 1986; Perryman and Lynn, 2002; Lockyer, 2007; Christiansen et al., 2016). Biologists have been measuring whales using photogrammetric methods since the 1980's (Klimley and Brown, 1983). The subject has been explored heavily in the subsequent decades (Dawson et al., 1995; Perryman and Lynn, 2002; Fearnbach et al., 2011). However, timing aerial surveys to coordinate with sea conditions, cloud conditions and optimal imaging conditions can be challenging, especially on migratory species such as gray whales (Perryman and Lynn, 2002).

More recently whale biologists have employed sUASs to measure whale morphology from the convenience of a boat. Durban et al. (2015) imaged killer whales with a DIY sUAS and estimated measurement bias to be 0.05 m with a standard deviation of 0.29 m. Following this study, Christiansen et al. (2016) conducted a much broader sUAS measurement campaign on humpback whales using a Splashdrone. Over 200 individuals were imaged and measured, and after thorough sensitivity analysis of error within and between images, it was determined that error did not prohibit precise body condition measurement. However, the method was highly dependent on coincidental imaging of both the reference object and the whale to ensure accurate image scaling.

Christiansen et al. (2016) and Durban et al. (2015) used different calibration methods and different whale measurement methods, highlighting a need for standardizing methods. Furthermore, a clearer presentation of methods and a more specific investigation into measurement error is needed to provide instruction to impending swarm of whale biologists who will use these techniques.

While seemingly divergent, these two topics provide a setting to explore the breadth of potential offered by sUAS systems within the broader field of environmental remote sensing. Swiss needle cast (SNC) disease presents an interesting case study for investigating the efficacy of sUAS disease detection surveys in natural setting because it affects a large population of Douglas-fir in the Oregon Coast range (Ritóková et al., 2016). Similarly, remote sensing of baleen whale morphometrics presents an interesting case study for investigating the metric accuracy of photogrammetry from low-cost sUASs in a dynamic oceanic setting.

The goal of this dissertation is to investigate the efficacy of using small unmanned aircraft systems to conduct remote sensing surveys in complex environmental conditions ranging from oceans to forested mountains using consumer grade, DIY and commercial equipment. The specific objectives are to: examine the efficacy of conducting SNC disease detection surveys with a sUAS, (2) determine if SNC detection accuracy varies by season, sensor and year, and (3) develop a method for conducting accurate morphometric analysis on baleen whales using a low-cost commercial sUAS. Objectives 1 – 3 of the dissertation are addressed in Chapters 2 – 4, respectively.

2 Individual Tree Disease Detection using a sUAS and a Consumer-grade Camera: A Case Study on Swiss Needle Cast Disease in Douglas-fir

Jonathan D. Burnett
Dave Shaw, Ben Smith, Michael G. Wing

Submitted to Forest Science

2.1 Abstract

A growing number of studies have investigated UAS remote sensing as a means of detecting foliar diseases in trees with a focus on cash crops in agricultural settings. In this study, a color camera and a near-infrared (NIR) converted camera were mounted on a UAS to image Douglas-fir stands infected with Swiss needle cast (SNC) disease at four sites in the mountains of northwest Oregon. The conventional method of SNC detection is via human interpreter onboard a manned aircraft with results at the acre-level. The aim of this study was to present a method for detecting and mapping SNC at the tree-level and compare disease detection reliability and accuracy among 36 different classification models for each site. Classification models were constructed from nine different spectral band combinations, some of which included NIR, and four different classification algorithms: randomForest, generalized additive models, generalized linear models (GAM), and stochastic gradient boosted machines. Kappa coefficients were > 0.58 for the best performing model at each site and positive predictive values for the best models were all > 0.70, indicating strong evidence of reliable and reasonably accurate SNC detection. No singular classification model emerged as the best, although GAM models were the most consistently reliable with the 22 of the 36 GAM models having kappa coefficients not significantly different from the best model at each site. Kappa coefficients of models that included NIR were not significantly different from those not including NIR, suggesting that NIR cameras may

not be necessary for SNC detection. Infection counts and tree-level disease maps were subsequently produced for each site to demonstrate the potential utility of the method for informing management operations. Several recommendations for broader implementation of method were identified, to include the need for beyond-line-of-site UAS operations and high endurance aircraft.

2.2 Introduction

Swiss needle cast (SNC) is a foliar disease in Douglas-fir caused by the native fungi *Phaeocryptopus gaeumannii* (Hansen et al., 2000). SNC has been intensifying in Oregon's Coast Range since the 1980's (Black et al., 2010; Ritóková et al., 2016). The disease is of special concern to the region and the state of Oregon because Douglas-fir (*Pseudotsuga menziesii*) is the major lumber producing species in the state (Brandt et al., 2006), and contributes heavily to the State's $12.7 billion annual industrial forest output (OFRI, 2012). SNC causes premature leaf abscission that reduces annual growth increment and can accumulate to end-of-rotation volume losses as high as 50% (Manter et al., 2000; Maguire et al., 2002). These projected volume losses lead to significant reductions in stand net present value (Kimberley et al., 2010). The dominant management strategy is to implement loss mitigation strategies such as stand conversion or thinning treatments that favor non-susceptible tree species such as hemlock or western redcedar (Shaw et al., 2011; Zhao et al., 2015).

The conventional method of mapping SNC is by aerial detection survey (ADS) using trained observers in an aircraft who map disease severity across western Oregon (MacLean and MacKinnon, 1996; Kanaskie et al., 2007; Johnson and Wittwer, 2008). A major advantage to ADS is the ability to immediately apply expert knowledge across a broad area and rapidly derive actionable products that describe disease extent and severity. However, SNC intensity in an area must sufficiently high for observers to detect from high altitude (> 400 m) while moving at velocities in excess of 130+ kph. Also, the spatial accuracy of results are at the sub-hectare level and not sufficient for individual tree assessments that are better suited for stand-level decision making. Furthermore, the spring phenology of the disease signature aligns with Oregon's rainy season which can limit manned flights and create poor canopy illumination for SNC detection. The use of small Unmanned Aircraft Systems (sUAS) offers a potential solution to these limitations by removing humans from the aircraft, providing operational flexibility to operate in conditions where a manned survey is not possible, conducting near canopy observations (120 m above ground level), and collecting data that are compatible for using supervised classification with modern machine learning algorithms (MLAs) to determine if individual trees contain visible signs of SNC infection.

A UAS equipped with a consumer-grade digital camera and flying lower than 120 m above ground level (AGL) can produce images with 2.5 cm ground sampling distance (GSD) (Wing et al., 2013). The images can be used individually for conventional photo interpretation and if there are enough overlapping images they can be compiled into a single georeferenced orthomosaic and digital surface model using Structure from Motion (SfM) (Verhoeven, 2011). Orthomosaics, especially when produced from linear sensors calibrated to surface reflectance, can be used in conventional remote sensing workflows that can attribute a classification (*e.g.* diseased/non-diseased) to pixels or objects using supervised classification (Laliberte and Rango, 2009; Liaw et al., 2002; Lillesand et al., 2014). The ability to use supervised classification methods for SNC detection has the potential to reduce analysis time and reduce error by allowing a trained interpreter to use their expertise to classify hundreds of training trees as diseased or non-diseased and extrapolate those training observations to subsequently classify tens of thousands or potentially even millions of trees. Furthermore, it is possible to add near-infrared cameras (NIR) to sUAS (Verhoeven, 2010). The addition of NIR has been shown to significantly improve the detection of foliar disturbance in trees and plants (Franke and Menz, 2007; Sankaran et al., 2010; Eitel et al., 2011).

Machine learning algorithms (MLAs) have been increasingly used in environmental remote sensing classification applications (Moisen et al., 2006; Rodriguez-Galiano et al., 2012; Lary et al., 2016; Gilbertson et al., 2017). MLAs relax the underlying assumptions of conventional parametric classifiers such as maximum likelihood, kmeans, or logistic regression classifiers to potentially improve classification accuracy at the cost of substantially increased computational time (Liaw et al., 2002; Wood, 2017). Furthermore, modern personal computing power has evolved such that a supercomputer is not necessary to utilize MLAs in many applications; for example, we used a used a Lenovo P50 Thinkpad Laptop for this study.

Classification of environmental remote sensing data frequently makes use of vegetation indices (VIs). VIs can simplify data complexity by reducing the number of explanatory variables through transformations, and can be related to a physiological plant component. The most commonly used VI is the normalized difference vegetation index (NDVI) (Carlson and Ripley, 1997). NDVI has been used to assess leaf chlorophyll content and greenness for numerous applications (Yoder and Waring, 1994; Carlson and Ripley, 1997; Zarco-Tejada et al., 2012). However, NDVI is derived from red and NIR data, the latter of which requires a specialized sensor. (Verhoeven, 2011). A similar and newer VI is the triangular greenness index (TGI) proposed by Hunt et al. (2011). TGI is similarly a measure of greenness, but has been shown to be sensitive to varying chlorophyll content in leaves, even when high leaf area results in saturated NDVI. In addition, TGI does not require a specialized NIR camera.

A few existing studies have used UAS remote sensing approaches to investigate foliar disease in tree plantations. Garcia-Ruiz et al. (2013) used MLAs in conjunction with multispectral imagery collected from a small UAS to detect Huanglongbing in crowns of a citrus tree plantation. This study is particularly pertinent to ours because Huanglongbing can produce a chlorotic response in infected leaves, similar to that of Douglas-fir leaves infected with SNC, however, the multispectral camera employed is much more expensive and specialized than consumer-grade cameras. Calderón et al. (2013) employed UAS equipped with a highly sophisticated hyperspectral scanner to conduct early detection and severity assessment of *Verticillium* wilt in two olive orchards and found that wilt could be reliability detected at the crown level. Their findings highlight the extent of disease detection that is possible with UAS although the method was applied to a limited setting (*i.e.*, two olive plantations), and required a specialized sensor that is too expensive to be readily accessible to many research or forest management entities.

2.2.1 Objectives

Our study investigated the efficacy and reliability of detecting SNC presence in Oregon coast Douglas-fir trees with a sUAS remote sensing technique. Tree-level disease detection methods are understudied and applying such a technique in combination with a low-cost UAS equipped with a low-cost consumer-grade camera in a mountainous industrialized setting is a unique approach that may be of interest to the research and commercial forestry communities alike. The primary goals of our study are to present a methodology that produces SNC detection maps for surveyed areas and conceptually explore the concept of broader implementation. The specific objectives to accomplish our goal are to: (1) examine whether trees can be classified as diseased or non-diseased with better than random chance, (2) discern if the use of a NIR camera significantly improves classification reliability, (3) determine whether physiologically meaningful VIs improve classification reliability, and (4) compare SNC detection reliability among two different MLAs and two different GLM implementations using nine different model specifications to examine the effect of classification model on detection reliability.

2.3 Methods

2.3.1 Study Area

Four 1.6-ha study sites were selected from areas in the Swiss Needle Cast Cooperative plot network (Ritóková et al., 2014). Criteria for site selection was based on prior knowledge of SNC presence, coverage from an existing Federal Aviation Administration Certificate of Authorization (COA) and the ability to maintain line of sight (LOS) to an unmanned aircraft flying 120 m AGL. Sites contained Douglas-fir cohorts ranging between 7 and 40 years of age and an elevation gradient between 2 and 100 m. Overhead images of the sites provide a sense of each site's vegetative condition as well as a visual context of each site's location (Figure 2.1). The Norton site was distinct from the other sites because it was a common garden planted for a USFS Douglas-fir seedling movement trial and was known to contain *Rhabdocline spp.,* a lethal needle cast disease that produces phenotypic responses (*e.g.* needle senescence in spring) similar to SNC (Stone, 1987; Hansen et al., 2000). We did not consider the confounding influence of *Rhabdocline spp.* a concern because it is unlikely a quantitative classifier could discern the difference between such similar disease signatures and the presence of *Rhabdocline spp.* would be of extreme interest to a forest manager because it is a known mortality agent. The survey was conducted in May 2015 to correspond with the phenological response of Douglas-fir to SNC infection (Manter et al., 2000) as well as the annual SNC ADS.

2.3.2 Equipment

The UAS selected for this study was a TurboAce Matrix quadcopter (Appendix Figure A.1) valued at $5000 USD with a motor-to-motor span of 800 mm and an all-up weight of 3 kg, including the battery, camera, and two-axis gimbal. The UAS was equipped with a Pixhawk autopilot capable of conducting GPS navigated flight. The sensor used for this study was a gimbal-stabilized Sony Nex 5T camera ($750 USD) equipped with a 20-mm focal length lens for collecting visible light in the red (R), green (G), and blue (B) wavelengths. An identical camera was converted to be sensitive to NIR wavelengths (Verhoeven, 2010). This resulted in a camera that was sensitive to G, R, and NIR wavelengths. The combination of the two cameras allowed the collection of imagery containing R, G, B and NIR data. Two cameras were used in tandem because the G and R channels on the converted camera are polluted by NIR (Verhoeven, 2010), thus confounding subsequent derivation of VIs and the accurate estimation of R, G, and B importance on reliable SNC detection.

Flights were planned on a ground control station (GCS) laptop running Mission Planner software (Oborne, 2015). Mission Planner estimates GPS navigation points along a grid to ensure previously specified overlap/sidelap can be achieved based on UAS altitude, and the following camera parameters: firing interval, focal length, sensor size in metric units, and number of pixels on the sensor. We used overlap and sidelap settings at both 80% to ensure there were enough images to conduct an SfM reconstruction of the scene (Agisoft, 2013). Flights were conducted at 120 m above ground level (AGL), resulting in images with 2.5 cm ground sampling distance (GSD), where GSD is the horizontal projected pixel area across the earth's surface (McGlone, 2013). Two flights were conducted per site within

15 minutes of one another to minimize sun-angle change because only one camera could be flown at a time.

Four to six 1 m x 1 m iron cross ground control points (GCPs) were placed at each study site to facilitate georeferencing of the final orthomosaics. The number of GCPs was limited by forest openings, road locations, and the total area that could be covered while ensuring visual LOS to the UAS. Two 0.6 x 0.6 m drywall panels painted with titanium dioxide and having a reflectance of 90% were placed in the imaging area of each site for calibrating all imagery to a common standard. All flight operations were conducted under Certificate of Authorization 2014-WSA-212-COA and were in accordance with all applicable Federal and State regulations. The specific timing of flights was scheduled to take place within +/- 1 hour of solar noon to minimize the effect of sun shading (McGlone, 2013); however, equipment malfunctions and rain showers delayed surveys at N02 and N154 to approximately 1500 local time. Flights were limited to either fully sunny conditions or fully cloudy to ensure consistent solar radiance during the surveys.

2.3.3 Processing and Analysis

The analysis of the imagery required the development of a three-stage workflow to extract the information necessary to conduct tree-level disease classification from UAS imagery. The primary components of the workflow are illustrated in Appendix Figure A.2. The following summary will only describe the most pertinent elements.

2.3.4 Stage 1: Pre-processing

Pre-processing in the context of this study comprised the steps and sub-routines required to produce initial data from UAS images for follow-on processing and analysis. This stage converted the raw image data into a multi-band ortho-rectified mosaic (OM) and digital surface model (DSM) that facilitated the creation of geographically explicit vector data used for subsequent disease classification and diseased tree mapping. Raw image processing was a multi-step process to convert the raw Sony demosaicked 12-bit intensity images into a three-color band (RGB) JPEG format image using Adobe Photoshop (Knoll et al., 2015). We used Photoshop because it contains lens distortion and chromatic aberration correction models for the specific lens and camera model used in this study, and will subsequently apply corrections to all images. Photoshop also facilitates manual hotspot correction (Pixel, 2011) and tuning of exposure (*i.e.*, gain) to maintain consistency among all images.

We created a custom linear profile to prevent the software from applying the sRGB linear gamma function (Gasparini and Schettini, 2004) when converting to JPEG format. An 8-bit JPEG format was chosen over 16-bit TIFF format to reduce memory usage and processing time in Photoscan and to reduce disk storage requirements for the thousands of images collected over the course of this study. Empirical comparison of linear images processed using the method described in Verhoeven (2010) was compared to our method with a qualitative (*i.e.*, visual) comparison of histograms for ten images from one site and was determined to be similar enough to justify the use of linear profile JPEG format images. Processed images were geotagged using the autopilot flight log and existing functionality in Mission Planner software.

Mosaic Processing combines the geotagged images into an OM containing R, G, B, and NIR grids using SfM. SfM is analogous to conventional soft-copy photogrammetry, but uses a multi-view stereo approach that is based on computer vision algorithms to identify conjugate points. Conjugate points are used to geometrically reconstruct a scene in 3D space and subsequently produce both a DSM and OM (Westoby et al., 2012). We used Photoscan (Pasumansky, 2017) for SfM processing due to the customizability and general acceptance in environmental sciences (Fonstad et al., 2013; Holman et al., 2016). Photoscan processing generally followed the method outlined in UNAVCO's tutorial (Shervals and Dietrich, 2016). Because collection of R, G, B, and NIR requires separate instruments, SfM processing was conducted individually for each site and camera type (*i.e.*, NIR and RGB). Hereafter RGB, will refer to the composited color data produced by the visible light and individual color band grids will be referenced using the appropriate letter acronym.

The RGB dataset was georeferenced using the GCPs. The NIR OM was registered to and composited with the RGB OM using the 'georeferencing' and 'composite' functionality in ArcGIS (ESRI, 2016). Color distortion and reduced resolution made georeferencing the NIR data directly challenging. The four-band composite was calibrated to a common reflectance standard using the empirical line calibration method (Smith and Milton, 1999) implemented into a custom script constructed in MATLAB (Mathworks, 2016). The reflectance standards were (1) the two 90% reflective panels described earlier and (2) shadows between trees (*i.e.*, 0% reflectance). Two VIs, NDVI and TGI were derived from the R, G, B, and NIR data in the OMs to determine if physiologically meaningfully indices improved classification reliability or simplified classification model specification to one or two variables. The first three principal components (PC1 – PC3) of R, G, and B were subsequently derived from the calibrated composite. Principal components were derived because the inherent collinearity of these variables can confound model fitting. The final results were the following individual spectral data rasters representing the spectral variables that were used in the subsequent classification models: 'R', 'G', 'B', 'NIR', 'TGI', 'NDVI', 'PC1', 'PC2', and 'PC3'.

2.3.5 Stage 2: Post-processing

Tree Segmentation began with the creation of a relative-height (*i.e.*, normalized) canopy height model (created in FUSION) in conjunction with both the SfM-derived DSM and a 10-m digital elevation model (DEM) sourced from the National Elevation Dataset (Gesch et al., 2002). Individual tree positions and circular crown areas were segmented from the CHM using FUSION CanopyMaxima (McGaughey, 2017) with the default parameterization. CanopyMaxima was chosen for the tree segmentation process because it has been shown to effectively extract dominant and co-dominant Douglas-fir trees from LiDAR surface models (Popescu and Wynne, 2004; Apostol et al., 2016) and it produces a geographically explicit tree list. Appendix Figure A.3 illustrates this process for site N02. Segmented tree points and crown areas were clipped to the site with a buffer to prevent half-trees and minimize the erroneous commission of trees resulting from edge effects. Segmentation commission errors were manually filtered in ArcGIS and omission errors were resolved by manually adding points and crown areas. Three hundred were randomly selected without replacement from the population of segmented trees for each site for the purpose of classification model training. Three hundred trees were selected to ensure there were sufficient observations to train models having five covariates without overfitting.

Segmented training trees were classified using photo interpretation of a 2.5 cm GSD OM for each site with the assistance of a trained ADS interpreter. The classification criteria included four levels (*i.e.*, classes): no visible sign of infection (NVSI), visible sign of infection (VSI), weak visible sign of infection (WVSI), and Other (Figure 2.2). The first three categories are intuitive while Other accounts for non-Douglas-fir trees, dead trees (*i.e.*, no leaves) and trees that cannot be readily categorized within the previous three classes. A convention using 'diseased' and 'non-diseased' was not appropriate because Douglas-fir trees have been shown to be infected with SNC without a visible response. Figure 2.2 illustrates the classification criteria for the scheme. Classification results were attributed to each training tree. Despite having four initial classes, we filtered out all training data that was not classified as NVSI or VSI to ensure training observations were pure, thus minimizing model confusion. Nested classification schemes or multinomial classification models were not used for this study due to the increased sample sizes required for robust model development, the added challenges associated with accounting for uncertainty, and because the primary study question was associated with a simple binomial response of VSI/NVSI (*i.e.*, diseased or non-diseased).

2.3.6 Stage 3: Model Fitting and Site Disease Classification

Spectral information was ascribed to each tree by averaging data from each spectral raster described previously within the crown area of each tree at each site. Spectral data were aggregated at the crown level because a pixel-level approach as applied in (Calderón et al., 2013) was not computationally feasible for the number of trees in our study nor is it realistic for broader implementation. Spectral data rasters were first resampled from native 2.5 cm GSD to 10 cm GSD to minimize processing time and memory utilization space without grossly over generalizing within crown variation. The 10 cm GSD was chosen to ensure > 20 pixels fall within the area of each tree crown area. Non-vegetated pixels (*e.g.* roads and shadows) were masked out of the spectral data rasters using NDVI thresholding. Mean intensity (*i.e.*, digital number) by spectral variable within each crown area was subsequently estimated and attributed to the associated tree.

2.3.7 Classification Models

The training data were used to construct binomial classification models for each site. Independent models were fitted to each site because local environmental conditions at each site were not consistent enough to justify fitting one model to all the data. We used four different supervised classification algorithms (Table 2.1), where generalized additive models (GAM) and generalized linear models (GLM) classified using parametric logistic regression methods with the commensurate underlying assumptions, and RandomForest (RF) and stochastic gradient boosted machine (GBM) were MLAs that conducted binomial classification using non-parametric methods with the primary assumption that covariates were not strongly correlated. These classification algorithms were selected for use in this study because they have been previously used in environmental remote sensing classification applications (Moisen et al., 2006; Prospere et al., 2014; Lopatin et al., 2016; Gilbertson et al., 2017). For clarity of the subsequent description, model parameterization refers to the non-varying components unique to each classification algorithm, and model specification refers to the combination of response and predictor variables used during model training. Pertinent parametrization variables and source references for each algorithm also appear in Table 2.1.

2.3.8 Model Specification and Parameterization

The values associated with parametrization variables were optimized to achieve the maximum Cohen's kappa coefficient (Congalton, 1991), hereafter referred to as kappa. Optimization occurs internally during model fitting and was conducted using the Caret package (Kuhn and Johnson, 2013; Kuhn, 2016) implemented in R (R Core Team, 2017). Caret constructs classification models from multiple combinations of parameters and subsequently discerns the combination resulting in the highest kappa. Kappa was chosen because it is a measure of reliability that estimates the proportion of agreement to observed response beyond agreement that is possible by chance alone (Sim and Wright, 2005); where chance in this study was determined by the proportion of VSI and NVSI trees within the training data. A kappa > 0 was indicative of a classifier that classified at better than random guess and a kappa = 1 was a perfect classifier. Kappa scores > 0.4 were considered reliable (Landis and Koch, 1977; Viera et al., 2005).

Nine different model specifications (Table 2.2) were identified to explicitly examine the effect of predictor combinations on classification reliability, where the predictors were the spectral variables described previously. Hereafter, the specifications in Table 2.2 will be referenced as S1, ..., S9.

There were notable differences among the selected classification algorithms, both in terms of function and in terms of parameterization, however, these were not the focus of this study. As such, we describe the relationship between the response and predictor variables for a given classifier using the general model form (Eq: 2.1).

$$y_{a_s} \sim \beta_{0_{a_s}} + \beta_{1_{a_s}}(x_{1_{a_s}}) + \ldots + \beta_{i_{a_s}}(x_{i_{a_s}}) \tag{2.1}$$

Where a is the classification algorithm (Table 2.1), s is the predictor specification (Table 2.2), i is i^{th} predictor and i = $1, \ldots,$ n when n is equal to the number of predictors in specification s, $y_{a_s}$is the binomial response of classification (*i.e.*, NVSI or VSI), $\beta_{0_{a_s}}$ is the intercept, and $\beta_{i_{a_s}}$ is a coefficient or function that relates predictor $x_{i_{a_s}}$ to the response.

2.3.9 Model Fitting

All combinations of classification algorithms and model specifications were fit to training data for each site, resulting in 144 classification models that are hereafter referred to as classifiers. Classifiers were evaluated using k-fold repeated cross-validation (CV) routines because CV is as effective as data-splitting without a reduction in inferential power (Harrell et al., 1996). Advanced spatial cross validation techniques such as spatial leave-one-out or (Le Rest et al., 2014) or spatial blocking (Roberts et al., 2017) were not employed because they were not implemented in the statistical analytical package we utilized; however, we did not expect CI estimates to be overly optimistic because there was no intent to extrapolate results beyond the spatial extent of each site. The caret package was used to conduct 30 repetitions of 10-fold CV for each classification algorithm, model specification, and model parameterization combination for each site to estimate mean kappa, and Positive Predictive

Value (PPV) statistics for each fitted model. PPV is the probability that any tree predicted to contain VSI truly contains VSI and was used to compare how well each classifier detected VSI (*i.e.*, SNC). We opted to do 30 repetitions of the *k*-fold CV because variance and bias have been shown to level off by 30 and more repetitions were computationally inefficient (Vanwinckelen and Blockeel, 2015). 95% CIs of kappa and PPV were derived from the 2.5% and 97.5% percentiles of the *k*-fold repeats. Using CIs to compare performance among models was appropriate because we ensured that repeated *k*-folds were identical for all models by using seeded randomizers (Vanwinckelen and Blockeel, 2015). Producing CIs by bootstrapping on *k*-fold repeats like Tong et al. (2012) was considered, but was determined to be too computationally expensive and produced similar intervals to the quantiles of repeats, albeit slightly less symmetric. The kappa coefficient cannot be compared across fitted models due to variations in individual classifier performance that change mkappa, so we calculated a proportion of kappa (*i.e.*, pkappa) by dividing kappa by maximum possible kappa (mkappa) (Sim and Wright, 2005).

2.3.10 Model Selection and Site-level Classification

Classifier combinations by site were quantitatively compared using the pkappa scores (reliability) and PPV scores (accuracy) to discern the best classifier for each site. PPV and Kappa The best classifier for each site was selected based on the highest minimum threshold that bounded the estimated pkappa score. The best classification model for each site was applied to the full population of segmented trees to assess VSI/NVSI for the entire site. Lastly, a spatial cluster analysis was conducted on the final results using the Density-based spatial clustering of applications with noise (Birant and Kut, 2007) implemented in R under the ‘dbscan’ package (Hahsler et al., 2017). The purpose of the cluster analysis was to group the VSI trees into generalized areas that might be used for guiding future management activities.

2.4 Results

A comparison of pkappa among classifiers within and between sites (Figure 2.3) suggested that there was not a single 'best' classifier, classification algorithm, or specification. The red bar indicates the best performer for each site, as defined by the classifier that had the highest lower confidence threshold. The best classifier and associated mean kappa with corresponding upper and lower CI for N02, N154, Norton and T01, were GAM + S4 (.57,.60,55), RF+S3 (.58,.60,.55), GBM+S2 (.86,.88,.83 and RF+S2 (.65,.68,.62), respectively. The blue bar indicates the confidence interval for the best performer at each site. The poorest performing classifiers were those based on S8 and S9; since pkappa for S8 was significantly lower than the best performer for 13/16 comparisons and pkappa for S9 was significantly lower than the best performer in all comparisons. The least consistent classifier was based on the GBM algorithm with 28/36 pkappa significantly lower than the best. Site specific pkappa was highly variable. The classifiers based on the parametric classification algorithms of GLM and GAM were the most consistent with 20/36 and 22/36, respectively, of classifiers producing pkappa scores not significantly different from the best performer. GAM + S3 and GAM + S5 were the only classifiers not significantly different from the best classifier at all sites, and of these, GAM + S5 was the most parsimonious. Subsequent results focus on the GAM + S2 classifier because there was no evidence that GAM + S2 differently from GAM + S3 or GAM + S5, it was more parsimonious, it did not include transformed predictors, and did not include any NIR predictors.

PPV was > 0.7 for all classifiers except those based on S8 and S9. PPV for the best classifiers ranged from 0.79 (N02) and 0.92 (Norton) and provided a measure of confidence that most trees classified as VSI were truly exhibiting VSI, regardless of site. The classifiers resulting in the highest PPVs did not always correspond with the classifiers having the highest pkappa because individual models were optimized to kappa and while related, the two metrics represented different components of classification performance. The estimated PPV for GAM + S2 classifier was never significantly different from the PPV of the best classifier at any site. We recognize that binomial classification results can also be described with receiver operating characteristic (ROC) curves, negative predictive performance, specificity, sensitivity and accuracy; however, these metrics all relate to PPV and

Kappa and we felt that presenting this information would provide unnecessary detail for our intended audience impede the presentation of results.

Variables of importance were compared for the best performing classifier at each site (Figure 2.4). There was no evidence to suggest that classifier reliability was improved by including NIR explicitly or indirectly (*i.e.*, NDVI) based on the following results: (1) three of four of the sites selected a best model that did not include a NIR specification (Figure 2.4), (2) there was no evidence that S3 performs any worse than the model specifications containing NIR (Figure 2.3), and (3) models based on NDVI (derived from NIR) consistently produced the least reliable classifications (Figure 2.3).

The number of trees classified as VSI for each site was estimated by applying the classifier based on the GAM + S2 classifier to the population of trees. Minimum and maximum likely SNC incidence rates were estimated based on the false positive rate. VSI rates of SNC range from 23% (T01 minimum) to 35% (N02 maximum) (Table 2.3). The number of trees classified as VSI ranged from 195 (T01) – 418 (N02). Maps of VSI trees (represented by triangles) show evidence of spatial clustering (Figure 2.5). Statistically significant spatial clustering was detected at all four sites with the number of clusters ranging from 7 (T01) to 14 (N02) and appear as polygons in the supporting maps (Figure 2.5).

2.5 Discussion

2.5.1 Classification Reliability

Despite the use of identical cameras, SNC detection reliability varied widely among classifiers and among sites. However, the best classifiers for each site, as well as the GAM + S2 classifier, produced kappa scores well above the 0.4 threshold that has been previously established as a metric for an acceptable classifier (Landis and Koch, 1977; Viera et al., 2005). SNC detection was most reliable at the Norton site; likely because many of the trees at the Norton site were exhibiting strong visible signs of infection from *Rhabdocline spp.* as well as SNC (Wilhelmi, 2016). The best classifier for the remaining sites were comparatively less reliable, likely because these were mature stands and variation of SNC phenology in mature crowns tended to be highly variable ranging from not visible, to subtle chlorosis and crown thinning, to highly chlorotic with a near-total loss of upper canopy foliage (Mulvey et al., 2013). While needle retention varies with height in the crown (Shaw et al., 2014) and greater needle retention in the lower crown could make overall assessment challenging. These irregularities can increase the likelihood of errors in the training data (*i.e.*, misclassification) or result in poor spectral separation between VSI and NVSI classes that produce models that are unable to distinguish between the two classes. Furthermore, the disease is not systemic so needles are infected individually. This can result in partial-crown damage (Mulvey et al., 2013) that may not be detectable when the spectral response is effectively attenuated by averaging spectral information across the crown area. A possible mitigation for this problem is pixel-level classification as was employed in Garcia-Ruiz et al. (2013) but the method is computationally intensive on 10 cm resolution imagery and may not be feasible for large area surveys.

The true positive rate of the best classifier at each site was > 0.75 and represented the probability that a tree classified as diseased was truly diseased. We can interpret the PPV as a type of certainty. Medical sciences use PPV and negative predictive value (true negative rate) to assess the validity of diagnostic tests (Parikh et al., 2008; Hajian-Tilaki, 2013). For example, the best performing classifier at site N02 detected 559 VSI trees. The PPV of the classifier was 0.76 to 0.80 (Table 2.3) and indicates that we expect 428 to 445 of these trees to be truly VSI. These results show that even at the worst performing site we have gleaned increased understanding about the intensity of the SNC at a level of detail not previously explored in literature. Furthermore, these tree counts could be used in conjunction with growth and yield modeling (Maguire et al., 2002) to determine if species conversion (Shaw et al., 2011; Zhao et al., 2015) is an economically viable alternative. Given the definition used for creating the VSI class, it is likely the presented method is most effective at detecting the most severely diseased trees, and is therefore most useful in assessing stands where soil conditions, climate and stand condition create a high risk scenario for SNC infestation.

2.5.2 Classifier Selection and NIR Importance

Examination of nine different model specifications (Table 2.2) and four different classification algorithms (Table 2.1) did not yield a superior combination (*i.e.*, classifier), although classifiers based on S8 and S9 were clearly the poorest performing. The poor performance of S8 and S9 indicates that neither chlorophyll content (TGI) or greenness (NDVI) adequately describe the variation in spectral signature between diseased and non-diseased crowns. The inadequacy may also be a function of the broad bandwidth of the camera color channels that ultimately may prevent detection of disease-related signatures the in green and red absorption bands (Clark and Lister, 1975; Nelson, 1983). The best classifier tended to change among sites, likely due to changes in light physics such as bidirectional reflectance. This remains a major source of uncontrolled error in UAS surveys (Rasmussen et al., 2016). Neither of the MLAs performed significantly better than the GAM or GLM classifiers. It is possible that restricting the RF and GBM models to 1000 trees was limiting to classification performance, however, more trees would have made the models computationally unwieldy, especially the GBM models. MLAs tend to be superior in situations where there are many predictors and the observation data do not conform to the underlying assumptions of parametric classifiers (Lopatin et al., 2016). MLAs may be suited to SNC detection when additional non-spectral predictors (*e.g.* dbh, aspect or precipitation) are included. While the results do not provide a conclusive decision on whether to incorporate MLAs for SNC detection we believe there is compelling evidence to recommend continued use of GAM classification algorithms.

Classifiers using NIR did not improve classification reliability or detection certainty above what was possible from models based on the most parsimonious specification (*e.g.* S2). The lack of NIR importance in any of the classifiers is likely associated with the wide sensitivity range of the converted camera (Verhoeven, 2011). Additionally, inconsistent NIR performance for foliar disturbance detection has been reported in previous studies (Sankaran et al., 2010). While our results do not support the use of a converted NIR consumer-grade camera for SNC detection, the value of a NIR sensor should not be negated without additional field trials using a narrow-band NIR sensor such as the MicaSense RedEdge (Patrick et al., 2017).

2.5.3 Disease Mapping

The primary objective of this study was to produce a map of individual trees that exhibited signs of SNC for each site. We used the GAM + S2 classifier to classify the population of trees for each site (Figure 2.5). Our rationale for the use of this classifier was described in the results section. The robust classification reliability and detection accuracy evaluation provided evidence that the SNC counts (Table 2.3) were not simply a random guess and thus can provide meaningful insight about SNC intensity across each site, despite the varied reliability among sites. The individual tree-level results facilitated the use of spatial clustering to identify areas containing high densities of infected trees. (Figure 2.5). These maps can be subsequently used to prioritize management operations.

2.5.4 Future Investigation

While our study presents a compelling method to detect SNC at the individual tree-level, we still lack understanding about: (1) detection consistency over time, (2) whether reliability and detection accuracy can be improved with a commercial grade sensor (*e.g.* MicaSense RedEdge), (3) if the methods employed facilitate reliable SNC detection during seasons (*e.g.* summer) when weather conditions are not limiting, (4) the quantitative relationship between needle retention and spectral signature, and (5) the functional relationship between needle retention and projected growth loss. Furthermore, there were limitations that prevent immediate implementation at a statewide scale. The data collected for this study took four field days to collect and months of preparation and processing that is similar to the effort required to fly the entire SNC aerial detection survey. The ability to fly beyond LOS and above 120 m (*e.g.* 500 m to produce the 10 cm GSD orthomosaics used in this study) would go a long way to improving efficiency. It is likely that a ScanEagle UAS (Moreland et al., 2015) or something of similar capability would be necessary to efficiently employ our methods over a large spatial extent. The other efficiency challenge that remains is producing a singular model for the entire survey area rather than developing specific models for each sub-area. This is only possible with strong control of the light physics (*e.g.* cloud cover, sun angle, calibrated surface reflectance, bi-directional reflectance function (Roujean et al., 1992) and the creation of a large (5000+ individuals) training network across the known infection area. Areas with significant mixed conifer components would also require species-level classification. Solutions to the aforementioned efficiency limitations and additional scientific inquiry into the outstanding knowledge gaps will be essential before these methods can be efficiently employed on a large scale.

2.6 Conclusion

Our results suugest that low-cost UAS remote sensing methods combined with logistic regression from generalized additive models can reliably detect (*i.e.*, better than random chance) SNC in individual coast Douglas-fir trees at the four sites we examined in western Oregon. Classifiers based on RF, GBM, GAM, and GLM algorithms can detect SNC with similar levels of accuracy, although classification reliability was highly variable depending on model specification and site. Classification models specified with R, G, and B covariates were as reliable as those based on NIR, NDVI, or TGI, indicating that complex data transformations and the added cost of a NIR camera was unwarranted. SNC infection maps can be produced at the individual-tree level and these results can be combined with a spatial cluster analysis to map focus areas to potentially increase the efficiency of field activities. While there are many limitations to implementation at a statewide scale, this novel study showed that SNC detection with a sUAS at the individual-tree scale is possible in smaller areas. This capability can provide decision making products related to SNC infection intensity at a level of detail that has not previously been available.

2.7 Figures

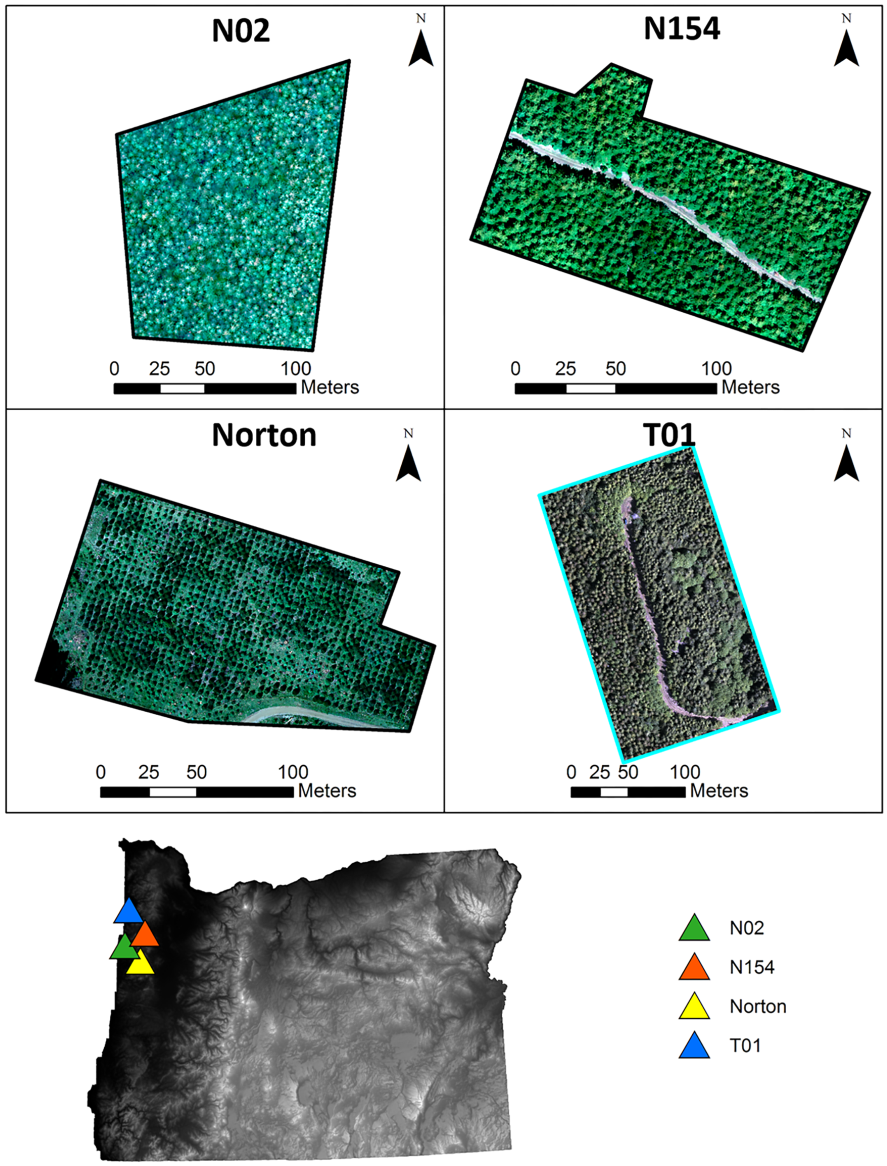


Figure 2.1: Study site map displaying the four sites selected for SNC disease detection. Names appear above the respective sites. Areas are approximately 1.6 ha and contain between 800 and 1300 individual Douglas-fir trees.

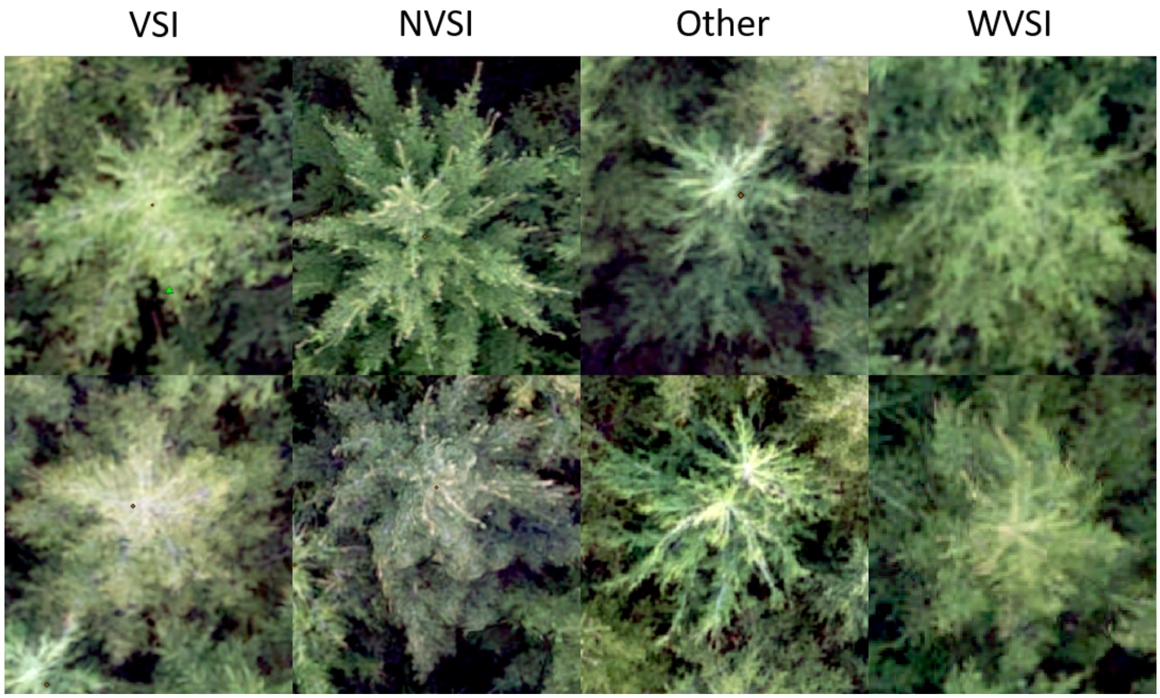


Figure 2.2: Classification training scheme depicting the criteria used to assign a class to any given training tree. Trees classified as having visible signs of infection (VSI) tended to exhibit signs of reduced leaf retention and yellowing of leaves. Trees having no visible signs of infection (NVSI) tended to have a full crown with little evidence of bare branches. Non-species, dead, and trees exhibiting unusual color and/or morphology were classified as 'Other' to keep the VSI and NVSI classes as distinct as possible. Trees classified as having weakly visible signs of infection (WVSI) tended to exhibit the coloration associated with SNC infection but no evidence of significantly reduced leaf area.

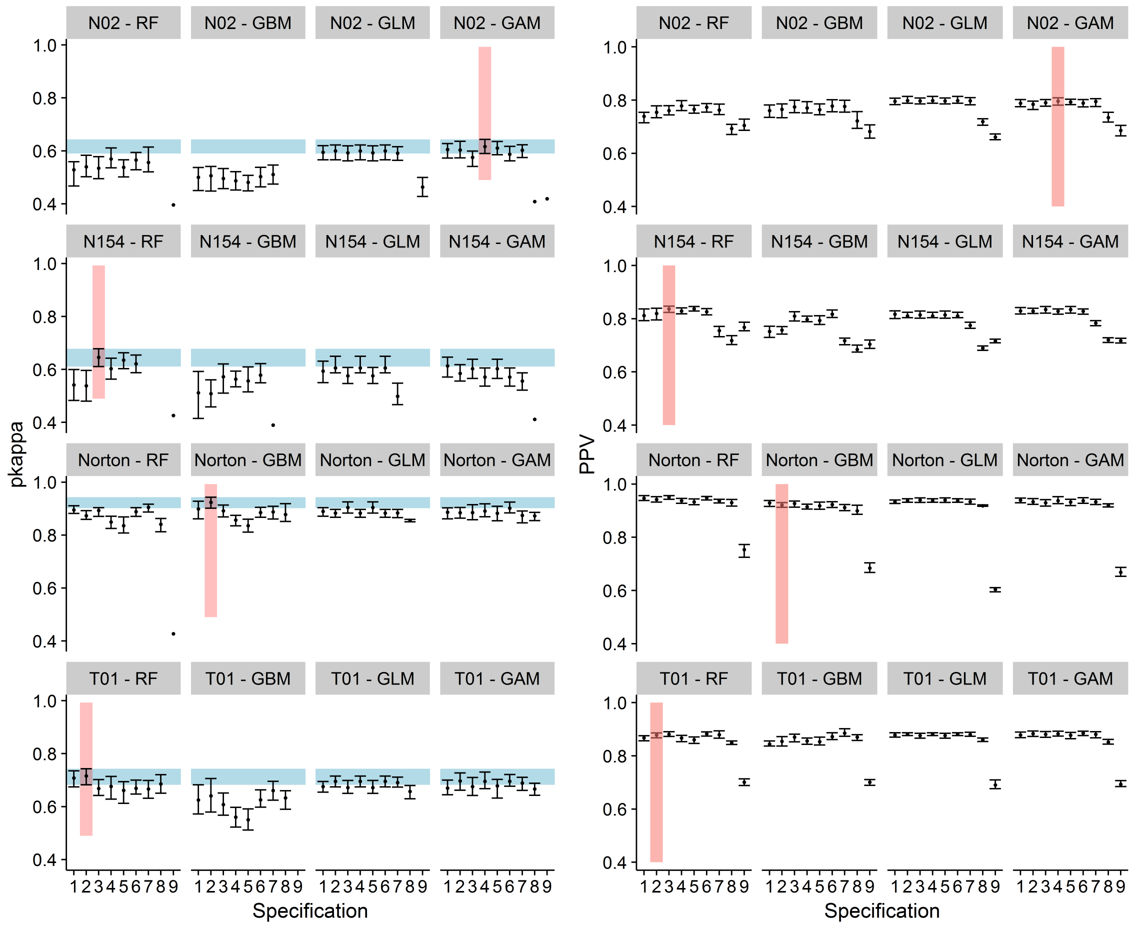


Figure 2.3: Estimated pkappa and PPV confidence intervals (CIs) for all combinations of site and classifier. The red bar indicates which model was best as determined by the highest lower pkappa CI and the blue bar is the CI for the best classifier at each site. The red bars on the PPV graph correspond to the best performing classifier based on the pkappa metric.

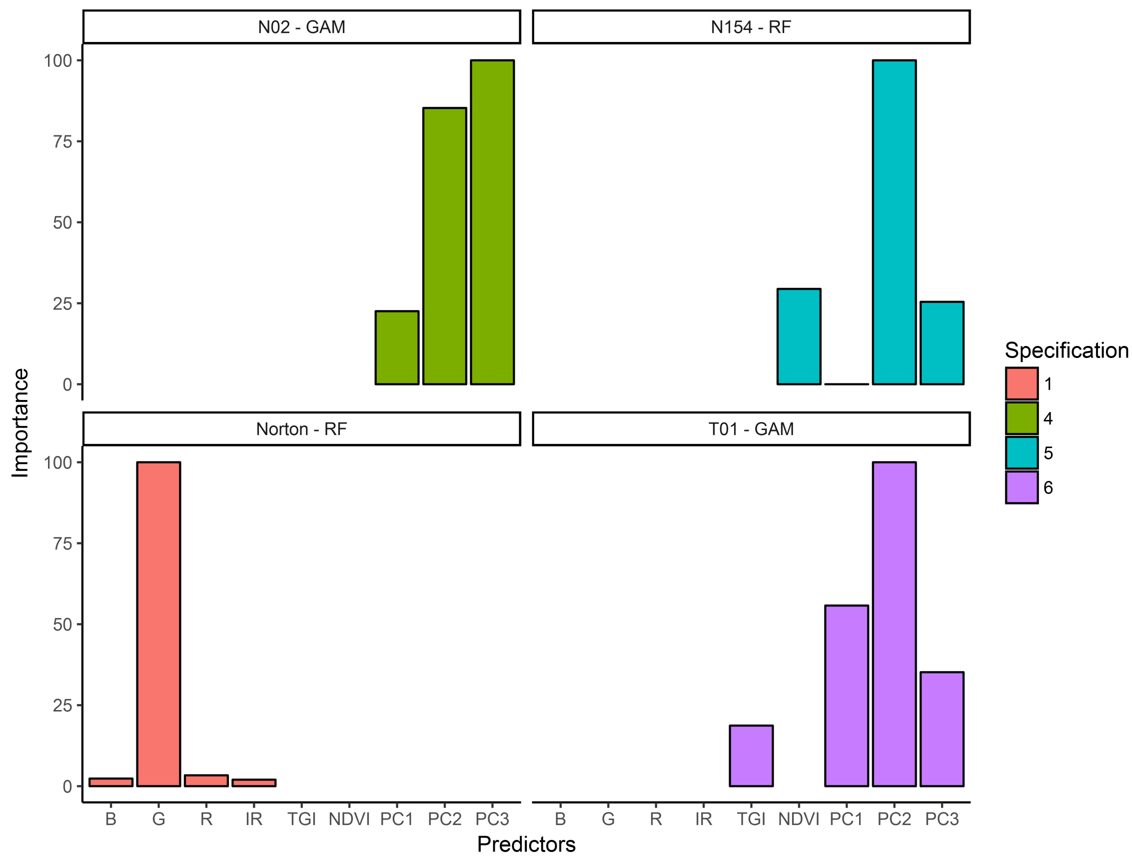


Figure 2.4: Variables of importance for best classifier identified for each site. None of the models contained explicit NDVI or near-infrared (NIR) predictors. Three of the four sites indicated strong influence from green reflectance (*e.g.* green and triangular greenness index). Flat lines indicate a predictor that is present in the specification but was unimportant for classification. Specification refers to the combination of explanatory variables used to fit the classification model. The different specifications are described in Table 2.2.

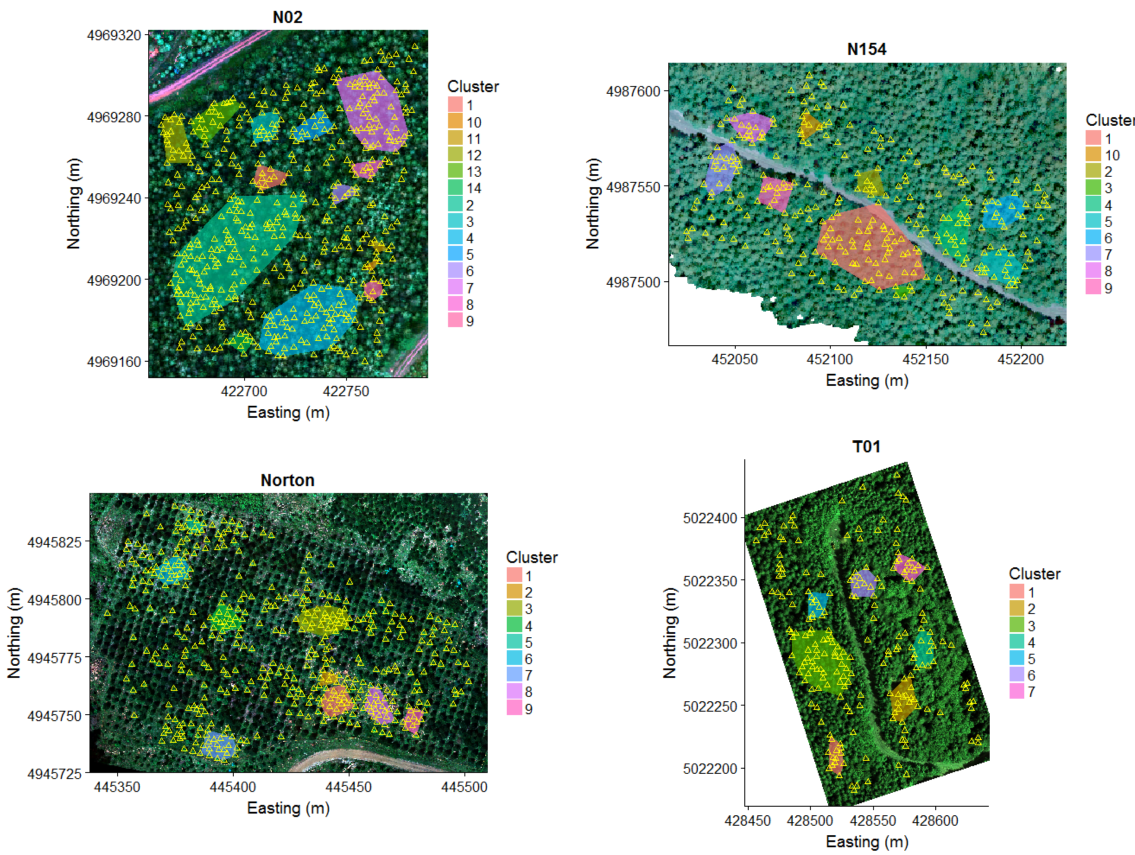


Figure 2.5: Maps of SNC detection results based on the GAM + S2 classifier for each site. These maps illustrate the ability to detect individual trees exhibiting visible signs of Swiss Needle Cast infection (*i.e.*, VSI) and the utilization of spatial clustering to focus field or management operations in areas of higher intensity. Clusters are areas where VSI trees tended to be clustered and represent potential focus areas for management operations

2.8 Tables

Table 2.1: Algorithms used for developing supervised classification models. The tuning parameters are used during model parameterization. The parameters containing a variable (v) are parameters that are optimized by iterative modeling functionality provided within the 'caret' package (Kuhn, 2016).

MLA	Tuning Parameters	Reference
RandomForest (RF)	# tree = 2000, mtry=v	Breiman (2001)
Generalized Linear Models (GLM)		
Generalized Additive Model (GAM	method=REML	Hastie and Tibshirani (1990)
Stochastic Gradient Boosted Machine (GBM)	# tree = v, interaction depth = v, shrinkage = v, min # obs = v	Friedman (2002)

Table 2.2: Nine different model specifications and the associated spectral explanatory variables.

Specification	Explanatory Variables
S1	G, R, B, IR
S2	G, R, B
S3	PC1, PC2, PC3, TGI, NDVI
S4	PC1, PC2, PC3
S5	PC1, PC2, PC3,NDVI
S6	PC1, PC2, PC3, TGI
S7	PCI, TGI, NDVI
S8	TGI
S9	NDVI

Table 2.3: Classification results for each site using specification 2 (S2) and the GAM classification algorithm. The 'Trees' column describes the number of Douglas-fir trees in the study site. Positive predictive value (PPV) is an assessment of the true positive rate and 1-PPV is an estimate of the false positive rate. VSI is the number of trees classified as having visible signs of infection for each site. VSI lower and VSI upper are VSI multiplied by the respective PPV lower/upper metric. For example, the best classifier for T01 detected 224 VSI trees, however, our bootstrapped PPV interval indicates that there are between 195 and 200 VSI trees. The VSI Rate (%) for each site is based on the VSI lower and indicates the severity of infection at each site.

Site	**Trees**	**Training Trees**	**PPV Lower**	**PPV Upper**	**VSI**	**VSI Lower**	**VSI Upper**	**VSI Rate (% min)**
N02	1265	300	0.76	0.80	559	428	445	33.8
N154	931	300	0.82	0.84	306	251	257	26.9
Norton	1245	300	0.92	0.95	462	427	437	34.2
T01	847	300	0.87	0.89	224	195	200	23.1

Acknowledgements

Funding for this project was provided through the USDA Forest Service, Region 6, Forest Health Protection, Special Technology Development Program.

3 Remote Sensing of Swiss Needle Cast Disease with a sUAS and a Narrowband Multispectral Camera

Jonathan D. Burnett
Dave Shaw, Ben Smith, Michael G. Wing

Submitted to Forest Science

3.1 Abstract

Swiss needle cast is a non-fatal foliar disease in Douglas-fir that reduces annual growth and stumpage value. Conventional detection methods rely on manned aerial detection surveys that are tedious, but modern revolutions in sUAS technology offer a potential alternative. Remotely sensing Swiss needle cast disease with a consumer grade camera onboard a sUAS has been shown to be reliable at four survey sites. The present study expands the scope to a total of eight sites imaged over two years and two seasons and includes a paired comparison between a multispectral camera and a consumer grade camera. Distinguishing between visibly diseased trees and those not visibly diseased was reliable at sites surveyed with the multispectral camera in May 2016 (kappa > 0.4) and paired detection surveys with a consumer grade camera produced similar results at all but two sites. Findings suggest that remotely sensed sUAS detection of Swiss needle cast is reliable under cloudy conditions, especially when using NDVI information derived from a narrowband multispectral camera. Additionally, summer surveys were less reliable and accurate than spring surveys and thus are not likely to be effective for Swiss needle cast detection. There was no evidence that foliage retention relates to classified diseased status of trees although, additional study is recommended due to the limited sample size ($n < 28$). Many regulatory, technical, and computational hurdles must be overcome before large scale implementation can be considered.

3.2 Introduction

Remotely sensing foliar disease with unmanned aircraft systems (UASs) has become increasingly prevalent (Garcia-Ruiz et al., 2013; Calderón et al., 2013; Dash et al., 2017; Patrick et al., 2017). UASs are uniquely adapted to this type of application because survey altitude and frequency can be customized to achieve the spatial and temporal resolution necessary to observe the desired phenomenon (Feng et al., 2015; Candiago et al., 2015). Small UAS technology is accessible to budget-strapped land managers and ecologists via the availability of low cost remote control (RC) aircraft parts, and affordable autopilot systems that can be assembled into a functional sUAS. (Koh and Wich, 2012). Inexpensive ($<$ \$1000) consumer-grade cameras can be integrated on UASs to keep costs low and the resulting imagery can be analyzed with conventional remote sensing workflows to detect a specific foliar phenomenon with acceptable reliability (Feng et al., 2015; Hunt et al., 2011; Martinelli et al., 2015). The sophistication of these low-cost systems can be further enhanced by deploying consumer-grade cameras modified to collect near-infrared (NIR) light to better isolate vegetative condition (Verhoeven, 2012; Rasmussen et al., 2016).

Our previous study (Burnett et al., 2017) applied the aforementioned innovations to an investigation of Swiss needle cast (SNC) disease detection using a low-cost sUAS equipped with a low-cost color and a low cost NIR camera. SNC (*Phaeocryptopus gaeumanni)* is a foliar disease endemic to Oregon's coast Douglas-fir (Hansen *et al.* 2000). SNC causes premature leaf senescence that can reduce annual growth increment, leading to end-of-rotation volume losses as high as 50% (Manter et al., 2000; Maguire et al., 2002). The phenology of SNC is such that aerial and field detection occurs in the spring time when infected needles become chlorotic and eventually abscise (Kanaskie et al., 2007; Ritóková et al., 2016). SNC was detectable at the tree-level with acceptable accuracy at the four sites surveyed using a consumer-grade color camera (Burnett et al., 2017). These results were presented with the caveat that investigation should be expanded to more sites and multiple time periods to demonstrate consistency. We also recommended experimenting with a narrowband multispectral camera (NMC) before discounting the potential advantage of collecting NIR imagery.

Narrowband multispectral cameras have the potential to improve the sensitivity of foliar disease detection surveys over consumer-grade sensors because a properly tuned camera, such as the MicaSense RedEdge (Patrick et al., 2017) or the Tetracam Mini (Candiago et al., 2015), is sensitive to the spectral wavelengths associated with maximal chlorophyll absorption (red and blue) and reflectance (green and NIR) (Clark and Lister, 1975; Nelson, 1983). It is this characteristic that has made Landsat imagery so effective for regional vegetation change surveys (Hilker et al., 2009; Kennedy et al., 2010). Furthermore, the NIR band on NMC tends to be tuned to a narrow region of the NIR spectrum that is just outside the red edge (Horler et al., 1983) rather than the broad region collected by a converted NIR sensor (Verhoeven et al., 2009).

A major advantage of employing NMC is the ability to derive vegetation indices (VIs) that relate to physiological conditions of the plant (or tree) in question. The most ubiquitous VI is the normalized differenced vegetation index (NDVI; Rouse et al. 1974). NDVI has been shown to be strongly correlated to green leaf biomass (Tucker, 1979), photosynthetically active radiation absorption (Huemmrich and Goward, 1997), leaf area index (Gitelson, 2004), and chlorophyll content (Hunt et al., 2011). The enhanced vegetation index (EVI) is an enhancement on NDVI that is less sensitive to saturation in canopies with high leaf area indices (Gao et al., 2000). The triangular greenness index (TGI) was developed by Hunt et al. (2011) as an alternative to NDVI. It is similarly associated with chlorophyll content, but does not require NIR information nor does it saturate in canopies with high leaf area. Plant senescence reflective index (PSRI) was shown to be strongly sensitive to leaf senescence (Gamon et al., 1995), which is a key sign of SNC. Small UASs have employed NMCs and associated VIs to reliably detect *Verticillium* wilt in olives (Calderón et al., 2013), tomato spot wilt in peanuts (Patrick et al., 2017), and *Flavescence doree* in grapevines (Albetis et al., 2017).

There appears to be little prior work investigating time-series or broad area (i.e., multiple sites, thousands of plants, or trees, *etc.,*) sUAS foliar disease surveys plants of species. This is likely due to the challenge of calibrating aerial imagery and accounting for potential confounding effects caused by lens vignetting, bi-directional reflectance, atmospheric absorption, shading and topographic effects (Smith and Milton, 1999; Gao et al., 2009), some of which have been shown to severely limit the extrapolation of sUAS survey results in certain conditions (Rasmussen et al., 2016).

The presented study incorporates the findings of our previous SNC detection study (Burnett et al., 2017) and vastly expands the scope of that work in terms of space and time. We conducted 22 SNC disease detection surveys across eight sites using a custom-built sUAS equipped with both a consumer-grade camera and NMC. Presently no studies exist to evaluate the efficacy of using a sUAS equipped with a narrowband multispectral camera to detect foliar disease in Douglas-fir or any commercial timber species. The objectives of our study were to (1) determine if SNC detection surveys with a NMC were more accurate and reliable than those conducted with a consumer grade camera, (2) evaluate the reliability of between year change detection at the site-level and tree-level, (3) investigate the plausibility of conducting summer SNC detection surveys with sUAS, and (4) examine the relationship between field-based foliage retention estimates and disease status as determined by remote sensing.

3.3 Methods

3.3.1 Study Area

The study area was comprised of eight 1.6-ha study sites. General stand condition and species composition of each site are summarized in Table 3.1. Six of the study sites, S1W, S1E, S2W, S2E, S2S, and S3W, were selected from areas in the Swiss Needle Cast Cooperative plot network (Ritóková et al., 2014). S1W and S3W corresponded with N02 and N154 from our earlier investigation (Burnett et al., 2017). Sites were chosen based on known SNC presence, the ability to maintain visual line of site between pilot and UAS, and containment within previously approved flight areas as specified in Federal Aviation Administration (FAA) Certificate of Authorization (COA) number 2014-WSA-212-COA.

3.3.2 Equipment

The sUAS selected for this study was a Tarot 650 quadcopter with a motor-to-motor span of 650 mm and an all-up weight of 3 kg, including the battery, gimbal, and sensors.

Data imaged in 2015 used a Sony Nex camera described in Burnett et al. (2017). The 2016 surveys incorporated a Sony A5100 24.3 MP camera with 20 mm lens to collect color imagery in broadband red, green and blue wavelengths. We did not deploy the NIR converted Nex camera from our previous study because previous findings suggested that NIR did not significantly improve SNC detection and thus did not warrant the added operational complexity (Burnett et al., 2017). 2016 surveys also utilized a Micasense RedEdge NMC. The RedEdge weighs 180 grams and has five spectral bands, blue, (B) green (G), red (R), red edge (RE), and near-infrared (NIR), centered over 475, 560, 668, 717 and 840 nm, respectively. B, G, R and NIR are approximately centered over Landsat 7 bands 1 – 4 (Chander et al., 2009), respectively. RedEdge integration was intuitive due to the included GPS for geotagging and an onboard intervalometer for automatic camera triggering. The A5100 and RedEdge sensors were mounted side-by-side to ensure both were exposed to the same lighting conditions and nadir-pointing was ensured by mounting the assembly to a two-axis gimbal.

The sUAS was equipped with a Pixhawk autopilot capable of conducting GPS navigated flight. Flights were planned on a ground control station (GCS) laptop running Mission Planner software (Oborne, 2015). Mission Planner estimated GPS navigation points along a grid to ensure a specified overlap and sidelap of imagery based on the following camera parameters: firing interval, focal length, sensor size in metric units, and number of pixels on the sensor. Overlap and sidelap were both set to 80% to ensure there were enough images to conduct a geometrically accurate reconstruction of the scene for each site (Agisoft, 2013). Flights were conducted at 120 m above ground level (AGL) resulting in image pixels with a ground sampling distance (GSD) of 2 cm and 8 cm for the A5100 and RedEdge cameras.

3.3.3 Field Operations

Four to six 0.2 m x 0.3 m iron cross ground control points (GCPs) were placed at each study site to facilitate georeferencing of the final orthomosaics. The number of GCPs was limited by forest openings, road locations, and the total area that could be covered while ensuring visual LOS to the sUAS. GCP positions were surveyed with a TOPCON GR3 RTK global positioning system. Positions were differentially corrected with OPUS corrections during postprocessing. Two 0.6 x 0.6 m drywall panels painted with titanium dioxide and having a reflectance of 90% were placed in the imaging area of each site for calibrating all imagery to a common standard. A 60% reflectance calibration panel was imaged with the RedEdge before and after each flight in accordance with the manufacturer's protocol. All flight operations were conducted under the guidelines specified in the COA and were also in accordance with all applicable Federal and State regulations.

The site-level details of surveys in terms of month, year and camera used are specified in Table 3.1. Flights took place within +/- 1 hour of solar noon to minimize the effect of sun shading (McGlone, 2013). Exceptions to this were 2015 S1E, S1W, S2W, S2S, S2E as well as 2016 S2E and S3N due to technical difficulties. Lighting conditions were limited to either fully sunny conditions or fully cloudy to ensure consistent solar radiance during the surveys (Table 3.1). The sites surveyed in August (Table 3.1) were flown with a 3DR Solo equipped with a custom gimbal and the RedEdge camera. Missions were planned with Mission Planner as described previously. No other sensor was flown in August. Trees climbed and evaluated for foliage retention were mapped with a Nikon Nivo total station during the August surveys.

3.3.4 Spectral Data Processing

Sony A5100 images from each survey were stitched into a single orthomosaic (OM) with Agisoft Photoscan. A digital surface model (DSM) was also created during this process. Photoscan processing followed the workflow described by Shervals and Dietrich (2016). Sony Nex images collected in 2015 were processed in an identical manner. DSMs were exported at 2.5 cm GSD and OMs were exported at 10 cm because 10 cm is the approximate GSD of the RedEdge OMs and was shown to be an effective resolution for SNC detection (Burnett et al., 2017). OMs were registered to GCPs during Photoscan processing.

Orthomosaics (OMs) were produced from the RedEdge images at each site using MicaSense's proprietary ATLAS web-based software. OMs for each site were registered to the reference image indicated in Table 3.1 using the image-to-image registration functionality in ArcGIS 10.3. RedEdge, A5100 and Nex OMs were subsequently calibrated with the empirical line method (Smith et al. 1999) using a custom Matlab (Mathworks, 2016) program. ELMs were fitted to the 90% reflective targets described previously and shadows between trees (assuming arbitrary 0.01% reflectance). Although RedEdge data was calibrated to surface reflectance during ATLAS processing, we used the ELM method to calibrate to the same spectral references as the Sony camera images to minimize systematic differences between OMs. OMs were decomposited into individual spectral grids (5 for RedEdge and 3 for Nex and A5100). Spectral grids listed in Table 3.2 were derived from the RedEdge spectral grids following calibration.

The remainder of processing occurred in the R programming environment (R Core Team, 2017). Digital surface models (DSMs) created from the reference surveys for each site (Table 3.1) were used to conduct individual tree segmentation using the method described in Burnett et al. (2017). Segmented trees provided a geospatially explicit crown area for each tree and a tree list for each site. Spectral data were associated with individual crown areas using the method previously described in Burnett et al. (2017). We randomly selected 300 trees from each site's tree list for classification model training. Three hundred trees were evaluated to ensure that classification models would not be overfitted when models contain many covariates. Each tree was classified as having either 'no visible sign of infection' (NVSI) or 'visible sign of infection' (VSI) using the criteria and visual interpretation method described in (Burnett et al., 2017). Trees were not classified simply diseased and non-diseased because Wilhelmi (2016) determined that trees could be infected with SNC without producing a visible response. We segmented one tree list and classified one set of 300 trees for each site under the assumption that tree inventory and disease status did not change from 2015 to 2016 or from May 2016 to August 2016.

3.3.5 Classification Model Development

The generalize additive model (GAM) implemented in R under the 'mgcv' package (Wood, 2017) was used within the caret environment (Kuhn and Johnson, 2013; Kuhn, 2016) to develop binomial classification models from each site's training datasets using the spectral covariate combinations specified in Table 3.3. Models were subsequently applied to the entire population of trees at each site to estimate the number of diseased (VSI) and non-diseased (NVSI) trees. The general form of the GAM model appears in equation 3.1. GAM was selected for this study because it was determined to be the most consistently reliable classification algorithm at all sites in our previous study (Burnett et al. 2017). For conciseness, any given specification 1 – 9 will be referred to as S1 – S9. Hereafter, the classification model resulting from the combination of the GAM classification algorithm and a given specification will be referred to as a classifier.

$$g\left(E\left(y_i\right)\right) = \beta_0 + f_1\left(x_1\right) + f_p\left(x_p\right) \tag{3.1}$$

Where $i = 1, ..., n$ trees, g is the log link function, y is the response 0 = NVSI and 1 = VSI, $\beta_0$is the intercept, $f_1, ..., f_p$ are the thin-plate spline smoothing functions relating each respective predictor (p) to the response variable, and x_1,...,x_p is mean spectral intensity of predictor (p) over the crown area of tree i.

We selected Cohen's Kappa coefficient (hereafter, kappa) and Positive Predictive Value (PPV) to assess and compare classification reliability and disease detection accuracy (Congalton, 1991; Parikh et al., 2008; Hajian-Tilaki, 2013) between classifiers in the comparisons described below. PPV and kappa best capture the intent of this study because kappa describes how well a classifier discriminates between VSI and NVSI and PPV is an estimate of how many trees classified as VSI are truly VSI.

3.3.6 Multispectral Camera Performance

We compared classifiers developed from several covariate specifications (Table 3.3) in terms of Kappa and PPV to discern an optimal classifier for RedEdge camera data using the imagery collected in May 2016. We defined an optimal classifier as one that achieved a kappa $>$ 0.4, PPV $>$ 0.7, and was not significantly different from the best performing classifier at each site. Once identified, we used the optimal classifier for the remainder of the RedEdge performance comparisons. The optimal classifier for the Sony Nex was identified in Burnett et al. (2017) and equivocal to the GAM + S1 classifier (Table 3.3). We assumed that the optimal classifier for the Sony A5100 camera was identical given the similarities between cameras.

Camera selection has implications on the cost and effectiveness of our method, and is especially pertinent in the context of operational implementation. To evaluate differences in both classification reliability and SNC detection accuracy, we conducted coincidental imaging with both the Sony A5100 camera and RedEdge NMC at six sites in 2016 (Table 3.1). Derived OMs were classified using the optimal classifier respective to each camera and results were compared using kappa and PPV.

Seasonal variation in SNC detectability was expected due to the phenology of Douglas-fir response to SNC. The premature senescence signature of SNC is evident in the spring, but fully masked by current-year leaf flush by late May or early June. However, the spring timing of the visible SNC signature coincides with western Oregon's rainy season. This can be disruptive to sUAS surveys (Burnett et al., 2017). Oregon's summers tend to be rain free with longer periods of high angle sun and, thus, present more optimal survey conditions than spring. We investigated SNC detectability during summer (*i.e.,* August) by comparing classification results between RedEdge NMC detection surveys conducted in both May and August 2016 at three different sites (Table 3.1).

3.3.7 Detection Consistency between Years

An effective SNC detection method should produce similar results between repeat surveys of the same sites when crown condition between surveys is relatively unchanged. Although some change is expected due to annual variation in weather patterns, only trees that have signatures near the boundaries between diseased and non-diseased classes would be expected to exhibit a change response between years because minor changes in foliage retention and foliage color may alter a given tree's position relative to the boundary. We tested consistency of our method by comparing classification results of detection surveys conducted with consumer-grade cameras in May 2015 and May 2016 (Table 3.1). Site-level consistency was evaluated with kappa and PPV scores. Individual tree consistency within a site was evaluated by comparing classified disease status of the same tree in both years. Due to the binomial classification method we used, there were four possible outcomes: *VSI to NVSI*, *NVSI to VSI*, *VSI to VSI* and *NVSI to NVSI*, with the former two outcomes being potentially undesirable results and the latter two providing evidence of stability (*i.e.,* no change). While camera models differed slightly between years, we assumed that sensor-induced differences were minimal because imagery was collected in linear RAW format (Verhoeven 2011) using identical lenses, and resulting OMs were calibrated to the same reflectance standard.

We expected that some sites would exhibit higher levels of change than can reasonably be attributed to SNC because none of the classifiers in our previous study were perfectly reliable (*i.e.,* kappa < 1). Moreover, different trees may be erroneously classified in repeat surveys due to variation in spectral response data between surveys. Probabilities associated with specific classes can be tuned in binomial classification models to improve the certainty of results (Harrell, 2015). We examined the effect of tuning classification probabilities by applying the following rule set to effectively create three different classes for each year: $>$ 75% VSI probability = VSI, $<$ 25% VSI probability = NVSI and 25% - 75% VSI probability = Unknown. Individual tree change between 2015 and 2016 surveys for the four sites (Table 3.1) was evaluated at the tree level. Each tree was assigned 1 of 9 possible classes, five of which are associated with a change from or to an uncertain status. The remaining four classes are, *VSI to NVSI*, *NVSI to VSI*, *VSI to VSI*, and *NVSI to NVSI*, where the first two classes are undesirable on a landscape where no change is expected and the latter two classes represent no detectable change.

3.3.8 Detected Status and Foliage Retention

Surveys from sites 1W, S2W, and S3W in 2015 covered experimental plots installed by the Swiss Needle Cast Cooperative (Ritóková et al., 2014).Ten trees on each plot were climbed in 2015 and manually evaluated for SNC disease status estimating needle retention in terms of years at the lower, middle, and upper crown of selected trees (Ritóková et al., 2014). We used the stem map constructed in August 2016 to form a spatial relationship between the VSI probability of any given tree as determined by the 2015 detection survey results and the estimated mean foliage retention of the mid-canopy of climbed trees. Results were plotted for visual trend assessment and a Pearson correlation coefficient (Mukaka, 2012) was calculated to quantify agreement between foliage retention and VSI probability.

3.4 Results

3.4.1 Multispectral Camera Comparisons

No singular model specification resulted in a clearly superior classifier at any of the sites (Figure 3.1). More than 95% of the training trees at S2E were classified as VSI. Classifiers based training sets with such high prevalence are not effective for discrimination so S2E was dropped from the study. Classifiers based on S4 and S7 were the only classifiers that were not significantly different from the best performing classifier for each site. S4 and S7 are differentiated only by the inclusion of NDVI (S4) and EVI (S7). S4 was selected as the optimal classifier for the remainder of this analysis due to the ubiquitous nature of NDVI in sUAS remote sensing. Classification and SNC detection reliability was above acceptable levels for all sites except S2W, however the CIs of kappa and PPV at these sites did include the minimum thresholds (kappa = 0.4, accuracy= 0.7).

The comparison of optimal classifiers between RedEdge and Sony A5100 cameras at the six sites imaged in 2016 (Figure 3.2) revealed evidence that datasets from both cameras provided similar discriminatory capability and detection accuracy. The two exceptions were comparisons at sites S3N and S3W where Red Edge discriminatory ability was substantially better than the A5100, although in terms of detection accuracy, Red Edge was only superior at S3N.

Comparison of classification performance between datasets collected with the RedEdge NMC in May and August of 2016 revealed evidence of substantially lower classification reliability (Figure 3.3a) and detection accuracy (Figure 3.3b) in summer. At S2S the classification reliability was scantly more reliable than a random guess with Kappa scores near zero. S1W was the sole exception with performance being similar (but still poorer in August) between the two months. In addition, reliability was similar between sites in May with kappa scores all near 0.5 but very dissimilar in August.

3.4.2 Detection Consistency between Years

The ability of classification models to reliably distinguish VSI from NVSI trees with imagery collected from consumer grade cameras at four different sites varied greatly between years and sites (Figure 3.4). 2016 reliability was better at S1E and S1W, and worse at S2W and S3W. A comparison of accuracy yielded similar trends, although differences tended to be lower in magnitude.

Adjusting probability thresholds associated with VSI/NVSI increased sensitivity of the classifiers to detect SNC at all sites in 2015 and 2016 (Figure 3.5a). The effect of changing class probability thresholds substantially reduced the number of trees detected as VSI (New VSI) at all sites in both years when compared to the number of trees detected as VSI with the default classification (Default) probability thresholds (Figure 3.5b). However, applying the adjusted classification thresholds notably reduced the number of undesirable changes detected between 2015 and 2016 at the four sites surveyed (Figure 3.6).

3.4.3 Detected Status and Foliage Retention

A qualitative comparison of foliage retention data from climbed trees to classification results from the A5100 imagery collected at S1W, S3W and S2S in 2015 revealed a relatively poor agreement (Figure 3.7a, b, c). Classification at S1W resulted in the most agreement between foliage retention estimates and disease status with six out of eight trees agreeing. Agreement at S3W (3.7b) and S2S (3.7c) was much worse with ratios of 4/8 and 2/8 trees, respectively. All trees at S2S had foliage retentions $<$ 3 years but only two of the trees were classified as VSI. Comparing foliage retention to the underlying classification probabilities for each tree (Figure 3.8) did not yield evidence of a quantifiable relationship ($R^2 < 0.01$).

3.5 Discussion

The individual tree-detection method rapidly estimated the disease status of nearly 6000 indvidiual Douglas-fir trees across eight sites. Our results corroborated those from Burnett et al. (2017) that accurate tree-level SNC detection is possible and that visibly affected trees (VSI) can be distinguished from those that appear unaffected (NVSI) with substantially better than random chance. The novelty of the present work is in the scope of inquiry afforded by conducting surveys at eight distinct sites although two sites were effectively removed from this analysis because they were so heavily diseased that training data representing diseased and non-diseased classes could not be produced.

3.5.1 Multispectral Camera Performance

Inclusion of NIR information in the form of NDVI consistently increased classification reliability and detection accuracy with the RedEdge NMC. Additionally, the inclusion of less commonly employed VIs, such as PSRI (S5), or TGI (S6) did not appear to improve reliability or accuracy over NDVI. The improved reliability and accuracy resulting from using NDVI contrasts with Burnett et al. (2017) who found that NIR in any form (to include NDVI) did not improve classification reliability with a consumer grade camera (Sony Nex). Direct comparisons between classifications of paired datasets collected simultaneously with the RedEdge and consumer grade camera (Sony A5100) suggest that a NMC can potentially improve reliability and to a lesser extent SNC detection accuracy. It is likely that the combination of utilizing the RedEdge NMC and inclusion of NDVI for classification increased the ability to discriminate between vegetative conditions because spectral bandwidths are narrower and centered on spectral wavelengths that are associated with plant physiological phenomenon (Tucker, 1979). Anecdotally, the RedEdge was easy to integrate into the sUAS owing to the integrated GPS and intuitive programming interface.

Summer surveys are likely to be less accurate or reliable as the conventional spring surveys. While the supporting dataset was small, the inconsistent performance of August detection surveys suggests that further investigation into feasibility of using the RedEdge NMC for summer detection of SNC is not warranted. Furthermore, August surveys experience an increased risk of being confounded by foliar responses to other damage agents, such as spruce budworm (Brubaker and Greene, 1979; MacLean and MacKinnon, 1996). The relatively small difference in both classification reliability and detection accuracy between May and August at S1W may indicate that the NDVI-based classifier was detecting distinctly different time invariant structural features (Candiago et al., 2015) between NVSI and VSI trees rather than the subtle ephemeral signature resulting from the combination of moderate needle retention and chlorotic needles. It is plausible that structural features like foliage retention (*i.e.,* leaf area index) may appear time invariant in stands where SNC infestation is so severe that numerous trees were averaging less than one year of needle retention.

3.5.2 Detection Consistency between Years

SNC detection with a consumer grade camera resulted in highly variable results between years and is suggestive of variable canopy conditions between years or the presence of confounding influences that shrouded observation of true canopy condition. It is unlikely that cameras produced the variations because they used the same lens, were flown at approximately the same altitude, were recording in raw format, and were calibrated to the same reflectance sources. Varied lighting conditions may be the culprit. Surveys at S1E and S1W produced similar results (in terms of reliability and accuracy) and both sites were imaged in cloudy conditions in 2015 and 2016 (Table 3.1). In contrast S2W and S3W surveys were imaged in cloudy conditions in 2015 and sunny conditions in 2016. Imaging in sunny conditions increases the confounding influence of uncorrected bi-directional reflectance (Rasmussen et al., 2016) and we observed many instances of crown shading. Imaging in sunny conditions also requires adjusting camera aperture and shutter speed to prevent sensor saturation. Furthermore, S2W and S3W are on a northwest facing aspect that further amplified topographic shading even at solar noon.

The number of trees changing status could be substantially reduced by tightening probability thresholds (Figure 3.5). The cost of employing this strategy to mitigate unnecessary change is the incidental creation of a third class containing trees of unknown classification. A cursory visual assessment of the S1W 2015 survey suggests a possible relationship between a tree receiving an ‘unknown’ status and visibly yellow crowns with little to no visible reductions in foliage retention. This type of tree would amount to a tree classified as weakly visible signs of infection (WVSI) in Burnett et al. (2017). It is likely that employing a multinomial classification strategy (Congalton, 1991; Goodchild, 1994) using the four classes defined in Burnett et al. (2017) would reduce the need to employ tightened probability thresholds. However, a multinomial classification model is much more complex and it may be that machine learning algorithms are better suited than GAM for handling the complex and possibly non-linear relationships (Lary et al., 2016).

3.5.3 Challenges to Large Scale Implementation

This study was heavily constrained by the same operational and regulatory limitations described in Burnett et al. (2017), most notably the requirement to fly under 120 m above ground level and to maintain visual line of sight on the aircraft severely impacted efficiency by reducing survey areas to < 10 acres in all but one location). To expand the presented methods across larger areas would require not only a bigger aircraft capable of flying for multiple hours but also a substantial revision of the methods. Future investigations related to broad-area implementation of our method should focus on (1) improving computational efficiency by using an object-oriented tree crown segmentation method (*e.g.* Strîmbu and Strîmbu (2015)), (2) establishing an efficient observation network for collecting training points, (3) improving understanding of the link between tree spectral response to SNC infection and tree physiological condition (4) identifying the detection window within the spring period that maximizes detection likelihood, (5) investigating optimal multinomial classifiers, and (6) examining the efficacy of employing our method over a large area now that FAA Part 107 (FAA, 2016) offers waivers to conduct beyond line of sight operations at altitudes greater than 120 m above ground level.

3.6 Conclusion

The Swiss needle cast detection method rapidly estimated the disease status of nearly 6000 indvidiual Douglas-fir trees across eight sites. SNC was detected accurately (PPV > 0.7) and trees having visible signs of infection were reliably distinguished from those that do not with sufficient reliability (kappa > 0.4) when using either the Sony A5100/Nex camera or the RedEdge NMC. However, the RedEdge consistently produced the most accurate SNC detection. Despite the RedEdge's increased sensitivity, summer SNC detection surveys do not appear feasible. Moreover, there was evidence that changing lighting conditions between surveys confounded accurate SNC detection. As such, we recommend imaging under clouds to minimize the effect (Rasmussen et al., 2016). We expect that the methods employed would be more broadly applicable to other foliar diseases where causal agents of premature leaf senesce can be easily separated by appropriate survey timing, such as oak wilt disease (Everitt et al., 1999) and even sudden oak death (Meentemeyer et al., 2004). Several technical, computational, and regulatory challenges must be overcome before large scale implementation of our method should be attempted.

3.7 Figures

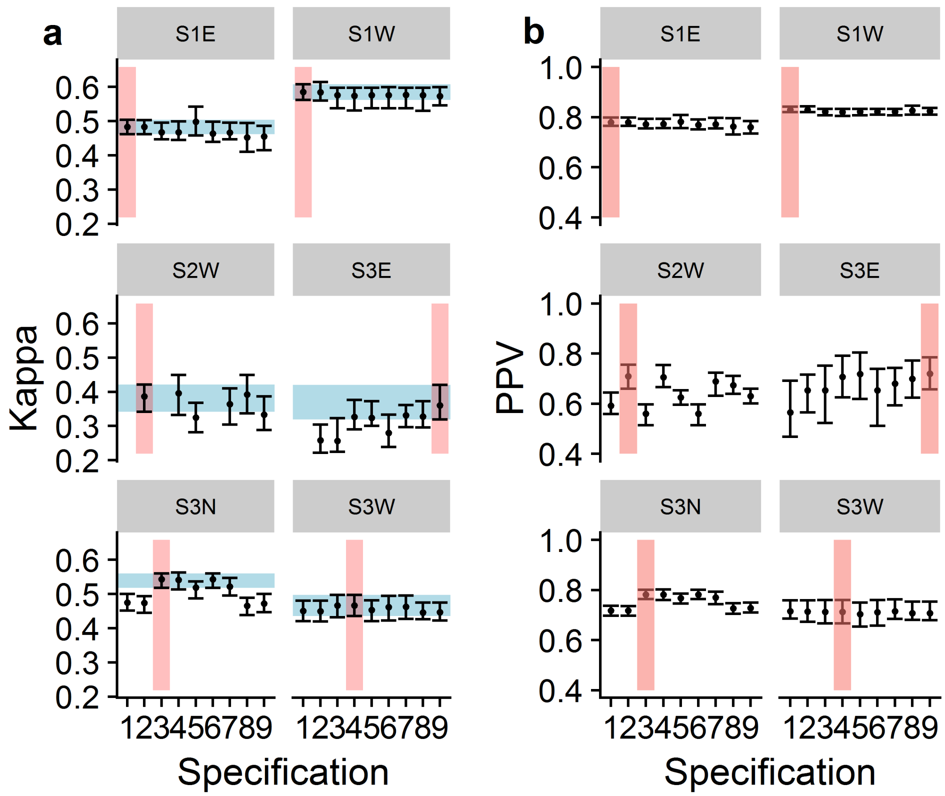


Figure 3.1: Comparison of the (a) Kappa coefficient and (b) Positive Predictive Value (PPV) for the nine different Generalized Additive Model (GAM) classification model specifications for each of the six sites imaged with the RedEdge multispectral camera in May of 2016.

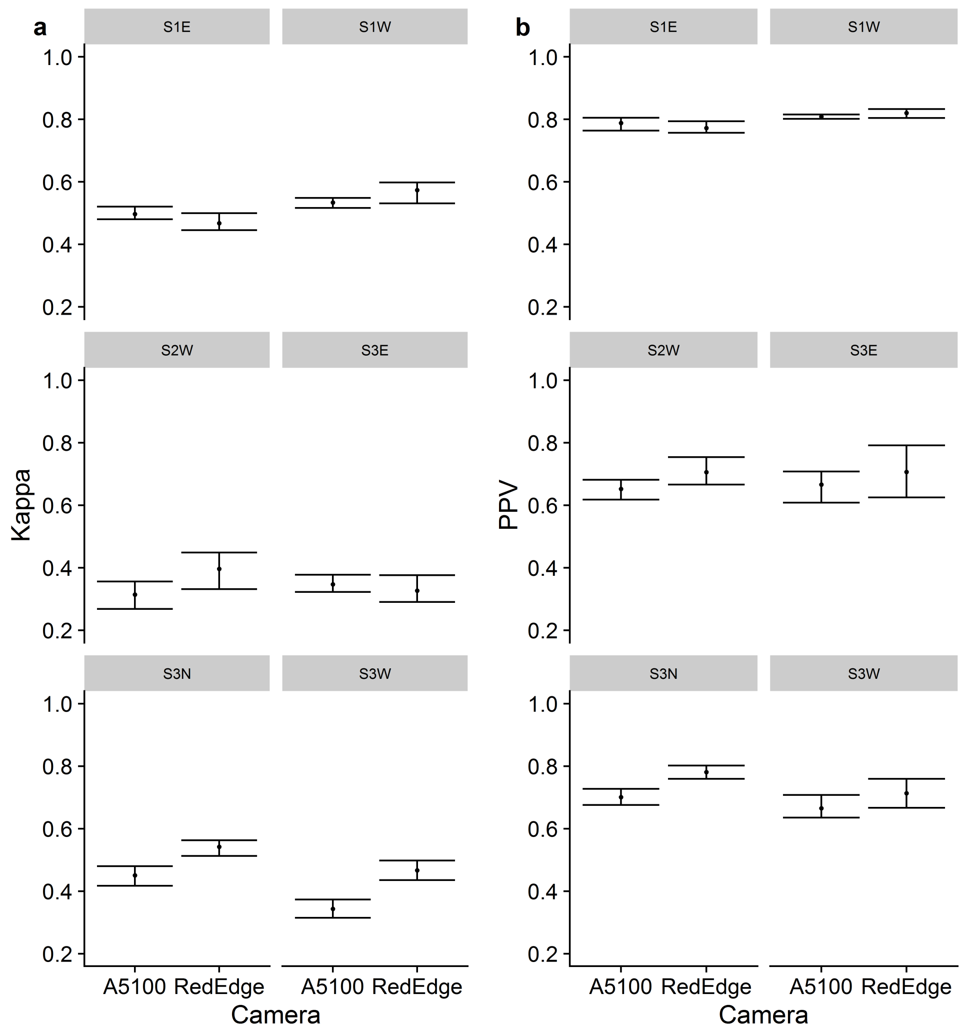


Figure 3.2: Between sensor comparison of the (a) Kappa coefficient and (b) Positive Predictive Value (PPV) for SNC disease classification of the six sites imaged with both the Sony A5100 camera and the RedEdge multispectral camera in 2016. Bars represent the 95% confidence intervals around the mean estimate.

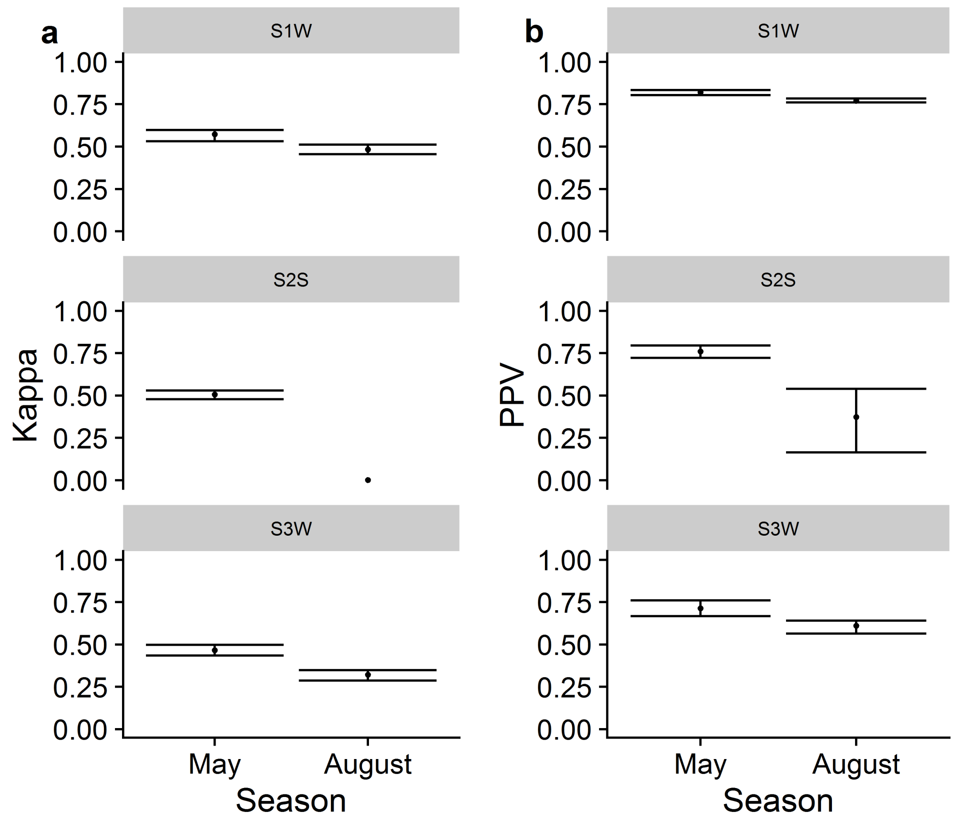


Figure 3.3: Comparison of the (a) Kappa coefficient and (b) Positive Predictive Value (PPV) for SNC disease classification of the three sites imaged with the RedEdge multispectral camera in both May and August of the year 2016. Bars represent the 95% confidence intervals around the mean estimate. Note that S2S May results are from the Sony Nex imagery collected in 2015 because S2S was not imaged with the RedEdge in May 2016 due to radio range limitations.

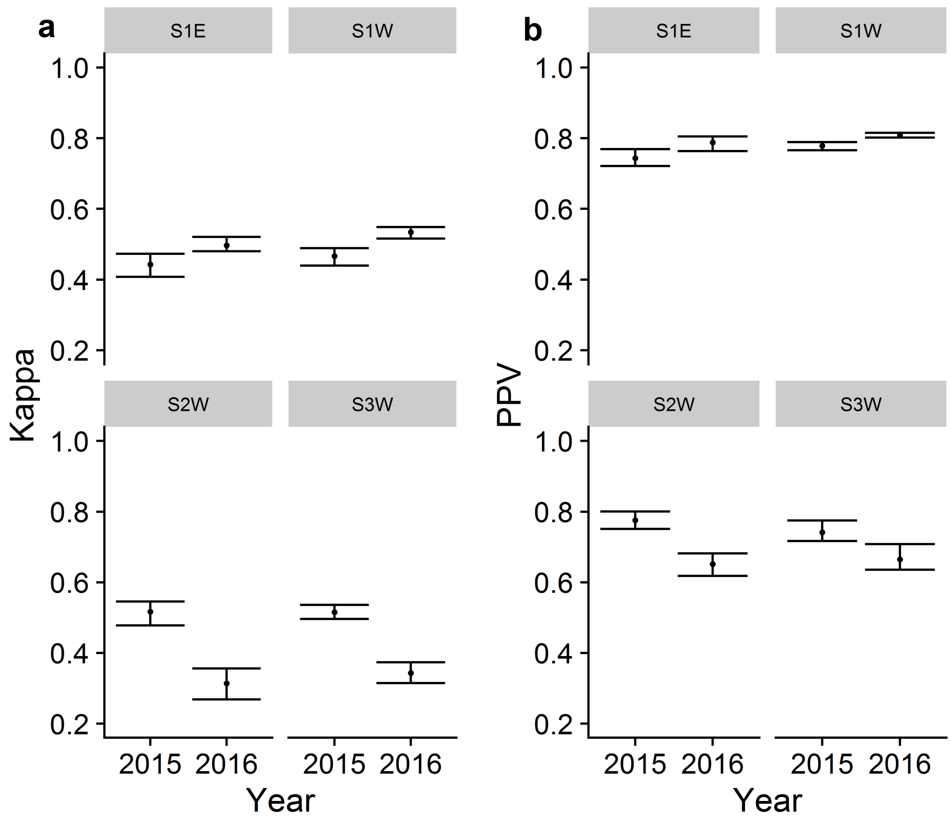


Figure 3.4: Between year comparison of the Kappa coefficient and Positive Predictive Value (PPV) for SNC disease classification of the four sites imaged in both 2015 and 2016. 2015 data are results from analysis on imagery collected with the Sony Nex camera and the 2016 data are results from the analysis on imagery collected with the Sony A5100 camera. Bars represent the 95% confidence intervals around the mean estimate.

Figure 3.5: Between year comparison of both (a) sensitivity and (b) # of VSI trees per site in the context of the default classification scheme and the adjusted threshold method described in the methods section. Plot headings are the site names corresponding to the four sites evaluated with the between year comparison in Figure 5. Default refers to the sensitivity of the classifier for a given site and sensor using the default probability threshold of >50% = VSI and < 50% = no visible signs of infection (NVSI). The 25% threshold refers to the specificity achieved by adjusting the visible signs of infection (VSI) classification threshold to those trees having a > 75% likelihood of being VSI and the NVSI classification threshold to those trees having a <= 25% of being VSI. Old VSI is the number of trees per site that were estimated to be VSI using the default probability threshold and the new VSI is the number of trees per site estimated to be VSI using the 25% probability threshold.

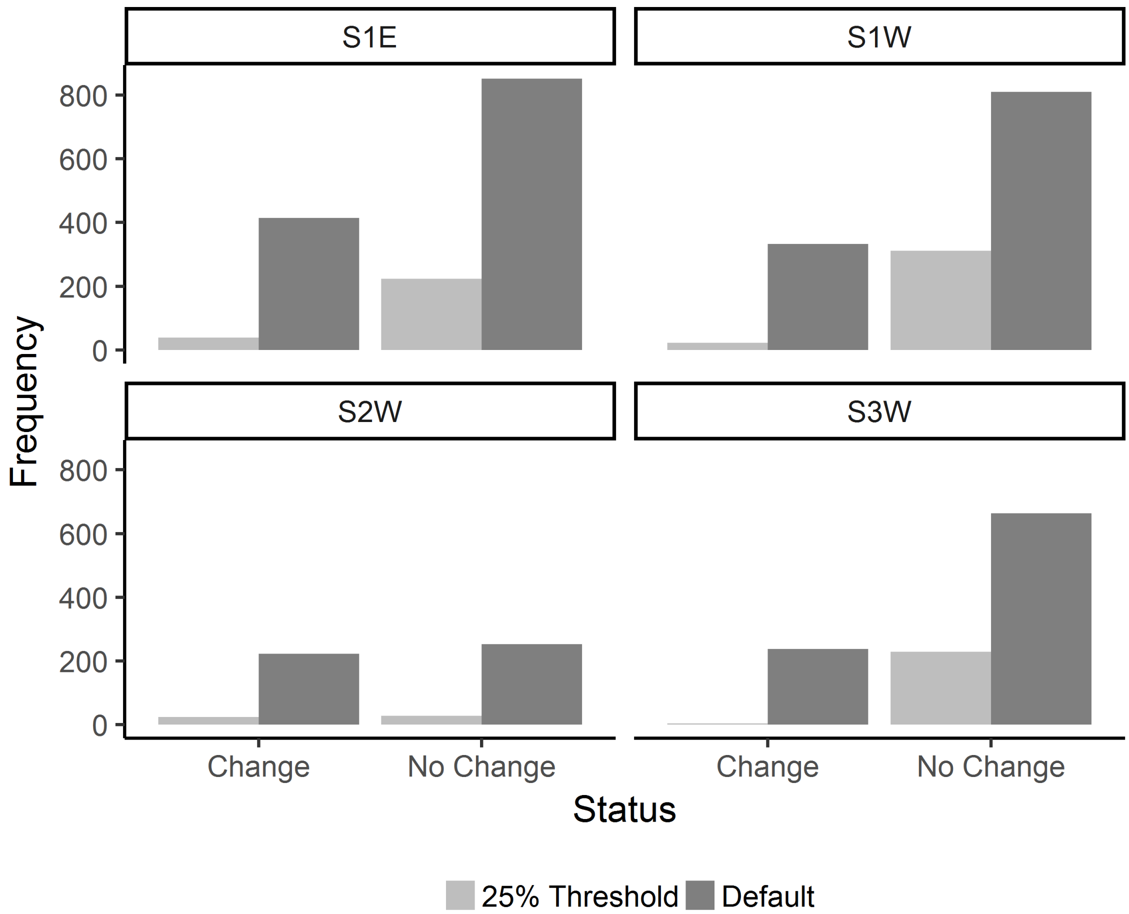
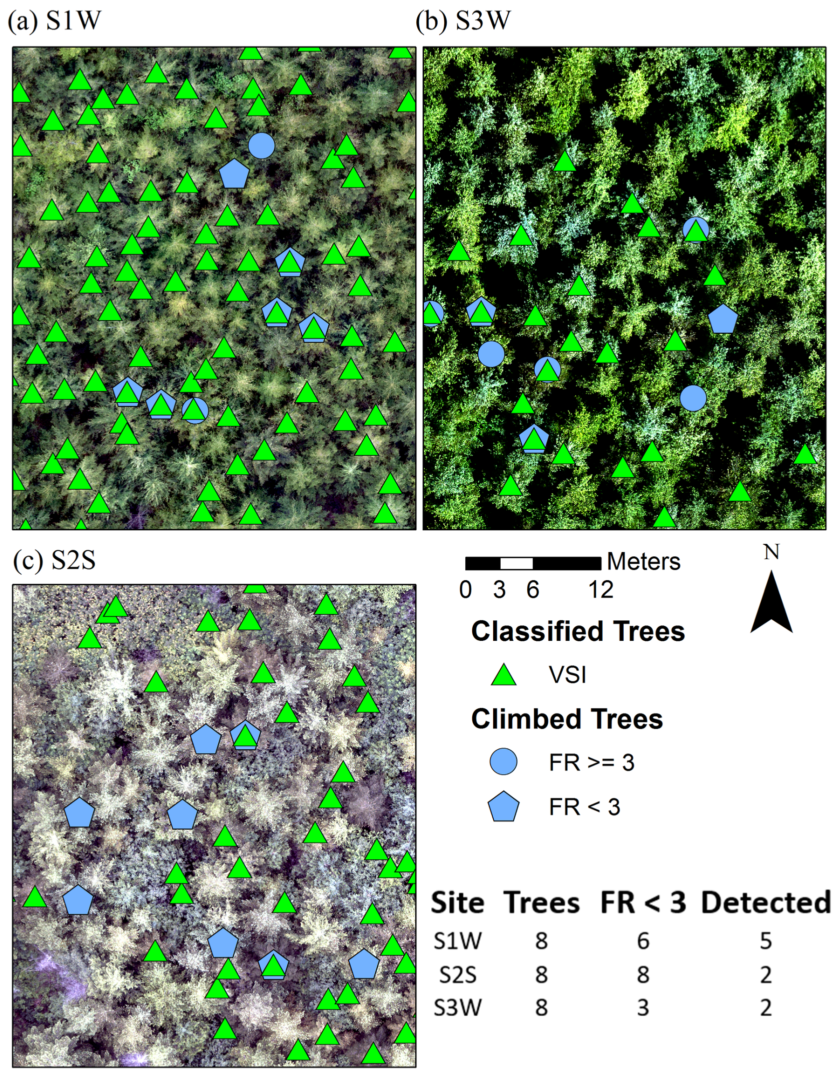


Figure 3.6: Classification change between 2015 and 2016 at each of the four sites imaged in both years. 'Change' refers to the number of trees that changed classification between years and 'No Change' refers to the number of trees that did not change classification between years. These results depict a method for reducing the frequency of potentially undesirable 'Change' occurrences by only selecting classification results of high certainty at the cost of adding a third class (not shown) of trees where change status is unknown because they fall outside the 25% thresholds for visible signs of infection (VSI) and no visible signs of infection (NVSI). Site name is annotated at the head of each plot. A graphical depiction of these results for site S2W appears in Appendix Figure A.4. '25% Threshold' and 'Default' are described in the caption for Figure 6.



Site	Trees	FR < 3	Detected
S1W	8	6	5
S2S	8	8	2
S3W	8	3	2

Figure 3.7: Map of trees classified as having visible signs of infection (VSI) compared to the estimated mid-crown average foliage retention (FR) of climbed trees (blue icons). The included table is a summarization of the comparison. FR $<$ 3 refers to trees estimated as having fewer than three years of foliage retention. The 'Detection' heading is the number of climbed trees at each site that were classified as VSI.

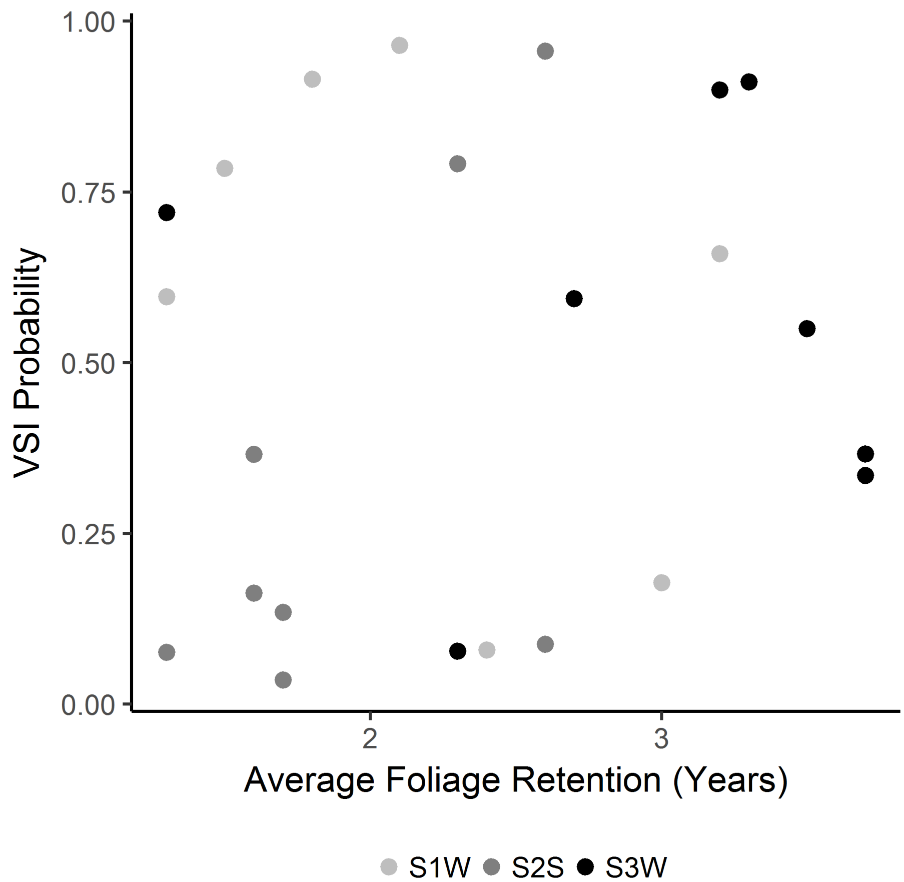


Figure 3.8: The relationship between the probability that a climbed tree exhibits visible signs of infection (VSI) and average mid-crown foliage retention of each climbed tree depicted in Figure 3.7. Point colors denote the site associated with a specific observation.

3.8 Tables

Table 3.1: General stand age by site and site-specific accounting of survey month and year as well as camera employed. Stands were primarily mature (M) although one stand was primarily comprised of saplings (S). Species indicates the dominant species on site. Most sites were dominant in Douglas-fir (DF) but a few sites contained large pockets of mixed alder (MA). The single mixed conifer (MC) site contained hemlock. Cloudy (Cldy) and sunny (Sun) indicate the lighting conditions at the time of the survey. Asterisks (*) indicate the dataset at each site that was used as the reference for image-to-image registration as well as the digital surface model used for individual tree segmentation. The year of the survey, month of the survey and camera used during the survey is denoted in the table cells directly under the cell labeled 'Year - Month - Camera'.

Sites:	**S1W**	**S1E**	**S2W**	**S2S**	**S2E**	**S3W**	**S3E**	**S3N**
Stand Age:	M	M	M	M	M	M	S	M
Species:	DF	DF	DF	MA	MA	MC	DF	DF
Year - Month - Camera	-	-	-	-	-	-	-	-
2015 - May - Nex	Cldy*	Cldy*	Cldy*	Cldy*	-	Sun	-	-
2016 - May - A5100	Cldy	Cldy	Sun	-	Sun*	Sun*	Sun*	Sun*
2016 - May - RedEdge	Cldy	Cldy	Sun	-	Sun	Sun	Sun	Sun
2016 - August - RedEdge	Sun	-	-	Sun	-	Sun	-	-

Table 3.2: Spectral grids used for covariates in classification models. Asterisk (*) indicates the spectral grids available from orthomosaics derived from Sony Nex and Sony A5100 imagery.

Spectral Grid	**Derivation**	**Reference**
Red* (R)	Direct Observation	-
Green* (G)	Direct Observation	-
Blue* (B)	Direct Observation	-
Near-infrared (NIR)	Direct Observation	-
1st Principal Component of RGB (PC1)	Principal component analysis	Venables and Ripley (2013)
2nd Principal Component of RGB (PC2)	Principal component analysis	Venables and Ripley (2013)
3rd Principal Component of RGB (PC3)	Principal component analysis	Venables and Ripley (2013)
Normalized Difference Vegetation Index (NDVI)	(NIR - R) / (NIR + R)	Rouse et al. (1974)
Triangular Greenness Index (TGI)	- 0.5*((668-475)*(R - G)-(668-560)*(R - B))	Hunt et al. (2011)
Enhanced Vegetation Index (EVI)	2.5 * (NIR - R) / (NIR + 6 * R - 7.5 * B + 1)	Huete et al. (2002)
Plant Senescence Reflection Index (PSRI)	R - G / NIR	Gamon et al. (1995)

Table 3.3: Specification combinations evaluated with the GAM classification algorithm. Covariate sources are described in Table 3.2.

Specification	**Covariates**
1	B, G, R
2	B, G, R, NIR
3	PC1, PC2, PC3
4	PC1, PC2, PC3, NDVI
5	PC1, PC2, PC3, PSRI
6	PC1, PC2, PC3, TGI
7	PC1, PC2, PC3, EVI
8	PC1, PSRI, NDVI, TGI
9	PC1, PSRI, EVI, TGI

Acknowledgements

Funding for this project was provided through the USDA Forest Service, Region 6, Forest Health Protection, Special Technology Development Program.

Acknowledgements

Funding for this project was provided through the USDA Forest Service, Region 6, Forest Health Protection, Special Technology Development Program.

4 Estimating Morphometric Attributes of Baleen Whales with Photogrammetry from Small UAS: A Case Study with Blue and Gray Whales

Jonathan D. Burnett
Leila Lemos, Dawn Barlow, Michael G. Wing, Todd Chandler, Leigh Torres

Submitted to Marine Mammal Science

4.1 Abstract

Small Unmanned Aircraft Systems (sUASs) are fostering novel approaches to marine mammal research, including baleen whale photogrammetry, by providing new observational perspectives. We collected vertical images of 89 gray and 6 blue whales using a commercially available sUAS to examine the precision and accuracy of image-based morphometry. Moreover, measurements from 192 images of a 1 m calibration object were used to examine variants of a scaling correction model. Results indicate that a linear mixed model including an error term for flight and date contained 0.17 m less error and 0.25 m less bias than no correction. We used the propagation uncertainty law to examine error contributions from scaling and image measurement (digitization) to determine that digitization accounted for 97% of total variance. Additionally, we present a new body size metric termed Body Area Index (BAI). BAI is scale-invariant and is independent of body length (R^2 = 0.11), enabling robust comparisons of body size within and among populations, and over time. With this study we present a three-program analysis suite that measures baleen whales and applies scale corrections to produce 11 morphometric attributes from sUAS imagery. The program is freely available to the community and is expected to improve processing efficiency and analytical continuity.

4.2 Introduction

Across the animal kingdom, ecologists develop and analyze various metrics to gauge response to environmental and anthropogenic change. A key measure of physiological health and response is animal body condition assessed through morphometrics. Morphometrics are the numerical expression of animal morphological characteristics that facilitate evidenced-based quantitative analysis of individual and population trends (Stower et al., 1960). Baleen whale morphometrics can be examined at an individual level to describe energy stores and reproductive capacity, and also at a population level to describe pervasive influences on population health and viability. Typically, larger individuals are considered to be in a better health state due to increased capacity for energy storage (Clutton-Brock and Sheldon, 2010; Christiansen et al., 2016). Baleen whales (*Mysticetes*) are long-lived, capital breeders that rely on energy stores to support reproductive and migratory life-history stages. Therefore, morphometric comparison of baleen whale body condition, across individuals and over time can reveal reproductive state, offspring growth rates, energetic capacity, body size demographic structure, and incidents of compromised health due to injury (Lockyer, 1986; Perryman and Lynn, 2002; Lockyer, 2007; Christiansen et al., 2016). Synoptic data on prey availability, ecosystem state, and acute impacts, such as entanglements or vessel strike, coupled with body condition data can reveal important data on response levels and recovery rates.

Measurements that facilitate morphometric analysis can be generally categorized as direct or indirect. Direct measurement of non-captive subjects typically involves destructive sampling such as whaling, or samples of opportunity such as stranding events (Norris, 1961; Forrester et al., 1980; Finley and Darling, 1990). These direct measurements tend to be the most accurate method as morphological parameters can be directly recorded using a measuring device. As such, the major source of potential error is attributed to the individual measurer. However, there are significant caveats to this approach including the ethics and feasibility of acquiring dead animals, along with small sample sizes and limited possibilities of individual re-measurement to discern changes in body condition. Conversely, photogrammetric morphometric acquisition methods use geometric principles to estimate parameters based on a scaling reference. This method provides a means of noninvasive data collection, the potential for collecting larger sample sizes, and the ability to re-measure individuals. However, indirect methods are estimated rather than directly measured and thus are subject to more sources of error such as ranging error, individual measurement error, and pixel mixing.

Photogrammetry dates to 1829 and has been defined as "... the science or art of obtaining reliable measurements by means of photographs" (Konecny, 1985; McGlone, 2013). Reliable or precise photogrammetric measurements are possible when certain physical parameters relating to the camera(s) are known. While a limited number of photogrammetric techniques can be applied to a single image (*e.g.*, single photo resection), stereo photogrammetry and 3D reconstruction relies on multiple, overlapping images of a scene acquired using either a moving camera (*e.g.*, on an aircraft) or multiple cameras offset by a known baseline distance. Camera viewing geometries can range from highly oblique to nadir (*i.e.*, direct overhead viewing). These viewing geometries can be achieved via camera placement on a boat deck, a crow's nest, or on an aircraft. The multi-camera method has been used to estimate lengths of hammerhead sharks (Klimley and Brown, 1983), bowhead whales (Cubbage and Calambokidis, 1987), and sperm whales (Dawson et al., 1995). Single image methods pioneered by Whitehead and Payne (1981) are more common and have been used to estimate morphometry of southern right whales (Best and Rüther, 1992), dolphins (Perryman and Lynn, 1993), gray whales (Perryman and Lynn, 2002), sperm whales (Jaquet, 2006), killer whales (Fearnbach et al., 2011), and whale sharks (Rohner et al., 2011).

Several sources of uncertainty can influence the precision and accuracy of morphometric measurements from photogrammetry. These sources have been previously documented (Perryman and Lynn, 2002; Jaquet, 2006; Fearnbach et al., 2011; Christiansen et al., 2016), and include: body flex, non-horizontal body position, light refraction on submerged body, deviations in camera roll, pitch (*i.e.*, pointing angle) and yaw as well as errors in reported range (*e.g.*, distance from camera to whale). Jaquet (2006) used a wooden plank of known length to create a regression model for the purpose of scaling altitude dependent pixel length measurements to real world units while simultaneously calibrating out ranging error and inherent error sources in the camera and lens system. These errors were determined to be negligible based on the low coefficient of variation (CV) of repeated measurements.

Jaquet (2006) also examined effects off angular error and determined that a camera position that deviated by 10 degrees off perpendicular resulted in $< 2.5\%$ underestimation of fluke width. Perryman and Lynn (2002) and Fearnbach et al. (2011) minimized the influence of uncertainty from body flex, body attitude, body submersion and camera tilt by filtering out images where these sources were evident to a substantial degree.

Cetacean photogrammetry was traditionally conducted via time and resource consuming manned aerial surveys, which can be cost prohibitive thus limiting repeated flights. However, recent technological advances have resulted in the affordable miniaturization of aircraft and camera systems culminating in the advent of small unmanned aircraft systems (sUAS) technology (Wing et al., 2013) in the early 2000s. The advent and accessibility of sUAS technology makes photogrammetric measurement of cetaceans more accessible, safe, cost-effective and repeatable.

Durban et al. (2015) demonstrated a single camera vertical photogrammetry method for measuring killer whales using a hexacopter equipped with a 25 mm focal length consumer grade camera. Christiansen et al. (2016) applied a similar method to assess the body composition of hundreds of humpback whales with a low-cost Splashdrone. They concluded that resulting measurements appeared to be robust to error within and between images, however, accurate image scaling required the scaling object (a ship) to be in close proximity to the whale. These two studies represent the beginning of a new era for baleen whale research driven by the increased accessibility offered by low cost sUASs.

The multitude of methods for assessing whale morphology and the coming tide of sUAS data that will result from ubiquitous application of low-cost minimal risk systems such as the Splashdrones and DJI Phantom sUASs, create a clear need to develop a standardized and repeatable method of conducting photogrammetric surveys and subsequent morphometric analyses. Furthermore, clearer field methodology needs to be presented to ensure that the broader community with access to low-cost aerial survey equipment ensures the safety of their personnel and the subject of interest, but also that data is collected in a manner that ensures reliable inferences can be achieved.

The goals of the presented study were to (1) establish methods for conducting accurate and repeatable sUAS photogrammetric surveys that do not require scaling objects to be co-located with the survey subject, (2) thoroughly evaluate sources for measurement uncertainty, (3) examine strategies to reduce measurement uncertainty, (4) develop standardized methods for extracting whale morphometrics from vertical sUAS imagery,and (5) disseminate these methods in the form of freely-available MATLAB and R scripts. We developed these methods with vertical sUAS imagery of Eastern North Pacific gray whales (*Eschrichtius robustus*) foraging off the coast of Oregon, USA and pygmy blue whales (*Balaenoptera musculus brevicauda*) foraging in the South Taranaki Bight of New Zealand. We collected the commonly accepted morphometrics for evaluating whale body condition (*e.g.*, length and width: Perryman and Lynn 2002) and incorporated additional width measurements at percentages of total length similar to the method presented by Christiansen et al. (2016). In addition, we introduce a length normalized surface area index that we refer to as body area index (BAI) that allows comparison of body size among whales similar to body mass index (BMI) in humans.

4.3 Methods

4.3.1 Study Area and Collection Methods

Small UAS overflights of blue whales occurred in the South Taranaki Bight region of New Zealand during the January-February period of 2016 as part of a larger project to describe the ecology of this population (Torres, 2013). Field methods were thoroughly documented in Torres et al. (2017). Six blue whales were imaged over four separate flights during this period. Small UAS overflights of gray whales occurred off the Oregon Coast, USA during the August – October period of 2016; 89 gray whales were imaged over 43 flights. The primary survey equipment for this study was a DJI Phantom 3 Pro sUAS. The camera has a 3.61 mm focal length and a pixel pitch of 0.0015 mm. Manual flight control of the aircraft was through the included remote control. Small UAS configuration and real-time camera output were available through an Apple IPAD Mini tablet ground station that was operating the DJI Go application.

The Phantom 3 Pro sUAS was chosen because the system is robust to cross winds even when traveling at 40 kph, pilot training is intuitive, and the aircraft can safely initialize on a moving platform (*e.g.*, boat). The Phantom 3 camera has low distortion compared to other similarly sized cameras (*e.g.*, GoPro), is stabilized by a 3-axis brushless gimbal, is capable of 4K video output, and can transmit a high definition real-time video sample to the pilot/observer. The video contains altitude and geolocation metadata recorded at 1 hz and camera directional pointing is controlled via remote control.

The sUAS was navigated such that the surfacing whales were centered in the camera field of view at altitudes between 25 m and 40 m above sea level (ASL) with a flight duration of $<$ 10 min. A calibration object of known length was centered in the frame and imaged from 10 m to 40 m during takeoff and landing for all flights similar to the method described in Durban et al. (2015), although they used a boat. Object lengths were 4.40 m and 1.00 m for the blue and gray whale flights, respectively. The product of the aerial survey is a 4k video of individual whales and calibration objects. Video format was chosen instead of still images because individual frames were high resolution (*e.g.*, 8 mp) and video increased the likelihood of capturing a whale in an ideal presentation. In addition, video facilitates a behavioral analysis to support a future study.

4.3.2 Body Area Index

Christiansen et al. (2016) demonstrated that intraseasonal body condition changes in humpback whales can be assessed with a body condition index (BCI). BCI captures width variation along the length of the whale by segmenting trapezoids along percentiles of body length and using sums of trapezoids to estimate flat dorsal surface area. Although BCI offers an approximation of surface area, we developed a more complete estimate of the flat dorsal surface area of a whale than trapezoid sums, by assuming a parabolic shape for approximately 40% of a whale's total length and combining the area under parabolas representing each side of a whale (Fig 4.1). We chose a parabolic model because it appeared to fit the 'average' whale well and the model provided a least-squares optimized fit that added objectivity to the subjectivity of clicking points to delineate the edge of the whale body. Furthermore, parabolic models have been successfully used to evaluate body condition in large-bodied mammals such as cows (Halachmi et al., 2008, 2013).

We evaluated the parabolic shape of each whale by orienting each image along a horizontal axis in pixel space from rostrum to tail where length was along the x-axis and width was along the y-axis. Eleven points were placed on the outline of each whale side between 15% and 65% of length at approximately 5% intervals. Parabolas were independently fit to each side, due to a lack of symmetry caused by mild variation in body presentation (*e.g.*, curvature). Parabolas were fit to the points using Eq. 4.1.

$$BW_i = \beta_0 + \beta_1(WL_i^2) + \varepsilon_i \tag{4.1}$$

Where BW_i is width at WL_i, where WL_i is length in units of pixels (p) at the ith percentage of WL. The goodness of parabolic fit was also evaluated between 10% and 70% to determine if more area could be included without compromising the quality of the model. Models were evaluated using R^2 and p-values. Surface area was estimated from the fitted parabolas using Eq. 4.2.

$$SA_p = \left(\int_{WL_{60\%}}^{WL_{20\%}} (Eq.4.1_{s1})dx \right) + \left(\int_{WL_{60\%}}^{WL_{20\%}} (Eq.4.1_{s2})dx \right) \tag{4.2}$$

Where SA_p is the surface area in p between 20% and 60% of WL and Eq. 4.1_{s1} and Eq. 4.1_{s2} are parabolic models for side 1 and side 2, respectively.

A length normalized BCI was appropriate for our study because it facilitated body condition comparison among individuals and between observations of the same individual in the same way that body mass index (BMI) is used to compare body condition among humans (Flegal et al., 2012). $BMI = mass(kg)/height^2$ (Gallagher et al., 1996); to emulate this we used surface area as a surrogate for body mass and estimated body area index (BAI) using Eq. 4.3.

$$BAI = \frac{SA_p}{(0.4 * WL_p)^2} * 100 \tag{4.3}$$

Where SA_p is the estimated surface area of the whale in units of p, WL_p is the estimated length of the whale and Eq. 4.1_{s1} and Eq. 4.1_{s2} represent the parabolic models fit for side 1 and side 2 of the whale, respectively. The length was multiplied by 0.4 because surface area was only captured across 40% of WL. The multiple of 100 keeps the index > 1 for most whales. There were two advantages to the proposed BAI metric. A unit-less index has the benefit of being scale invariant and is not influenced by scaling errors that may arise during photogrammetry efforts, as described later. Also, as a length normalized index, BAI can be used to compare size of an individual between two time periods, size between two individuals, size relative to the larger population, and size relative to an established standard. Furthermore, a mean BAI for a population can be estimated from β_1 in Eq. 4.4.

$$SA_i * 100 = \beta_1((WL_i * 0.4)^2) + \varepsilon_i \tag{4.4}$$

Where SA_i is the SA of whale i, WL_i is the length of whale i, ε_i is error, and the intercept is zero to ensure SA is zero when $WL = 0$. Mean BAI (mBAI) = β_1 and is the mBAI for the population of whales used to train the model.

We developed five simulated change scenarios to examine the sensitivity of BAI to change at the population and individual level. We established a desired change sensitivity threshold at 10% because 10% is the low-end range of gray whale seasonal change observed by Rice and Wolman (1971). We simulated changes by applying the respective changes in WL and SA specified in Table 4.1 to the raw (*i.e.*, uncorrected) morphometric data. The simulation scenarios were intended to represent biologically meaningful change events. BAI1 is the unchanged scenario representing the current BAI of the whale. BAI estimates for each scenario and whale were analyzed for significant differences ($p < 0.05$ level) using a pairwise Bonferonni adjusted t-test of BAIs with the 'pairwise.t.test' function in R. Population sensitivity significance was assessed by comparing differences in mBAI using the SE of β_1 in the fitted models for each of the simulations.

4.3.3 Photogrammetric Method

We employed a vertical image photogrammetric method that used camera to subject distance (*i.e.*, range) to scale images from pixels to meters, similar to a previous method (Jaquet, 2006; Fearnbach et al., 2011). Aircraft altitude above sea level (ASL) was used as a surrogate for range with the assumption that images were captured at nadir, the subject was imaged at the sea surface and ASL was zero at sea level. Morphometric attributes were measured in pixels in each of the whale images using the *Whale Measurements* program developed in MATLAB (MATLAB and Image Processing Toolbox Release 2016b, The MathWorks, Inc., Natick, Massachusetts, United States). Pixel lengths were converted to metric lengths via ground sampling distance (GSD) using Eq. 4.5.

$$GSD_i = d_c * (H'_i + \varepsilon_{hi})/f_c \tag{4.5}$$

Where GSD_i was the ground projected horizontal surface distance represented by one side of a square image pixel (Comer et al., 1998), and $H_i^{'}$ was the ASL of the sUAS at the ith observation. d_c was the physical dimension of one side of a square pixel on the sensor chip in units of mm (*i.e.*, pixel pitch) and f_c was focal length in units of mm; both parameters were fixed and specific to camera c. The term ε_{hi} is the bias in ASL at the ith observation and accounts for the likelihood that aircraft barometer was zeroed above sea level. Figure 4.2 is a graphical depiction of these parameters.

Once GSD was calculated, objects were scaled from a pixel length measurement to a metric length estimate with Eq. 4.6.

$$L'_{ki} = (GSD_i) * OL_{ki}^p \tag{4.6}$$

Where $H_{ki}^{'}$ was the scaled length of object k at observation i, GSD_i was calculated for observation i, and OL_{ki}^p was the pixel length of object k at the ith observation.

4.3.4 Sources of Uncertainty in Photogrammetric Measurements

The primary sources of uncertainty in measurement estimates derived from vertical imagery (*i.e.*, nadir-pointing) are those related to the assumption of zero camera tilt, the uncertainty in the above ground level (AGL) flying height (and, hence, the image scale), and the uncertainty in the analyst digitization of the whale on the imagery (Dolan et al., 1978). Tilt uncertainty is the degree of uncertainty of the true pointing angle of the camera when the camera was assumed to be pointing nadir. Camera tilt was excluded from this analysis after five aircraft initializations on a level surface resulted in $< 3°$ of tilt error, which is under the conventional threshold warranting explicit correction (Philpot and Philipson, 2012; US Army Corps of Engineers, 2015). Ranging uncertainty in the context of this study refers to the uncertainty associated with estimating true sUAS AGL ($H^{'}$). $H^{'}$ is the only variable required for estimating GSD (Eq 4.5) and ultimately scaling images. Several factors contribute to ranging uncertainty and include: wind-driven changes in local barometric pressure, ocean swells, zeroing aircraft altitude above sea level (*e.g.*, ship deck), imaging a subsurface whale and imprecision inherent in low-cost barometric sensors. We define analyst digitization uncertainty as the uncertainty associated with estimating true object length in image space. Based on this definition, there are two primary drivers of digitization uncertainty: (1) the deviation from the assumption that the object being measured is flat level and perfectly orthogonal to the camera and (2) the uncertainty associated with an analyst manually measuring an object on an image.

4.3.5 Mitigating the Influence of Uncertainty Sources with Linear Modeling

A commonly employed ranging (*e.g.*, altitude) correction model and requires an object of known length (*e.g.*, calibration object) to calculate what we refer to as empirical GSD (eGSD)(Jaquet, 2006). eGSD is calculated by dividing known calibration object length (OL) in meters by estimated calibration object length and calculates the ASL ($H^{'}$) from which the object must have been imaged based on the geometric relationship between fixed camera parameters (*e.g.*, f_c) and OL^p in image space (*i.e.*, length in pixels). The correction model accounts for systematic error in ranging that results from zeroing the barometric altimeter above sea level and from an altimeter that exhibits bias. We calculated eGSD and regressed it against observed ASL ($H_i^{'}$) using Eq. 4.7 to estimate a corrected GSD (*cGSD*) using the 'lm' function in R.

$$cGSD_i = B_0 + B_1(H'_i) + \varepsilon_i \tag{4.7}$$

The *cGSDs* estimated from this model can be considered unbiased and observations from multiple flights and days can be aggregated to increase sample size and thus power, when certain underlying assumptions are met. In addition to the conventional assumptions associated with linear modeling (*e.g.*, independent and normal distribution of observations), Eq. 4.7 assumes (1) that the altimeter was always zeroed to the same height above sea level,

(2) local environmental barometric pressure did not change over the duration of the flight and (3) the pixel length measurements used to calculate eGSD were unbiased and precise. Violations of assumptions 1 and 2 lead to bias in the resulting GSD estimates.

Error in Eq. 4.7 may not be completely random due to unique conditions associated with range estimation for each individual flight per day (abbreviated as 'Date-Flight'), such as different take off locations, wind conditions, and ocean swell. We examined the possibility of a 'Date-Flight' effect on the relationship between GSD and altitude to determine if calibration should be performed on a per-day and flight basis. Calibration data were nested by 'Date-Flight' and repeated measures analysis of variance (ANOVA) (Girden, 1992; Weinfurt, 2000) with an error term for 'Date-Flight' was conducted on the raw calibration observation data to examine the effect of both altitude on the 'Date-Flight' grouping on GSD.

To account for the possibility that 'Date-Flight' has a significant effect on GSD, we modified Eq. 4.7 to be a linear mixed model (LMM) that includes a 'Date-Flight' error term (Eq. 4.8), using the 'lmer' function in the lme4 package in R (De Boeck et al., 2011; Bates et al., 2014).

$$cGSD_{ij} = (B_0 + u_{0j}) + (B_1 + u_{1j})\left(H'_{ij}\right) + \varepsilon_{ij} \tag{4.8}$$

Where $cGSD$ is estimated at H'_{ij} for the ith observation in 'Date-Flight' group j. u_{0j} is a random effect to account for changing intercept (*i.e.*, bias) by 'Date-Flight' and u_{1j} is a random effect accounting for changing slope (*e.g.*, barometric instability) for each 'Date-Flight' group.

Eq. 4.7 and Eq. 4.8 are both sensitive to violations of the assumption that OL measurements are accurate. Since digitization error is one of the primary error sources in photogrammetric measurements, it is highly likely that this assumption is violated in most instances. The geometric relationship between H' and OL^p is such that log10 transformed pixel length decreases as log10 transformed H' increases. Any deviation from 1:1 linearity suggests the presence of digitization error. We smoothed OL^p by estimating the mean pixel length of the calibration object at a given H' using the linear regression in Eq. 4.9.

$$log_{10}(OL_i^p) = (B_0) + (B_1) * log10\left(H_i^{'}\right) + \varepsilon_i \tag{4.9}$$

This model is appropriate for estimating OL^p if no substantial deviations from assumptions 1 and 2 described previously are evident. If these assumptions are violated, then a modified model form that includes a 'Date-Flight' error term is necessary. To account for this possibility we created linear mixed model specified in Eq. 4.10

$$log_{10}(OL_{ij}^p) = (B_{0j} + u_{0j}) + (B_{1j} + u_{1j}) * log10\left(H_{ij}^{'}\right) + \varepsilon_{ij} \tag{4.10}$$

When OL is unstable, smoothed OL^P from Eq. 4.9 and Eq. 4.10 can be subsequently used to estimate eGSD used in the linear regression Eq. 4.7 and Eq. 4.8, respectively. Eq. 4.11 and Eq. 4.12 were created to reflect the use of smoothed OL^P to estimate eGSD for the purpose of performance comparison. Eq. 4.11 and Eq. 4.12 are parameterized identically to Eq. 4.7 and Eq. 4.8.

$$cGSD_i = B_0 + B_1(H'_i) + \varepsilon_i \tag{4.11}$$

$$cGSD_{ij} = (B_0 + u_{0j}) + (B_1 + u_{1j})\left(H_{ij}^{'}\right) + \varepsilon_{ij} \tag{4.12}$$

4.3.6 Evaluation of Linear Model Corrections with Calibration Data

The linear models were applied to the calibration data in different combinations to facilitate comparison of estimated mean OL (Eq. 4.3) among the five correction methods specified in Table 4.2. The precision of mean estimated length was assessed using mean root mean squared error (Eq. 4.13) and bias. Coefficient of variation (CV) was calculated to assess relative measurement variation by correction method. 95% confidence intervals (CIs) were determined for all estimates using 2.5% and 97.5% quantiles from a nonparametric bootstrap of n = 1000 replicates. Non-parametric bootstrapping has been shown to be an effective estimator of standard error when sample sizes are small and true variance is unknown (Efron and Tibshirani, 1986). CIs for M2 through M5 included GSD model prediction uncertainty. Eq. 4.7 and Eq. 4.9 used a SE derived from the bootstrapping method described in Davison and Hinkley (1997) in conjunction with the 'boot' function in R Carlson and Ripley (1997) and the LMMs (Eq. 4.8, Eq. 4.10,Eq. 4.11 and Eq. 4.12) used 'predictinterval' function within 'merTools' (Knowles et al., 2016; Knowles and Frederick, 2016). The CIs for the LMMs do not account for the variance associated with the random error term (*e.g.*, 'Date-Flight'). CIs were used in conjunction with an ANOVA to discern difference of mean OL among correction methods.

$$mRMSE_s = \frac{\sum_{i=1}^{m_s} \sqrt{\frac{\left(\sum_{i=1}^{n_s} (x_{sij} - x_t)^2\right)}{n_{sj}}}}{m_s} \tag{4.13}$$

Where $mRMSE_s$ is the mean RMSE for the sth correction method, m_s is the number of flights for a given method s, n_{sj} is the number of observations within a given correction method and 'Date-Flight' (j), x_{ij} is the ith observation for a given method and jth 'Date-Flight' and, x_t is the true length of object t. The linear models were evaluated for substantial departures of the assumptions of constant variance and symmetric error distribution using residual plots.

4.3.7 Comparison of Linear Model Corrections on Whale Morphometrics

The purpose of the linear corrections was to ultimately to improve the precision and accuracy of whale morphometric measurements. The performance of a given correction method with the calibration data was expected to be a good indication of how the correction method would estimate any given morphometric attribute when applied to actual whale measurements. The GSD correction models (Eq. 4.7, Eq. 4.8, Eq. 4.11 and Eq. 4.12) trained in the correction methods in Table 4.2 were applied to the whale observation data to estimate *cGSD*. The means of scaled morphometric attributes were calculated for each whale. True morphometric attribute length was unknown, so the five GSD correction methods were evaluated with CV, CIs and graphically with boxplots. CIs were calculated using the nonparametric bootstrapping method previously described. The linear models were evaluated for substantial departures of the assumptions of constant variance and symmetric error distribution using residual plots.

4.3.8 Sensitivity Analysis

As ranging error (*e.g.*, AGL altitude) and analyst digitization error (*e.g.*, pixel length measurements) are two of the largest sources of uncertainty in vertical photogrammetric measurement estimates, we examined the relative impact of each of these components on object length estimates in the calibration data for both the blue and gray whale data sets. We conducted a sensitivity analysis using the special law of propagation of variance to estimate total propagated uncertainty when estimating object length using Eq. 4.5 and Eq. 4.6, assuming $H^{'}$ and OL^{P} are independent (Ghilani, 2011). Total propagated uncertainty was evaluated using Eq. 4.14. We estimated total propagated uncertainty at 15 and 40 m because approximately 95% of the imaging of the calibration objects occurred within this altitude range and these two extremes were expected to illuminate potential altitude-dependent trends in individual parameter influence on total uncertainty. Direct estimation of σ was not possible because Eq. 4.14 must be evaluated at a fixed altitude and the calibration objects were imaged across a range of altitudes. The variance of $H^{'}$ was determined using the SE of the predictor in Eq. 4.11 and the variance of OL^{P} was determined using the SE of the predictor in Eq. 4.9. LMMs (Eq. 4.10 and Eq. 4.12 were not used due a dependence on 'Date-Flight'. Eq. 4.11 was chosen for estimating $H^{'}$ variance because it was expected to be less influenced by variability within OL^{p} measurements.

$$\sigma_{OL'} = \pm\sqrt{\left(\frac{l}{f_c} * d_c\right)^2 * \sigma^2_{H'} + \left(\frac{H'}{f_c} * d_c\right)^2 * \sigma^2_{OL^p}} \tag{4.14}$$

Where, $\sigma_{OL'}$ is the standard uncertainty of the estimated object length, $\sigma^2_{H'}$ is the square of the standard deviation of ASL and $\sigma^2_{OL^p}$ is the square of the standard deviation of the number of pixels.

4.3.9 Image Extraction

The basis for measurement of individual whales in this study was the nadir pointing 4K video taken during sUAS flight. For each whale sighting, five full-resolution frames were extracted from the video using the 'snapshot' functionality in VLC Media Player (Version 2.2.4) following the pertinent recommendations of the error mitigation strategies described above. Five calibration images were also extracted during both take-off and landing of each flight. Effort was made to ensure the survey objective was centered, in focus, and that the full range of the altitude gradient was represented within the set of 10 images per flight.

4.3.10 Image Analysis

We developed a three-program analytical framework for photogrammetric morphometric analysis for the purposes of minimizing sources of analytical error and standardizing morphology measurements across multiple analysts and images. The first program is titled *Whale Calibration Object Measurement* and was developed in MATLAB for the purpose of standardizing the measurement of calibration objects as well as the pertinent outputs that facilitate mitigating the effects of uncertainty through linear modeling. The program prompts the user for the following inputs related to the sUAS camera, calibration object and specific flight: f_c, d_c, sighting number, flight number, date, object name, known OL in mm, observe $H^{'}$, and height difference between sUAS initialization location and the calibration object (parameterized as ε_{hi} in Eq. 4.5). The user is guided through an interactive measuring process and a summary table is produced that includes the prompted inputs, OL^p, and GSD (as calculated from Eq. 4.5).

The second program, titled *Whale Measurements*, was also developed in MATLAB, for the purpose of standardizing the measurement of morphometric attributes in vertical whale images collected from sUAS, and standardizing the output measurements for subsequent scaling and error assessment in a third program described later. The program interactively requests an input image, and guides the user through several processes in the following sequence: (1) prompts user for details pertinent to the specific camera, date, flight, sighting and individual whale, (2) crops the image to the subject of interest, (3) aligns the whale lengthwise across a horizontal axis with an origin of 0,0 to simplify calculations, (4) guides the user through a series of measurements to collect the morphometric attributes specified in Table 4.3, (5) prompts the user for flight ASL ($H^{'}$) and the vertical difference between sUAS initialization location and sea level parameterized as ε_{hi} in Eq. 4.5) and (6) outputs a table of results and an image (*e.g.*, Fig 4.1) containing the subject with an overlay of pertinent metrics.

The third program, titled *Whale Quantitative Analysis*, was developed in R 3.3.1 (R Core Team, 2017) to process the summary tables produced in the first two programs to provide measurement data and error estimates. This program queries the user for directories that contain the summary tables produced in the previous two programs. Data are automatically grouped by date, flight and sighting under the assumption that all data shares a common calibration object and camera. The program builds the GSD correction models from the calibration data and applies the corrections to both the calibration data and whale morphometric data to create the five GSD correction methods in Table 4.2 for each whale. Summary tables of the calibration data include RMSE, bias, CIs, CVs and the estimates of total propagated uncertainty (Eq. 4.14) grouped by correction method. Summary tables of each whale and associated morphometric measurements include the estimated scaled value, CIs and CVs. The results presented below are based primarily on data collected during sUAS survey of 89 gray whales in Oregon, USA collected in 2016 and also include results from the data collected during sUAS survey of 6 pygmy blue whales in New Zealand collected in 2016.

4.4 Results

4.4.1 Gray Whale Calibration Object Correction Method Comparisons

Image measurements of the 1 m calibration object were examined in the context of the GSD correction methods. The blue whale calibration data were similarly analyzed but the results were excluded from this manuscript to economize space since the trends were largely similar. Data were filtered due to high variability (CV $>$ 15%) of estimated scaled object length. Twenty-two of the 193 observations were removed due to uncorrected estimated lengths > 1.97 SDs $(t, = 0.025, df = 192)$ from the mean of all measurements.

The linear models used for the GSD correction methods (Table 4.2) were visually evaluated for violations of non-constant variance and non-random error using plots of predicted vs residuals. Eq. 4.7 and Eq. 4.11 residuals displayed structural trends that indicate a violation of the assumption of error non-heteroscedasticity. Non-heteroscedasticity indicators were not visible in the LMMs (Eq. 4.8 and Eq. 4.12) residuals which suggests 'Date-Flight' grouping accounted for the non-random error.

The effect of a 'Date-Flight' grouping of calibration data was examined to determine if a per-flight imaging of the calibration object was necessary. A 'Date-Flight' grouping conceptually accounts for systematic variances in altitude that ultimately increase uncertainty in GSD and subsequent scaled-length estimates. Results of the repeated measures ANOVA indicated a significant effect of 'Date-Flight' on *GSD* at the $p < 0.05$ level $[F(1,39) = 2.93 * 10^{31}, p < 0.0001]$. Mean Squared Error of 'Date-Flight' accounted for approximately 27% of the overall variance in *GSD*. The remaining variance in *GSD* was attributed to altitude which was 73% of the variance in *GSD* at the $p < 0.05$ level $[F(1,132) = 2.64 * 10^{32}, p < 0.0001]$. These results suggest that it is appropriate to include 'Date-Flight' as a random error term in the linear model correction of *GSD* as reflected in Eq. 4.8. The same per-flight variance can be expected to influence the pixel length smoothing model (Eq. 4.9), justifying the inclusion of the 'Date-Flight' term in Eq. 4.10. These results suggest that a GSD correction including 'Date-Flight' as an explicit error term is appropriate (*e.g.*, M3 and M5).

Results of the calibration object correction methods listed are presented in Fig 4.3. Tabulated results that include mRMSE, bias, CV, CIs and estimated mean *OL* appear in Appendix Table A.1.

Measurement error (RMSE) and bias (over/under estimation) were the primary metrics for examining model performance among the five alternatives in Table 4.2. M1 (uncorrected) resulted in the largest RMSE and bias of the five alternatives. The inclusion of the 'Date-Flight' term reduced RMSE in M5 and M3 compared to M4 and M3, respectively. CV and CIs of M2 – M5 were larger than M1 due to model-induced uncertainty. CIs for M4 and M5 were narrower than those for M2 and M3 as a result of the improved predicted performance when using smoothed OL^p estimates for calculating eGSD. Fig 4.3b further corroborates the improved estimation performance of including 'Date-Flight' error term by the absence of the skew evident in the M2 and M4 observations. M5 has the lowest RMSE and bias of the alternatives as a result of accounting for pixel length variation when estimating the eGSD used to train the model as well as accounting for the 'Date-Flight' effect. Operationally, these results suggest that per-flight systematic sources of uncertainty (*e.g.*, initialization height, swell, *etc.*) influenced ASL ($H^{'}$) estimation at a level of significance that warrants continuing per-flight imaging of a calibration object.

4.4.2 Comparison of Correction Methods on Gray Whale Measurements

Whale morphometric attribute estimates (Table 4.3) based on the five correction methods in Table 4.2 displayed similar trends as calibration object results above. Figure 4.4a shows estimated WL by correction method for 9 of the 89 whales imaged. Appendix Table A.2 displays the CV and mean of WL for these whales. The no-correction (*i.e.*, uncorrected) method (M1) estimated a longer WL for 6 of the 9 gray whales depicted in Figure 4.4a, which was consistent with the relationship between M1 and the M2-M5 in the calibration data. The CV of the WLs for these six whales was $< 5\%$. However, for Whale 1, Whale 4 and Whale 5, WL in M3 was greater than WL in M4. This break in the trend was attributed to high CV ($> 5\%$) which was indicative of a significant and uncorrected digitization and/or ranging error. M1 WLs for these three whales were not significantly different from M5. In contrast, the M1 WLs for the other six whales was always greater than M5. This trend was consistent with the relationship between M1 and M5 in the calibration object data and provides evidence that M5 was appropriately correcting the morphometric estimates.

4.4.3 Gray Whale Morphometric Measurement Correction Method Comparisons

An identical analysis on the six blue whales surveyed in New Zealand was conducted to demonstrate the applicability of the software tools to another baleen whale species. WL estimates of each whale for the five correction methods appear in Figure 4.4b. Methods M2-M5 produced effectively identical WL estimates for each whale indicating a consistent zeroing of the aircraft altitude and very little discernible bias in the digitization process among whales and flights. These results further corroborate the evidence that M5 produces accurate results even when 'Date-Flight' influences are negligible. Eight individuals are depicted in Figure 4.4b because Whale 2 and Whale 4 were duplicates of Whale 1 and Whale 3, respectively. Whale 1 M1 WL is 18.77 m and Whale 3 M1 WL is 18.20 m, similarly, Whale 2 M1 estimated length is 18.33 m and whale 4 M1 is 19.33 m. Differences were significant ($p < .05$) and just outside of the 95% confidence intervals. These whales were imaged over two flights and the results were kept separate to illustrate the how a small number of image observations ($n < 4$) can produce measurements with low variation that contain undetectable bias and a false sense of certainty, reinforcing the need to analyze a minimum of five good frames (or images) per whale.

4.4.4 Total Propagated Uncertainty

The results of the total propagated uncertainty analysis appear in Table 4. As has been observed in the previous results, the gray whale data was more variable than the blue whale data. We attribute this difference in variability to two key differences: (1) blue whales tended to be centered and fully elongated in the images more frequently than gray whales, likely due to behavioral differences between species (foraging gray whales are more bendy than blue whales at the surface), and (2) the calibration reference used during the blue whale study was an object on the vessel at water-level that was less susceptible to pitching and yawing from ocean swells than the 1 m board used for the gray whale study. The relative contribution of $\sigma^2_{H'}$ was lower than $\sigma_O L^{'}$ for both gray whale and blue whale data which indicates that the barometric altimeter used to estimate $H^{'}$ is linear and relatively stable. The large OL^P values were a result of digitization error that was a function of user error during the digitization process and poor quality images (e.g., glare, off-center imaging). The variability in OL^p was clearly discernible in the Pixel length vs altitude plot (Appendix Figure A.5). If bias in $H^{'}$ had not been corrected from Eq. 4.7, $\sigma^2_{H'}$ would have been more influential at 15 m than 40 m due to the fact that the influence of $\sigma^2_{H'}$ increases as $bias/H^{'}$ increases.

4.4.5 BAI assessment

The parabolic models used to derive surface area for the BAI calculation were examined using adjusted R^2 and visual analysis of the agreement between each parabola and the side of the whale. R^2 values ranged from 0.29 to 0.98 with no observable relationship between poor R^2 and a particular side or particular whale. Poor R^2 was associated with poor whale edge visibility that frequently led to erroneous point placement and tended to increase the uncertainty of a whale's BAI (*e.g.*, wide Whale 5 CI vs narrow Whale 9 CI). BAI is scale invariant and thus unaffected by uncertainty associated with ranging error, so a comparison among correction methods was unnecessary. BAI allows comparisons among and between populations so it is important that BAI provide a size metric that is independent of WL. To evaluate independence we calculated Pearson correlation coefficients between both WL and SA, and WL and BAI and determined BAI ($R^2 = 0.11$) was substantially more independent of WL than SA ($R^2 = 0.11$).

We examined the sensitivity of BAI to detecting a 10% change in individual whale size using the change simulation scenarios in Table 4.1. Difference was assessed with a pairwise t-test comparison (Table A.1). The estimated BAIs by whale and scenario (for the 9 whales in Figure 4.4) appear in Figure A.6.

Individual BAI change from BAI1 to BAI2, BAI3, BAI4 and BAI5 was discernible in 62.2%, 28.1%, 28.1% and 27.0% of the 89 whales, respectively. Despite the inconsistent performance of BAI for detecting individual change, using SA directly to discern change from SA1 to SA3, SA4 and SA5 was only successful for 17% of the whales. SA2 was not evaluated because SA does not change. The pairwise comparisons in Table 4.5 and Table 4.6 examine how well change scenarios can be discerned from one another. The generally poorer performance in SA alone was likely the result of the added uncertainty from scaling (*e.g.*, GSD error). However, the one case where SA was superior to BAI was distinguishing SA4 from SA5, which was more frequently discernable than BAI4 from BAI5. In terms of BAI, these two scenarios produced nearly identical BAIs but very different SAs. These results suggest that using BAI to detect 10% change in individual whale size may not be reliable but will likely perform better than using SA directly. However, the fact that SA could adequately capture the simulated difference between BAI4 and BAI5 suggests that the SA metric should not be eliminated from evaluation when examining individual trends.

However, BAI appeared to perform well when examining change at the population level (Figure 4.5 and Appendix Figure A.6). mBAI's (red line Figure 4.5) of scenarios BAI2 – BAI4 were distinguishable from the mBAI of BAI1, although the large SEs obscured individual whale changes across the scenarios. The increased sensitivity at the population level was a function of discriminatory power offered by the larger sample size (*e.g.*, n = 89).

4.5 Discussion

Results indicated that accounting for ranging error with some type of calibration object is imperative. Uncorrected object lengths contained substantial bias (Figure 4.3) that exceeded the bias reported of the calibration object reported in Durban et al. (2015), although their method used a much longer calibration object that was more robust to confounding scaling through movement on sea surface, and the precision of the barometric altimeter was likely superior to that of the DJI Phantom. The results from the M5 estimates compared to the other four correction methods further suggest that smoothing pixel lengths of the calibration object prior to creating the GSD correction model resulted in less erroneous estimates of scaled length. While there were instances where M5 estimates of WL were not significantly different from WL estimates in M1, these exceptions tended to be associated with high levels of variability in the observations. High variability in the observations was likely a function of non-strict adherence to the optimal imaging recommendations presented by Christiansen et al. (2016) (*e.g.*, whale not centered in the camera during flight) but could potentially be overcome by substantially increasing the number of video frames or images analyzed for each whale.

The analysis of total propagated uncertainty showed that digitization error remained the largest source of uncontrolled error, further corroborating the need for more observations per whale and ensuring images are of high quality. Our uncertainty analysis also showed the importance of correcting for bias in altitude estimates with a calibration object. These findings lead us to recommend that surveys should be conducted at the highest reasonable altitude to minimize the $bias/altitude$ ratio when bias in altitude cannot be corrected from a calibration object.

We condensed our recommendations in the form of a 'uncertainty mitigation protocol' as a convenience for what we believe will be a rapidly growing community of whale photogrammetrists. This protocol is not specific to our study or even whales, but rather, is broadly applicable to any study where the subject is a surfacing animal and the survey aircraft is a sUAS with a nadir pointing camera.

4.5.1 Uncertainty Mitigation Protocol

1. Power-up (*i.e.*, initialize) the sUAS from the same location on the watercraft every time to minimize influence of random error in ranging uncertainty.

2. Measure the vertical distance between power–up location and water level and add that distance to reported altitudes to account for bias in ranging caused by initializing above sea level (Durban et al., 2015).

3. To the extent feasible, image over flat water in non-windy and swelly conditions to minimize ranging uncertainty.

4. Only measure images/frames where the whale is at water surface, fully elongated with no curvature (Perryman and Lynn, 1993; Fearnbach et al., 2011; Christiansen et al., 2016) to minimize digitization error.

5. Measure 5+ images of the same subject from each flight to evaluate variation.

6. Keep subject centered in the camera to minimize error associated with lens distortion and scale non-uniformity induced by camera tilt error.

7. Image a calibration object every flight. Object should rigid and be located as close to sea level and be as long as possible. Longer objects are more robust to uncertainty induced by imaging in swelly conditions.

8. Image from the highest safe and legal altitude that ensures adequate level of detail, to minimize the influence of altitude bias on scaling error.

9. Apply a GSD correction model like that in Eq. 4.12 that accounts for flight-level altitude variances.

The final objective of this study was to develop and present a length-independent body condition metric, which we term Body Area Index (BAI), to facilitate comparison of whale body condition over time, among and between populations. We demonstrated that BAI was more independent of WL than SA and that population level changes were detected well below the 10% threshold we established. The scale invariant property of BAI is especially valuable in surveys where scaling error cannot be controlled with a calibration object as was evidenced by the increased sensitivity of BAI to detecting change in body size over that of using SA directly. The results of the individual BAI change sensitivity analysis were inconclusive because change could be detected in some whales and not others. We attributed this inconsistency to high within whale BAI variability relative to the low sample size (*e.g.*, five images per whale).

When individual change detection is necessary, we recommend conducting multiple flights over the same whale and performing a power analysis to determine how many observations (*e.g.*, images/frames) will be necessary to discern change at the desired level of sensitivity. Several useful characteristics of BAI are demonstrated in Figure 6. For example, in scenario 1, Whale 10 is clearly larger than the others with a mBAI of 56 and a length of 13.5 m. Conversely, Whales 2 and 4 have a similar length (12.6 m), but both have a mBAI of 34. Based on this large difference, we can infer that Whales 2 and 4 likely have reduced fat reserves relative to Whale 10. The status of calves can be similarly evaluated. Many of the whales < 10 m in length have mBAIs equal to or higher than the population mean (red line) suggesting that these surveyed calves have slightly elevated fat reserves compared to the population.

A potential limitation to BAI is the underlying assumption of a parabolic shape that is used to estimate surface area. Pregnant or severely emaciated whales may present forms that deviate from that of a parabola. Additional investigation is necessary to discern the appropriateness of the method on those individuals. The method presented here would potentially be improved upon by incorporating convex hull algorithms (Barber et al., 1996) or automatic segmentation algorithms (Misimi et al., 2008))

4.6 Conclusion

This study presented a length-normalized body size index (BAI) to facilitate comparison among individuals and populations and can be used to describe population size change trends. We examined the effectiveness of models used to correct error in scale image measurement and determined the most precise and accurate model was a LMM containing a 'Date-Flight' error term and regressed on eGSD values that were derived from smoothed pixel length estimates. We subsequently determined that analytical digitization error was the largest source of uncertainty in scaled measurement estimates and developed an 'Uncertainty Mitigation Protocol' to help future studies avoid controllable sources of uncertainty. We also developed a three-program analytical suite for obtaining 11 morphometric attributes of free-swimming baleen whales from vertical sUAS imagery. Our findings suggest that sUAS photogrammetry from a DJI Phantom 3 Pro is a precise method to assess baleen whale body size when there are sufficient observations of an individual whale and uncertainty from ranging error can be controlled by imaging an object of known length every flight. Future studies will determine the broader applicability of our provided framework, but believe similar results can be achieved on any species exhibiting similar morphological characteristics as gray and blue whales. We expect that future studies will focus on automatic whale edge delineation in images and further investigate the applicability of BAI for ecological inference.

4.7 Figures

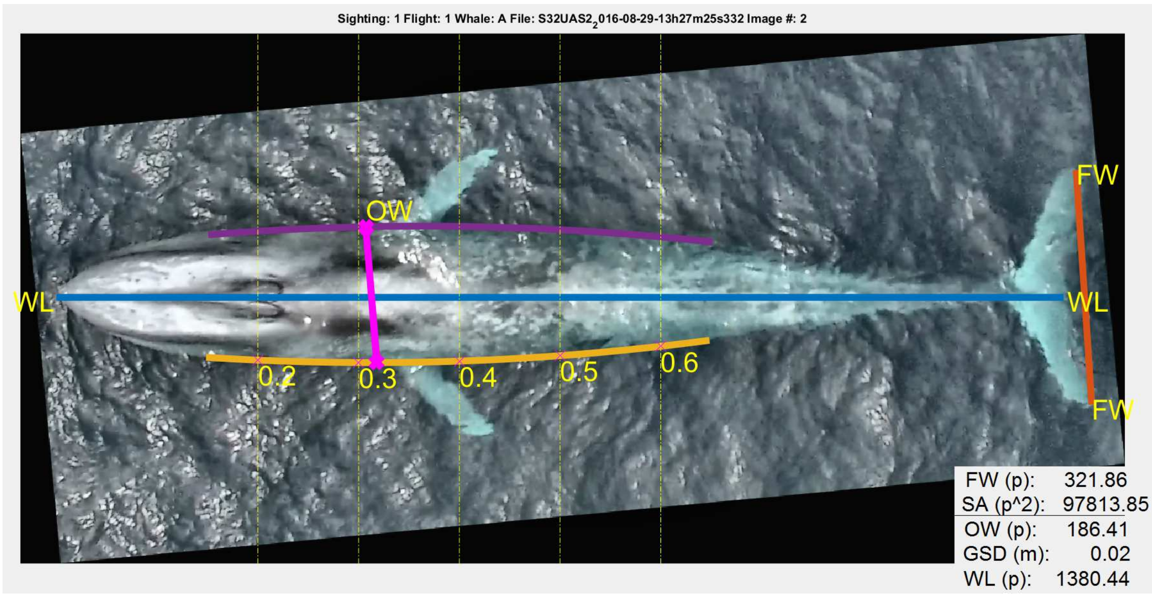


Figure 4.1: Output image from *Whale Measurements* program displaying the morphometric attributes that were measured, including fluke width (FW) and whale length (WL). Parabolas fit by Eq. 4.1 are also depicted. OW is the 'optimized width' estimated by the points on the parabolas nearest to the end points of interpreter defined manual width.

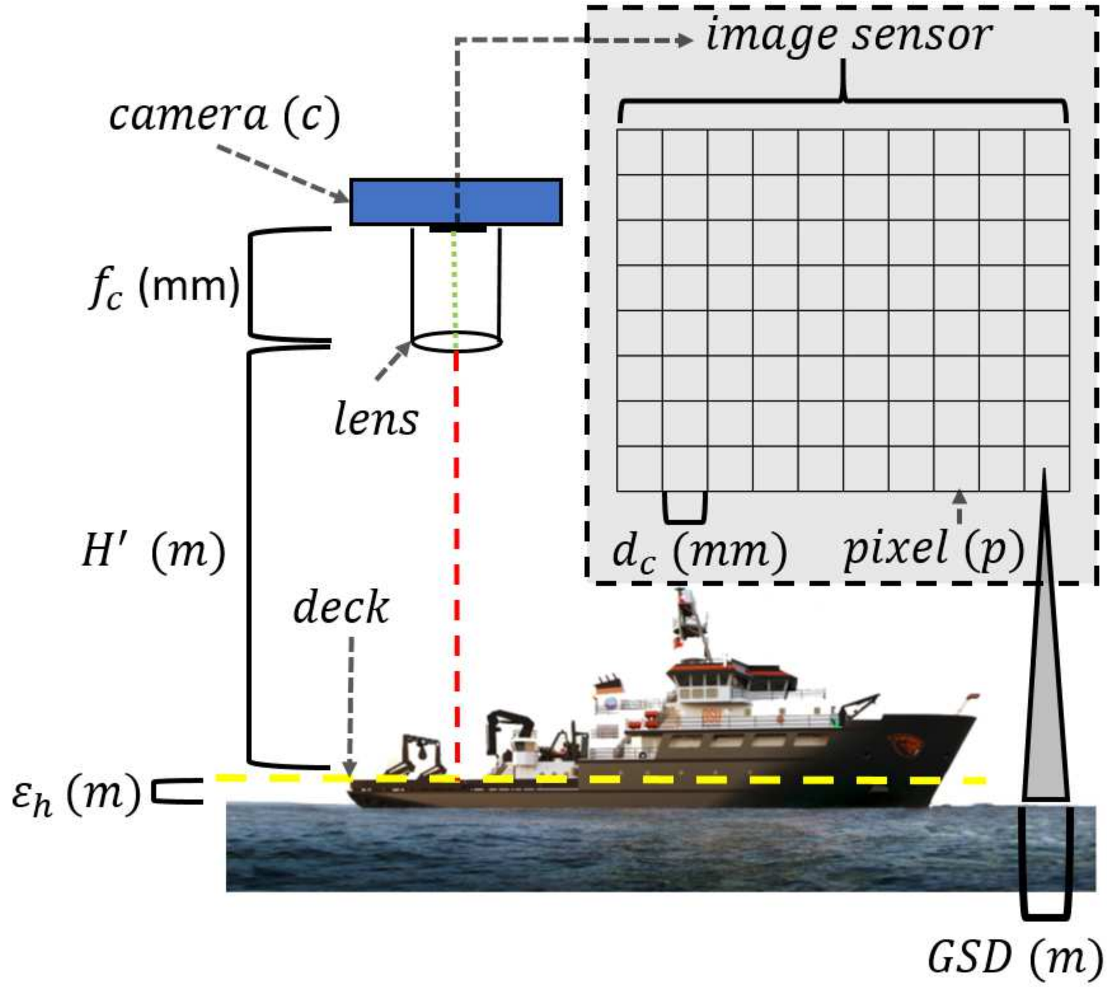


Figure 4.2: A graphical depiction of the pertinent parameters used to estimate scaled object lengths from measurements. *GSD* = ground sampling distance (*i.e.*, ground distance of one pixel), d_c = pixel pitch, $H^{'}$ = altitude, and f_c = focal length.

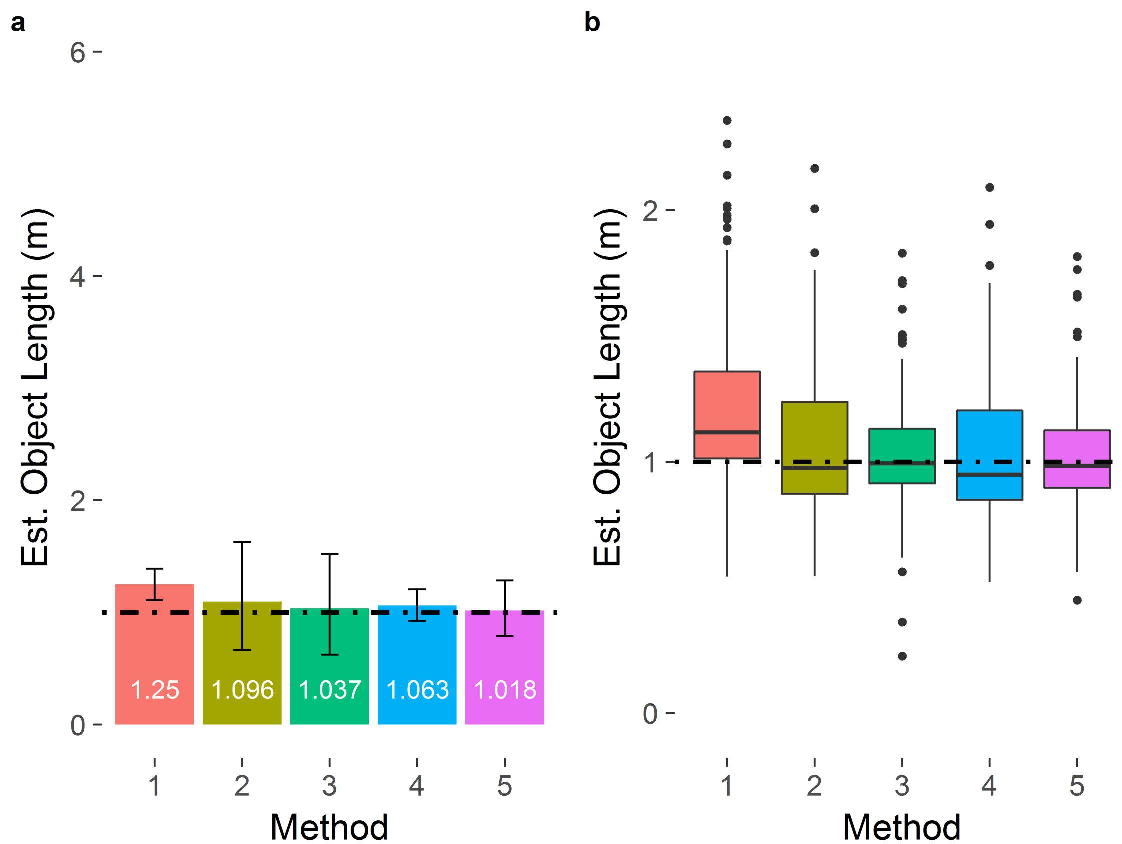


Figure 4.3: Estimated object lengths of the five correction methods listed in Table 4.2. (a) is a bar plot of mean estimated lengths and error bars are 95% nonparametric bootstrapped CIs. The dashed horizontal line is actual object length and (b) boxplot of estimated lengths for individual observations. Method 1 (M1) is the uncorrected method.

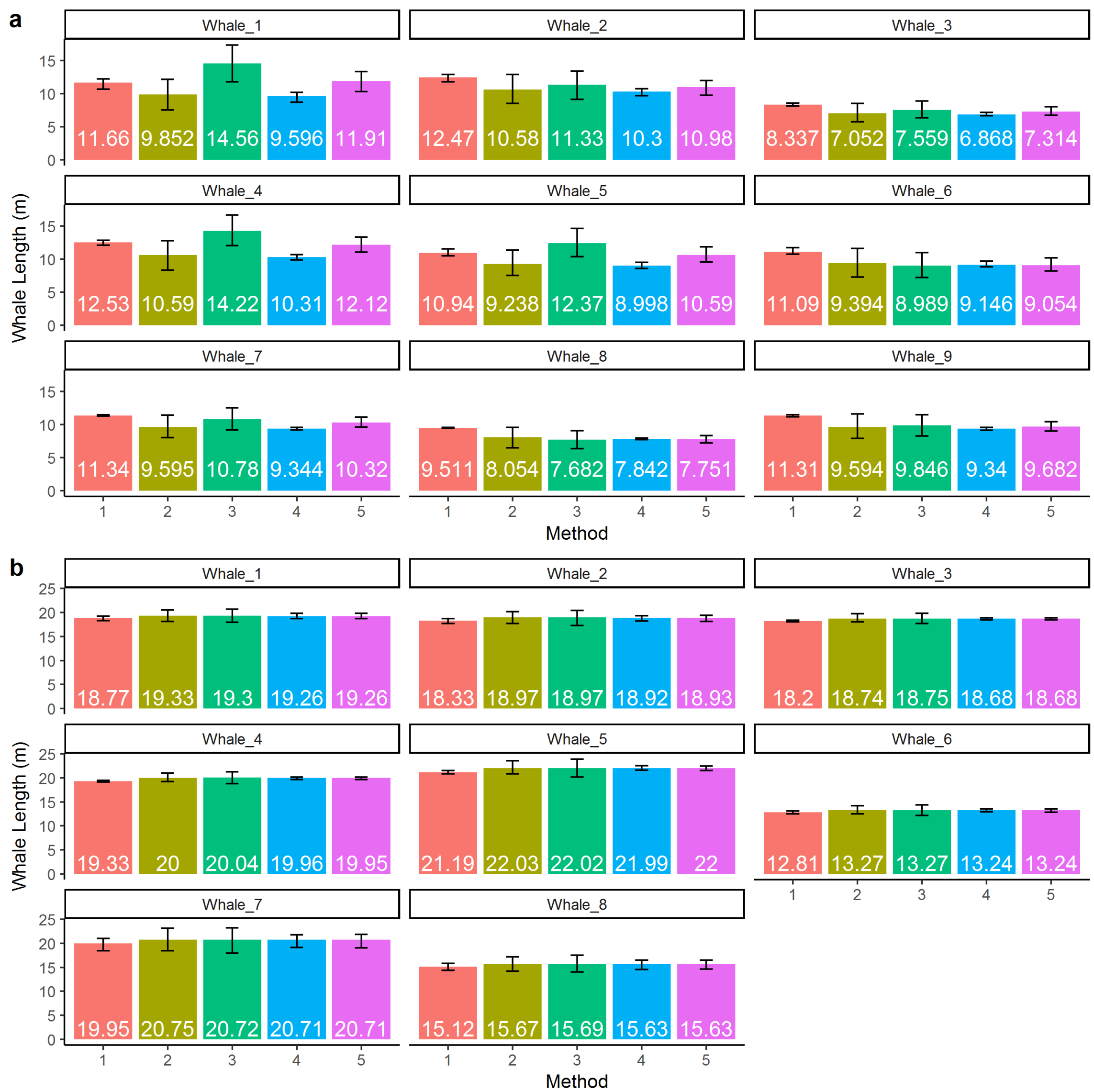


Figure 4.4: Estimated WL for each of the five GSD correction methods (Table 4.2 for each of nine arbitrarily selected gray whales (a) and each of the six blue whales imaged (b). Measurements for eight whales appear in Figure 4.4b because Whale 1 and Whale 2 are the same individual imaged over two flights; the same is true for Whale 3 and Whale 4. Bars are the 95% bootstrapped confidence intervals.

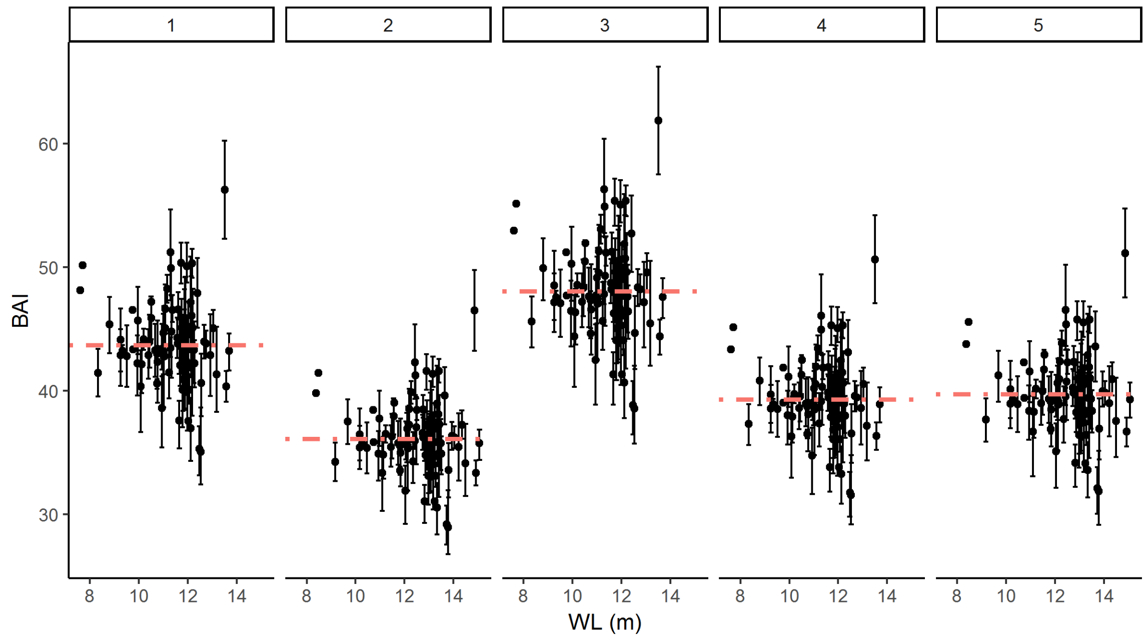


Figure 4.5: Relationship between body area index (BAI) and estimated whale length (WL) for the gray whales ($n = 89$), ordered by the five change scenarios listed in Table 4.1. Headings denote the scenario number. The red dashed line is estimated mean BAI (mBAI) for the scenario and is derived from the slope term in Eq. 4.4. Error bars are the 95% CIs around the BAI for each whale. Points missing bars have insufficient observations to derive a CI.

4.8 Tables

Table 4.1: Simulated change scenarios devised to discern how changing whale length (WL) and surface area (SA) influence body area index (BAI) estimates.

Scenario	**WL**	**SA**
BAI1	Unchanged	Unchanged
BAI2	+10%	Unchanged
BAI3	Unchanged	+10%
BAI4	Unchanged	-10%
BAI5	+10%	+10%

Table 4.2: Different correction methods examined for estimating GSD and eGSD using the regression models developed previously. eGSD describes how eGSD was calculated for the purpose of training the model listed under the GSD column and applies strictly to the calibration object estimates. GSD describes the GSD estimation method being used in Eq. 4.6 to estimate scaled length and is applicable in the context of both the calibration data and the whale morphometric measurement data.

Method	**eGSD**	**cGSD**
M1	NA	Eq. 4.5
M2	$OL \div OL^P$	Eq. 4.7
M3	$OL \div OL^P$	Eq. 4.8
M4	$OL \div$ Eq. 4.9	Eq. 4.11
M5	$OL \div$ Eq. 4.10	Eq. 4.12

Table 4.3: The morphometric attributes produced by the analytical programs and their descriptions.

Parameter	**Description**
WL	Whale length —rostrum to notch in tail
MW	Manual width —manual measurement of width at widest point
OW	Optimized width —width at point on parabola nearest MW
FW	Tail width —tip to tip fluke width
W20	Width at 20% of WL from rostrum
W30	Width at 30% of WL from rostrum
W40	Width at 40% of WL from rostrum
W50	Width at 50% of WL from rostrum
W60	Width at 60% of WL from rostrum
SA	Surface area between 20% and 60% of WL
BAI	Body Area Index

Table 4.4: Results of the total propagated uncertainty analysis, and supporting input parameters for Eq. 4.14 in context to estimating the length of the calibration object (OL) from vertical imagery. Heading names refer to the respective data set (*e.g.*, Gray = gray whale; Blue = blue whale) and numbers in the headings represent the altitude (*e.g.* 15 and 40 m) at which total propagated uncertainty was evaluated. Actual object length is abbreviated as $OL^{'}$. % Uncertainty is Total Uncertainty / $OL^{'}$.

Parameter	**Gray 15**	**Gray 40**	**Blue 15**	**Blue 40**
$H^{'}(m)$	15	40	15	40
OL^{p}	164.2	69.7	701.2	238.8
$\sigma_{OL^{p}}$	51.5	19.4	37.5	12.9
$\sigma_{H'}$	0.15	0.15	0.14	0.14
$\sigma^2_{OL^{p}}$	99%	99%	97%	97%
$\sigma^2_{H'}$	1%	1%	3%	3%
$OL^{'}(m)$	1	1	4.41	4.41
Total Uncertainty (m)	0.29	0.32	0.24	0.21
% Uncertainty	29%	32%	5%	5%

Table 4.5: Comparison of body area index (BAI) for all five change scenarios in Table 4.1 using Bonferonni adjusted pairwise *t*-test. Percentages are the ratio of the 89 whales that exhibited a significant change ($p < 0.05$ at 95% significance) when comparing the BAI computed in the respective change scenario specified in the column heading to the BAI computed in the change scenario specified in the row name.

	BAI1	**BAI2**	**BAI3**	**BAI4**
BAI2	65.2%	NA	NA	NA
BAI3	28.1%	77.5%	NA	NA
BAI4	28.1%	13.5%	70.8%	NA
BAI5	27.0%	19.1%	70.8%	0.0%

Table 4.6: Comparison of surface area (SA) for all five change scenarios in Table 4.1

	SA1	**SA2**	**SA3**	**SA4**
SA2	NA	NA	NA	NA
SA3	16.9%	16.9%	NA	NA
SA4	16.9%	16.9%	59.6%	NA
SA5	16.9%	16.9%	70.8%	59.6%

Acknowledgements

Gray whale overflights were authorized under NOAA/NMFS permit #16011 issued to John Calambokidis. Special thanks to Dr. Chris Parrish for editing and advice on analyzing photogrammetric accuracy. We also acknowledge the contributions of Alessandra Hossley, Ashley Leupold, and Taylor Mock for their efforts in data processing.

5 Discussion

The overarching goal of this dissertation was to explore the efficacy of remote sensing surveys with sUAS in complex, dynamic environments, and highlight the unique potential these instruments offer to conducting scientific inquiry. Chapter 2 presented a novel method to detect Swiss Needle Cast disease using a Generalized Additive Model to classify survey data collected with an sUAS equipped with a consumer grade camera. Chapter 3 investigated the efficacy of the presented SNC detection method in context to surveys conducted with both a consumer grade and commercial camera, as well as surveys conducted in two different months and in two different years. Chapter 4 introduced an error minimization strategy and standardized method for conducting low-cost sUAS photogrammetric survey and morphometric analysis of blue and gray whales. The three chapters combined present a robust assessment of sUAS remote sensing methods in remote and dynamic environments. The disease detection and morphometric analysis methods presented here are novel to their respective subfields within environmental remote sensing and demonstrate the unique capability of sUAS.

The primary objectives of Chapter 2 were to: (1) examine whether trees can be classified as diseased or non-diseased with better than random chance, (2) discern if the use of a NIR camera significantly improves classification reliability, (3) determine whether physiologically meaningful VIs improve classification reliability, and (4) compare SNC detection reliability among two different MLAs and two different GLM implementations using nine different model specifications to examine the effect of classification model on detection reliability. The individual-tree SNC detection method in Chapter 2 demonstrated the ability to reliability (kappa > 0.4) distinguish visibly diseased from non-visibly diseased trees, and specify the diseased status of a given tree accurately (PPV > 0.7). The added complexity and cost of flying a second NIR camera was determined to be unnecessary because NIR spectral information did not improve reliability or detection accuracy. Moreover, the inclusion of vegetation indices relating to crown structure (NDVI) and chlorophyll content (TGI) did not improve the accuracy or reliability of detection surveys. The inability of NIR and VIs to improve detection was attributed to noisy background added by the broadband sensitivity of the consumer grade cameras employed in the study. Finally, the algorithm used for detection had significant implications on detection reliability and accuracy depending on the covariates specified. However, the generalized additive model using simple, untransformed, visible spectral information was the most consistently reliable across the four sites (kappa > 0.4). The presented individual-tree SNC detection method was heavily dependent on accurate training data and tended to perform best in homogenous Douglas-fir stands. Extension of the method to other forest types, especially mixed species forests, or types with a wide age-class distribution, requires further investigation.

The primary objectives of Chapter 3 were to: (1) determine if SNC detection surveys with a multispectral camera were more accurate and reliable than those conducted with a consumer grade camera, (2) evaluate the reliability of between year change detection at the site-level and tree-level, (3) investigate the plausibility of conducting summer SNC detection surveys with sUAS, and (4) examine the relationship between field-based foliage retention estimates and disease status as determined by remote sensing. The addition of the multispectral did not improve results significantly improve the reliability or accuracy of SNC detectability at all sites. Results were similar between the consumer grade camera and the multispectral camera at most sites. However, the multispectral camera facilitated improved SNC detectability on sites where confounding influences of perspective effects (*i.e.*, bi-directional reflectance) were apparent. Between year change detection was heavily influenced by external factors such as changing lighting condition, and resulted in several hundred trees changing disease status between years at each of the four sites. However, employing more strict classification thresholds utilizing the underlying individual-tree disease status probabilities resulted in very few trees changing status. Findings suggest that summer surveys are not plausible, likely due to spring leaf flush and lighting conditions masking the evidence of reduced needle retention in infected crowns. There was no evidence of a relationship between foliage retention and individual-tree probability of disease status revealed no evidence of a relationship. Lighting conditions heavily influenced detection reliability, but the effect can be minimized by using a narrowband multispectral camera and enforcing tighter classification probability thresholds. The primary contributions of this study are the corroboration of the effectiveness of the individual-tree SNC detection method over an expanded scope that includes two sensors, two survey years and more sites. This findings of this study also provide recommendations regarding the survey timing as well as optimal lighting conditions and sensors.

The primary objectives of Chapter 4 were to (1) establish methods for conducting accurate and repeatable UAS photogrammetric surveys that do not require scaling objects to be co-located with the survey subject, (2) thoroughly evaluate sources for measurement uncertainty, (3) examine strategies to reduce measurement uncertainty, (4) develop standardized methods for extracting whale morphometrics from vertical sUAS imagery, (5) introduce a length normalized surface area measure that is robust to scaling error, and (6) disseminate these methods in the form of freely-available MATLAB and R scripts. The method presented demonstrated the ability to make precise and accurate estimates of an object's length when linear mixed models were employed to correct altitude induced bias when scaling pixel measurements to metric units. The sensitivity analysis revealed that measurement error, resulting from poor image quality and subjectivity during image measurement, was the largest contributor to total uncertainty. A mitigation protocol was devised from recommendations existing within available literature and experience from conducting photogrammetry operations on an ocean-going vessel. The standardized method for extracting morphometrics was presented in the form of a program that reduced enumerator fatigue, minimized opportunities for transcription error, and standardized whale measurements by guiding the analyst through measurements. The normalized body surface area measure, labeled body area index (BAI), was shown to be scale invariant (and thus robust to scaling errors), and independent ($R^2 = 0.11$) of body length. BAI provided a means of comparing whale body condition within (*i.e.*, over time) and between individuals and was shown to be more sensitive to detecting changes in body condition than surface area alone.

Image quality and resulting morphometric measurements in this study were heavily influenced by the choice to use 4K video and by measuring images of whales that were off-center or bending. These factors combined with the non-stationary nature of gray whales substantially reduced measurement precision. The primary outputs of this research are (1) a new morphometric attribute, BAI, that is similar to BMI in humans and provides a similar level of comparability among whales, (2) a software suite that standardizes morphometric measurements, and (3) an uncertainty mitigation protocol that provides guidelines to the broader community for how to minimize error induced during survey operations.

Despite the generally positive findings from all three studies, there is room to improve the methods employed to produce more useful results. The SNC detection study would benefit from an image-based crown canopy segmentation method that better captures the unique shape of each crown and better ensures that corresponding spectral information is included in the disease detection assessment. Furthermore, the method would benefit from additional study on the functional relationship between SNC detectability and the combination of crown color and crown structure. This information could be combined with customized multispectral cameras to better isolate light spectra that relate to detectability. In context to the baleen whale study, results could be improved by using full resolution still images rather than frames from the 4K video. Furthermore, employing the error mitigation protocol during future data collections would likely improve the precision of the resulting measurements. Finally, the method could be vastly improved in terms of efficiency and objectivity by utilizing automatic segmentation methods to delineate whales within the images.

5.1 Conclusions

Environmental remote sensing with sUAS was shown to produce detailed and accurate survey results in remote and dynamic environments. The innovative individual-tree Swiss needle cast detection method demonstrated the power of sUAS surveys to accurately detect diseased trees and reliably distinguish visibly infected trees from non-visibly infected trees. This information can be presented in the form of a map to guide management activity or be used to conduct economic analyses to inform species conversion decisions The photogrammetric survey of baleen whale morphometrics revealed the potential to conduct precise morphometric analysis in a repeatable fashion on two different baleen whale species with a low-cost commercial sUAS. This information can be used to assess population dynamics over time and in response to environmental phenomenon. The standardized method presented in the form of guidelines and analytical software offers guidance to the rapidly growing field of baleen whale photogrammetrists. The standardization method benefits the broader community by providing a free analytical tool that improves efficiency and perhaps more importantly, increase the transferability of data between studies.

In addition to the successes of the sUAS remote sensing methods employed in the the Swiss needle cast and baleen whale case studies, these studies revealed a several challenges related to environmental remote sensing with sUAS. Regulatory limitations that limit flights to a 120 m AGL ceiling and horizontal distances between pilot and aircraft to visual line of sight restrict efficient utilization. Furthermore, the limited endurance of multirotor sUASs (*e.g.*, < 30 minutes) further impedes efficiency such that surveys are limited in scope in terms of area covered or number of whales imaged in a single flight. However, within the United States regulations are relaxing to allow limited waivers to some of the regulatory restrictions which will vastly improve efficiency when surveys employ high endurance (*e.g.*, fixed-wing) UASs. Equipment reliability, pilot and analyst expertise will likely continue to plague operations but many will view the trade-offs compared to manned flights as acceptable because sUAS operations have a near-zero chance of endangering human lives.

Small UAS technology will continue to develop and become more accessible to scientists conducting environmental remote sensing. The rapidly increasing sophistication of low-cost consumer systems will drive new innovations in environmental remote sensing. Future studies will benefit from beyond line of sight UAS operations and long-endurance aircraft to test the efficacy of the methods presented here on a scale that truly compares with what is currently possible with manned aircraft.

Bibliography

Agisoft (2013). Agisoft PhotoScan User Manual Professional Edition. Ver. 1.0, Agisoft LLC, St. Petersburg, Russia.

Albetis, J., Duthoit, S., Guttler, F., Jacquin, A., Goulard, M., Poilvé, H., Féret, J.-B., and Dedieu, G. (2017). Detection of Flavescence dorée Grapevine Disease Using Unmanned Aerial Vehicle (UAV) Multispectral Imagery. *Remote Sensing*, 9(4):308.

Apostol, B., Lorent, A., Petrila, M., Gancz, V., and Badea, O. (2016). Height Extraction and Stand Volume Estimation Based on Fusion Airborne LiDAR Data and Terrestrial Measurements for a Norway Spruce. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 44(1):313.

Barber, C. B., Dobkin, D. P., and Huhdanpaa, H. (1996). The Quickhull Algorithm for Convex Hulls. *ACM Trans. Math. Softw.*, 22(4):469–483.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *arXiv:1406.5823 [stat]*.

Bedell, E., Leslie, M., Fankhauser, K., Burnett, J., Wing, M. G., and Thomas, E. A. (2017). Unmanned aerial vehicle-based structure from motion biomass inventory estimates. *Journal of Applied Remote Sensing*, 11(2):026026.

Best, P. B. and Rüther, H. (1992). Aerial photogrammetry of southern right whales, Eubalaena australis. *Journal of Zoology*, 228(4):595–614.

Birant, D. and Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering*, 60(1):208–221.

Black, B. A., Shaw, D. C., and Stone, J. K. (2010). Impacts of Swiss needle cast on overstory Douglas-fir forests of the western Oregon Coast Range. *Forest Ecology and Management*, 259(8):1673–1680.

Brandt, J. P., Morgan, T. A., Dillon, T., Lettman, G. J., Keegan, C. E., and Azuma, D. L. (2006). Oregon's forest products industry and timber harvest, 2003. Gen. Tech. Rep. PNW-GTR-681, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Brubaker, L. B. and Greene, S. K. (1979). Differential effects of Douglas-fir tussock moth and western spruce budworm defoliation on radial growth of grand fir and Douglas-fir. *Canadian Journal of Forest Research*, 9(1):95–105.

Burnett, J. D., Shaw, D., Smith, B., and Wing, M. G. (2017). Individual Tree Disease Detection using a sUAS and a Consumer-grade Camera: A Case Study on Swiss Needle Cast Disease in Douglas-fir. *Forest Science - In Review*.

Calderón, R., Navas-Cortés, J. A., Lucena, C., and Zarco-Tejada, P. J. (2013). High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. *Remote Sensing of Environment*, 139:231–245.

Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., and Gattelli, M. (2015). Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. *Remote Sensing*, 7(4):4026–4047.

Carbonneau, P. E. and Dietrich, J. T. (2017). Cost-effective non-metric photogrammetry from consumer-grade sUAS: Implications for direct georeferencing of structure from motion photogrammetry. *Earth Surface Processes and Landforms*, 42(3):473–486.

Carlson, T. N. and Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62(3):241–252.

Chander, G., Markham, B. L., and Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113(5):893–903.

Christiansen, F., Dujon, A. M., Sprogis, K. R., Arnould, J. P., and Bejder, L. (2016). Noninvasive unmanned aerial vehicle provides estimates of the energetic cost of reproduction in humpback whales. *Ecosphere*, 7(10).

Clark, J. B. and Lister, G. R. (1975). Photosynthetic Action Spectra of Trees: I. Comparative Photosynthetic Action Spectra of One Deciduous and Four Coniferous Tree Species as Related to Photorespiration and Pigment Complements. *Plant Physiology*, 55(2):401–406.

Clutton-Brock, T. and Sheldon, B. C. (2010). Individuals and populations: The role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, 25(10):562–573.

Comer, R., Kinn, G., Light, D., and Mondello, C. (1998). Talking digital. *Photogrammetric Engineering and Remote Sensing*, 64(12):1139–1142.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1):35–46.

Cubbage, J. C. and Calambokidis, J. (1987). Size-class segregation of bowhead whales discerned through aerial stereophotogrammetry. *Marine Mammal Science*, 3(2):179–185.

Dash, J. P., Watt, M. S., Pearse, G. D., Heaphy, M., and Dungey, H. S. (2017). Assessing very high resolution UAV imagery for monitoring forest health during a simulated disease outbreak. *ISPRS Journal of Photogrammetry and Remote Sensing*, 131:1–14.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*, volume 1. Cambridge university press.

Dawson, S. M., Chessum, C. J., Hunt, P. J., and Slooten, E. (1995). An inexpensive, stereophotographic technique to measure sperm whales from small boats. *Report of the International Whaling Commission*, 45(43):1–436.

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., Partchev, I., and others (2011). The estimation of item response models with the lmer function from the lme4 package in r. *Journal of Statistical Software*, 39(12):1–28.

Dolan, R., Hayden, B., and Heywood, J. (1978). A new photogrammetric method for determining shoreline erosion. *Coastal Engineering*, 2:21–39.

Durban, J. W., Fearnbach, H., Barrett-Lennard, L. G., Perryman, W. L., and Leroi, D. J. (2015). Photogrammetry of killer whales using a small hexacopter launched at sea. *Journal of Unmanned Vehicle Systems*, 3(3):131–135.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, pages 54–75.

Eitel, J. U. H., Vierling, L. A., Litvak, M. E., Long, D. S., Schulthess, U., Ager, A. A., Krofcheck, D. J., and Stoscheck, L. (2011). Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. *Remote Sensing of Environment*, 115(12):3640–3646.

ESRI (2016). ArcMap. Arcgis desktop ver. 10.3.1, Environmental Systems Research Institute, Redlands, CA.

Everitt, J. H., Escobar, D. E., Appel, D. N., Riggs, W. G., and Davis, M. R. (1999). Using Airborne Digital Imagery for Detecting Oak Wilt Disease. *Plant Disease*, 83(6):502–505.

FAA (2016). CFR 14 Part 107—Small Unmanned Aircraft Systems. Technical report, Federal Aviation Administration.

Fearnbach, H., Durban, J. W., Ellifrit, D. K., and Balcomb III, K. C. (2011). Size and long-term growth trends of endangered fish-eating killer whales. *Endangered Species Research*, 13(3):173–180.

Feng, Q., Liu, J., and Gong, J. (2015). UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. *Remote Sensing*, 7(1):1074–1094.

Finley, K. J. and Darling, L. M. (1990). Historical data sources on the morphometry and oil yield of the bowhead whale. *Arctic*, pages 153–156.

Flegal, K. M., Carroll, M. D., Kit, B. K., and Ogden, C. L. (2012). Prevalence of obesity and trends in the distribution of body mass index among us adults, 1999-2010. *Jama*, 307(5):491–497.

Fonstad, M. A., Dietrich, J. T., Courville, B. C., Jensen, J. L., and Carbonneau, P. E. (2013). Topographic structure from motion: A new development in photogrammetric measurement. *Earth Surface Processes and Landforms*, 38(4):421–430.

Forrester, D. J., Odell, D. K., Thompson, N. P., and White, J. R. (1980). Morphometrics, parasites, and chlorinated hydrocarbon residues of pygmy killer whales from florida. *Journal of Mammalogy*, 61(2):356–360.

Franke, J. and Menz, G. (2007). Multi-temporal wheat disease detection by multi-spectral remote sensing. *Precision Agriculture*, 8(3):161–172.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378.

Gallagher, D., Visser, M., Sepulveda, D., Pierson, R. N., Harris, T., and Heymsfield, S. B. (1996). How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups. *American journal of epidemiology*, 143(3):228–239.

Gamon, J. A., Field, C. B., Goulden, M. L., Griffin, K. L., Hartley, A. E., Joel, G., Penuelas, J., and Valentini, R. (1995). Relationships Between NDVI, Canopy Structure, and Photosynthesis in Three Californian Vegetation Types. *Ecological Applications*, 5(1):28–41.

Gao, B.-C., Montes, M. J., Davis, C. O., and Goetz, A. F. H. (2009). Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sensing of Environment*, 113:S17–S24.

Gao, X., Huete, A. R., Ni, W., and Miura, T. (2000). Optical–Biophysical Relationships of Vegetation Spectra without Background Contamination. *Remote Sensing of Environment*, 74(3):609–620.

Garcia-Ruiz, F., Sankaran, S., Maja, J. M., Lee, W. S., Rasmussen, J., and Ehsani, R. (2013). Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. *Computers and Electronics in Agriculture*, 91:106–115.

Gasparini, F. and Schettini, R. (2004). Color balancing of digital photos using simple image statistics. *Pattern Recognition*, 37(6):1201–1217.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., and Tyler, D. (2002). The National Elevation Dataset. *Photogrammetric Engineering & Remote Sensing*, 68(1):5–32.

Ghilani, C. D. (2011). *Adjustment Computations: Spatial Data Analysis*. John Wiley & Sons.

Gilbertson, J. K., Kemp, J., and van Niekerk, A. (2017). Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques. *Computers and Electronics in Agriculture*, 134:151–159.

Gillins, M. N., Gillins, D., and Parrish, C. (2016). Cost-Effective Bridge Safety Inspections Using Unmanned Aircraft Systems (UAS). *Geotechnical and Structural Engineering Congress 2016*.

Girden, E. R. (1992). *Anova: Repeated Measures*. Number 84. Sage.

Gitelson, A. A. (2004). Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology*, 161(2):165–173.

González-Jorge, H., Martínez-Sánchez, J., Bueno, M., and Arias, P. (2017). Unmanned Aerial Systems for Civil Applications: A Review. *Drones*, 1(1):2.

Goodchild, M. F. (1994). Integrating GIS and remote sensing for vegetation analysis and modeling: Methodological issues. *Journal of Vegetation Science*, 5(5):615–626.

Hahsler, M., Piekenbrock, M., Arya, S., and Mount, D. (2017). Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms. Ver. 1.1-1.

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*, 4(2):627–635.

Halachmi, I., Klopčič, M., Polak, P., Roberts, D. J., and Bewley, J. M. (2013). Automatic assessment of dairy cattle body condition score using thermal imaging. *Computers and Electronics in Agriculture*, 99:35–40.

Halachmi, I., Polak, P., Roberts, D. J., and Klopcic, M. (2008). Cow Body Shape and Automation of Condition Scoring. *Journal of Dairy Science*, 91(11):4444–4451.

Hansen, E. M., Stone, J. K., Capitano, B. R., Rosso, P., Sutton, W., Winton, L., Kanaskie, A., and McWilliams, M. G. (2000). Incidence and Impact of Swiss Needle Cast in Forest Plantations of Douglas-fir in Coastal Oregon. *Plant Disease*, 84(7):773–778.

Harrell, F. E. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer. Google-Books-ID: 94RgCgAAQBAJ.

Harrell, F. E., Lee, K. L., and Mark, D. B. (1996). Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Statistics in Medicine*, 15(4):361–387.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Wiley Online Library.

Hayhurst, K. J., Maddalon, J. M., Neogi, N. A., and Vertstynen, H. A. (2016). Safety and Certification Considerations for Expanding the Use of UAS in Precision Agriculture. 31 Jul. - 3 Aug. 2016, United States.

Hilker, T., Wulder, M. A., Coops, N. C., Linke, J., McDermid, G., Masek, J. G., Gao, F., and White, J. C. (2009). A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, 113(8):1613–1627.

Holman, F. H., Riche, A. B., Michalski, A., Castle, M., Wooster, M. J., and Hawkesford, M. J. (2016). High Throughput Field Phenotyping of Wheat Plant Height and Growth Rate in Field Plot Trials Using UAV Based Remote Sensing. *Remote Sensing*, 8(12):1031.

Horler, D. N. H., Dockray, M., and Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2):273–288.

Huemmrich, K. F. and Goward, S. N. (1997). Vegetation canopy PAR absorptance and NDVI: An assessment for ten tree species with the SAIL model. *Remote Sensing of Environment*, 61(2):254–269.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1):195–213.

Hugenholtz, C., Brown, O., Walker, J., Barchyn, T., Nesbit, P., Kucharczyk, M., and Myshak, S. (2016). Spatial Accuracy of UAV-Derived Orthoimagery and Topography: Comparing Photogrammetric Models Processed with Direct Geo-Referencing and Ground Control Points. *GEOMATICA*, 70(1):21–30.

Hunt, E. R., Daughtry, C. S. T., Eitel, J. U. H., and Long, D. S. (2011). Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index. *Agronomy Journal*, 103(4):1090–1099.

Jaquet, N. (2006). A simple photogrammetric technique to measure sperm whales at sea. *Marine Mammal Science*, 22(4):862–879.

Johnson, E. W. and Wittwer, D. (2008). Aerial detection surveys in the United States. *Australian Forestry*, 71(3):212–215.

Kanaskie, A., Sprengel, K., and Overhulser, D. (2007). Swiss Needle Cast Aerial Surveys, 1996 – 2007. 2007 Annual Report, Swiss Needle Cast Cooperative, Corvallis, OR.

Kennedy, R. E., Yang, Z., and Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. *Remote Sensing of Environment*, 114(12):2897–2910.

Kimberley, M. O., Hood, I. A., and Knowles, R. L. (2010). Impact of Swiss Needle-Cast on Growth of Douglas-Fir. *Phytopathology*, 101(5):583–593.

Klimley, A. P. and Brown, S. T. (1983). Stereophotography for the field biologist: Measurement of lengths and three-dimensional positions of free-swimming sharks. *Marine Biology*, 74(2):175–185.

Knoll, T., Knoll, J., and Others (2015). Adobe Photoshop CC 2015. Ver. 2015.5, Adobe Systems Inc.

Knowles, J. E. and Frederick, C. (2016). Prediction intervals from merMod objects.

Knowles, J. E., Frederick, C., and Knowles, M. J. E. (2016). merTools: Tools for analyzing mixed effect regression models.

Koh, L. P. and Wich, S. A. (2012). Dawn of Drone Ecology: Low-Cost Autonomous Aerial Vehicles for Conservation. *Tropical Conservation Science*, 5(2):121–132.

Konecny, G. (1985). The international society for photogrammetry and remote sensing-75 years old, or 75 years young. *Photogrammetric Engineering and Remote Sensing*, 51(7):919–933.

Kuhn, M. (2016). The caret Package. Ver. 6.0-76.

Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*. Springer Science & Business Media.

Laliberte, A. S. and Rango, A. (2009). Texture and Scale in Object-Based Analysis of Subdecimeter Resolution Unmanned Aerial Vehicle (UAV) Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(3):761–770.

Landis, J. R. and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.

Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., and Bretagnolle, V. (2014). Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. *Global Ecology and Biogeography*, 23(7):811–820.

Liaw, A., Wiener, M., and others (2002). Classification and regression by randomForest. *R news*, 2(3):18–22.

Lillesand, T., Kiefer, R. W., and Chipman, J. (2014). *Remote Sensing and Image Interpretation*. John Wiley & Sons.

Lockyer, C. (1986). Body fat condition in northeast atlantic fin whales, balaenoptera physalus, and its relationship with reproduction and food resource. *Canadian Journal of Fisheries and Aquatic Sciences*, 43(1):142–147.

Lockyer, C. (2007). All creatures great and smaller: A study in cetacean life history energetics. *Journal of the Marine Biological Association of the United Kingdom*, 87(04):1035–1045.

Lopatin, J., Dolos, K., Hernández, H. J., Galleguillos, M., and Fassnacht, F. E. (2016). Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. *Remote Sensing of Environment*, 173:200–210.

MacLean, D. A. and MacKinnon, W. E. (1996). Accuracy of aerial sketch-mapping estimates of spruce budworm defoliation in New Brunswick. *Canadian Journal of Forest Research*, 26(12):2099–2108.

Maguire, D. A., Kanaskie, A., Voelker, W., Johnson, R., and Johnson, G. (2002). Growth of young Douglas-fir plantations across a gradient in Swiss needle cast severity. *Western Journal of Applied Forestry*, 17(2):86–95.

Manter, D. K., Bond, B. J., Kavanagh, K. L., Rosso, P. H., and Filip, G. M. (2000). Pseudothecia of Swiss needle cast fungus, Phaeocryptopus gaeumannii, physically block stomata of Douglas fir, reducing CO2 assimilation. *New Phytologist*, 148(3):481–491.

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L. R., Davis, C. E., and Dandekar, A. M. (2015). Advanced methods of plant disease detection. A review. *Agronomy for Sustainable Development*, 35(1):1–25.

Mathworks (2016). MATLAB. Ver. r2016b, The MathWorks Inc., Natick, MA.

McGaughey, R. J. (2017). FUSION/LDV: Software for LIDAR Data Analysis and Visualization. Ver. 3.6, United States Forest Service Pacific Northwest Research Station, Seattle, WA.

McGlone, J. C. (2013). *Manual of Photogrammetry*. American Soc. for Photogrammetry and Remote Sensing.

Meentemeyer, R., Rizzo, D., Mark, W., and Lotz, E. (2004). Mapping the risk of establishment and spread of sudden oak death in California. *Forest Ecology and Management*, 200(1):195–214.

Millar, C. I., Stephenson, N. L., and Stephens, S. L. (2007). Climate Change and Forests of the Future: Managing in the Face of Uncertainty. *Ecological Applications*, 17(8):2145–2151.

Misimi, E., Erikson, U., and Skavhaug, A. (2008). Quality Grading of Atlantic Salmon (Salmo salar) by Computer Vision. *Journal of Food Science*, 73(5):E211–E217.

Moisen, G. G., Freeman, E. A., Blackard, J. A., Frescino, T. S., Zimmermann, N. E., and Edwards, T. C. (2006). Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling*, 199(2):176–187.

Moreland, E. E., Cameron, M. F., Angliss, R. P., and Boveng, P. L. (2015). Evaluation of a ship-based unoccupied aircraft system (UAS) for surveys of spotted and ribbon seals in the Bering Sea pack ice. *Journal of Unmanned Vehicle Systems Virtual Issue*, 01(01):114–122.

Mukaka, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal*, 24(3):69–71.

Mulvey, R. L., Shaw, D. C., Filip, G. M., and Chastagner, G. A. (2013). Swiss Needle Cast. Technical Report FS/R6/RO/FIDL#181-13/001, USDA Forest Service Forest, Portland, OR.

Nelson, R. F. (1983). Detecting forest canopy change due to insect activity using Landsat MSS. *Photogrammetric Engineering and Remote Sensing*, 49(9):1303–1314.

Norris, K. S. (1961). Standardized methods for measuring and recording data on the smaller cetaceans. *Journal of Mammalogy*, pages 471–476.

Oborne, M. (2015). Mission Planner. Ver. 1.3.20.

OFRI (2012). The 2012 Forest Report. Technical report, Oregon Forest Resources Institute, Portland, OR.

Olsen, J. L. (2002). Oregon State University's Integrated Pest Management Program for the Oregon Hazelnut Industry. *HortTechnology*, 12(4):623–625.

Otten, M. R. M., Haufler, J. B., Winterstein, S. R., and Bender, L. C. (1993). An Aerial Censusing Procedure for Elk in Michigan. *Wildlife Society Bulletin (1973-2006)*, 21(1):73–80.

Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., and Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, 56(1):45–50.

Pasumansky, A. (2017). Photoscan Professional. Ver. 1.3.1.4, Agisoft LLC.

Patrick, A., Pelham, S., Culbreath, A., Holbrook, C. C., Godoy, I. J. D., and Li, C. (2017). High throughput phenotyping of tomato spot wilt disease in peanuts using unmanned aerial systems and multispectral imaging. *IEEE Instrumentation Measurement Magazine*, 20(3):4–12.

Perryman, W. L. and Lynn, M. S. (1993). Identification of geographic forms of common dolphin (delphinus delphis) from aerial photogrammetry. *Marine Mammal Science*, 9(2):119–137.

Perryman, W. L. and Lynn, M. S. (2002). Evaluation of nutritive condition and reproductive status of migrating gray whales (schrichtius robustus) based on analysis of photogrammetric data. *Journal of Cetacean Research and Management*, 4(2):155–164.

Philpot, W. D. and Philipson, W. R. (2012). *Remote Sensing Fundamentals*. Chapter.

Pixel, L. (2011). Digital Infrared Photography Primer. Infrared conversions, ir modifications & photography tutorials | life pixel ir, Life Pixel LLC.

Poland, T. M. and McCullough, D. G. (2006). Emerald Ash Borer: Invasion of the Urban Forest and the Threat to North America's Ash Resource. *Journal of Forestry*, 104(3):118–124.

Pomfret, K. D. (2016). Sector Insights. *Photogrammetric Engineering & Remote Sensing*, 82(9):657–658.

Popescu, S. C. and Wynne, R. H. (2004). Seeing the trees in the forest. *Photogrammetric Engineering & Remote Sensing*, 70(5):589–604.

Prospere, K., McLaren, K., and Wilson, B. (2014). Plant Species Discrimination in a Tropical Wetland Using In Situ Hyperspectral Data. *Remote Sensing*, 6(9):8494–8523.

R Core Team (2017). R: A language and environment for statistical computing. Ver. 3.3.3, R Foundation for Statistical Computing, Vienna Austria.

Rasmussen, J., Ntakos, G., Nielsen, J., Svensgaard, J., Poulsen, R. N., and Christensen, S. (2016). Are vegetation indices derived from consumer-grade cameras mounted on UAVs sufficiently reliable for assessing experimental plots? *European Journal of Agronomy*, 74:75–92.

Rice, D. W. and Wolman, A. A. (1971). *The Life History and Ecology of the Gray Whale (Eschrichtius Robustus)*. Number 3. American society of mammalogists.

Ritóková, G., Shaw, D., Maguire, D., Mainwaring, D., Browning, J., Gourley, M., Filip, G., and Kanaskie, A. (2014). SNCC Research and Monitoring Plot Network. Annual Report 2014, Swiss Needle Cast Cooperative, Corvallis, OR.

Ritóková, G., Shaw, D. C., Filip, G., Kanaskie, A., Browning, J., and Norlander, D. (2016). Swiss Needle Cast in Western Oregon Douglas-Fir Plantations: 20-Year Monitoring Results. *Forests*, 7(8):155.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, pages n/a–n/a.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67:93–104.

Rohner, C. A., Richardson, A. J., Marshall, A. D., Weeks, S. J., and Pierce, S. J. (2011). How large is the world's largest fish? Measuring whale sharks Rhincodon typus with laser photogrammetry. *Journal of Fish Biology*, 78(1):378–385.

Roujean, J.-L., Leroy, M., and Deschamps, P.-Y. (1992). A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *Journal of Geophysical Research: Atmospheres*, 97(D18):20455–20468.

Rouse, J. W., Schell, J. A., Deering, D. W., and Harlan, J. C. (1974). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Technical Report NASA/GSFC Type III Final Report, NASA, Greenbelt, Maryland.

Sankaran, S., Mishra, A., Ehsani, R., and Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, 72(1):1–13.

Shaw, D. C., Filip, G. M., Kanaskie, A., Maguire, D. A., and Littke, W. A. (2011). Managing an epidemic of swiss needle cast in the Douglas-fir region of Oregon: The role of the Swiss needle cast cooperative. *Journal of Forestry*, 109(2):109–119.

Shaw, D. C., Woolley, T., and Kanaskie, A. (2014). Vertical Foliage Retention in Douglas-Fir Across Environmental Gradients of the Western Oregon Coast Range Influenced by Swiss Needle Cast. *Northwest Science*, 88(1):23–32.

Shervals, K. and Dietrich, J. T. (2016). Structure from Motion (SfM) Agisoft PhotoScan processing guide. Webpage, UNAVCO.

Sim, J. and Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3):257–268.

Smith, G. M. and Milton, E. J. (1999). The use of the empirical line method to calibrate remotely sensed data to reflectance. *International Journal of Remote Sensing*, 20(13):2653–2662.

Stone, J. K. (1987). Initiation and development of latent infections by Rhabdocline parkeri on Douglas-fir. *Canadian Journal of Botany*, 65(12):2614–2621.

Stower, W. J., Davies, D. E., and Jones, I. B. (1960). Morphometric studies of the desert locust, schistocerca gregaria (forsk.). *The Journal of Animal Ecology*, pages 309–339.

Strîmbu, V. F. and Strîmbu, B. M. (2015). A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104:30–43.

Sturrock, R. N., Frankel, S. J., Brown, A. V., Hennon, P. E., Kliejunas, J. T., Lewis, K. J., Worrall, J. J., and Woods, A. J. (2011). Climate change and forest diseases. *Plant Pathology*, 60(1):133–149.

Tang, L. and Shao, G. (2015). Drone remote sensing for forestry research and practices. *Journal of Forestry Research*, 26(4):791–797.

Tong, E. N. C., Mues, C., and Thomas, L. C. (2012). Mixture cure models in credit scoring: If and when borrowers default. *European Journal of Operational Research*, 218(1):132–139.

Torres, L. G. (2013). Evidence for an unrecognised blue whale foraging ground in new zealand. *New Zealand Journal of Marine and Freshwater Research*, 47(2):235–248.

Torres, L. G., Barlow, D. R., Hodge, K., Klinck, H., Steel, D., Baker, C. S., Chandler, T., Gill, P., Ogle, M., Lilley, C., Bury, S., Graham, B., Sutton, P., Burnett, J., Double, M., Olson, P., Bott, N., and Constantine, R. (2017). New Zealand blue whales: Recent findings and research progress. Recent findings and research progress. SC/67A/SH/02, International Whaling Commission.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127–150.

US Army Corps of Engineers (2015). Photogrammetric and lidar mapping. Technical Report Engineer Manual 1110-1-1000, US Army Corps of Engineers.

Vanwinckelen, G. and Blockeel, H. (2015). Look before you leap: Some insights into learner evaluation with cross-validation. pages 3–20.

Venables, W. N. and Ripley, B. D. (2013). *Modern Applied Statistics with S-PLUS*. Springer Science & Business Media.

Verhoeven, G. (2011). Taking computer vision aloft – archaeological three-dimensional reconstructions from aerial photographs with photoscan. *Archaeological Prospection*, 18(1):67–73.

Verhoeven, G. J. (2012). Near-Infrared Aerial Crop Mark Archaeology: From its Historical Use to Current Digital Implementations. *Journal of Archaeological Method and Theory*, 19(1):132–160.

Verhoeven, G. J., Smet, P. F., Poelman, D., and Vermeulen, F. (2009). Spectral Characterization of a Digital Still Camera's NIR Modification to Enhance Archaeological Observation. *IEEE Transactions on Geoscience and Remote Sensing*, 47(10):3456–3468.

Verhoeven, G. J. J. (2010). It's all about the format – unleashing the power of RAW aerial photography. *International Journal of Remote Sensing*, 31(8):2009–2042.

Viera, A. J., Garrett, J. M., and others (2005). Understanding interobserver agreement: The kappa statistic. *Fam Med*, 37(5):360–363.

Wallace, L., Lucieer, A., Malenovský, Z., Turner, D., and Vopěnka, P. (2016). Assessment of Forest Structure Using Two UAV Techniques: A Comparison of Airborne Laser Scanning and Structure from Motion (SfM) Point Clouds. *Forests*, 7(3):62.

Watts, A. C., Ambrosia, V. G., and Hinkley, E. A. (2012). Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use. *Remote Sensing*, 4(6):1671–1692.

Weinfurt, K. P. (2000). Repeated measures analysis: Anova, manova, and hlm. In *Reading and Understanding More Multivariate Statistics*. American Psychological Association, Washington, DC.

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., and Reynolds, J. M. (2012). 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179:300–314.

Wilhelmi, N. (2016). *The Effects of Seed Source and Planting Environment on Douglas-Fir (Pseudotsuga Menziesii) Foliage Diseases*. Thesis, Oregon State University, Corvallis, OR.

Wing, M., Burnett, J., Johnson, S., Akay, A., and Sessions, J. (2014). A Low-cost Unmanned Aerial System for Remote Sensing of Forested Landscapes. *International Journal of Remote Sensing Applications*, 4(3):113.

Wing, M. G., Burnett, J., Sessions, J., Brungardt, J., Cordell, V., Dobler, D., and Wilson, D. (2013). Eyes in the sky: Remote sensing technology development using small unmanned aircraft systems. *Journal of Forestry*, 111(5):341–347.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*. CRC press.

Yoder, B. J. and Waring, R. H. (1994). The normalized difference vegetation index of small Douglas-fir canopies with varying chlorophyll concentrations. *Remote Sensing of Environment*, 49(1):81–91.

Zarco-Tejada, P. J., González-Dugo, V., and Berni, J. A. J. (2012). Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. *Remote Sensing of Environment*, 117:322–337.

Zhao, J., Maguire, D. A., Mainwaring, D. B., and Kanaskie, A. (2015). The effect of within-stand variation in Swiss needle cast intensity on Douglas-fir stand dynamics. *Forest Ecology and Management*, 347:75–82.

APPENDIX

A Supporting Tables and Figures

Figure A.1: Matrix sUAS in-flight with Sony NEX 5T camera.

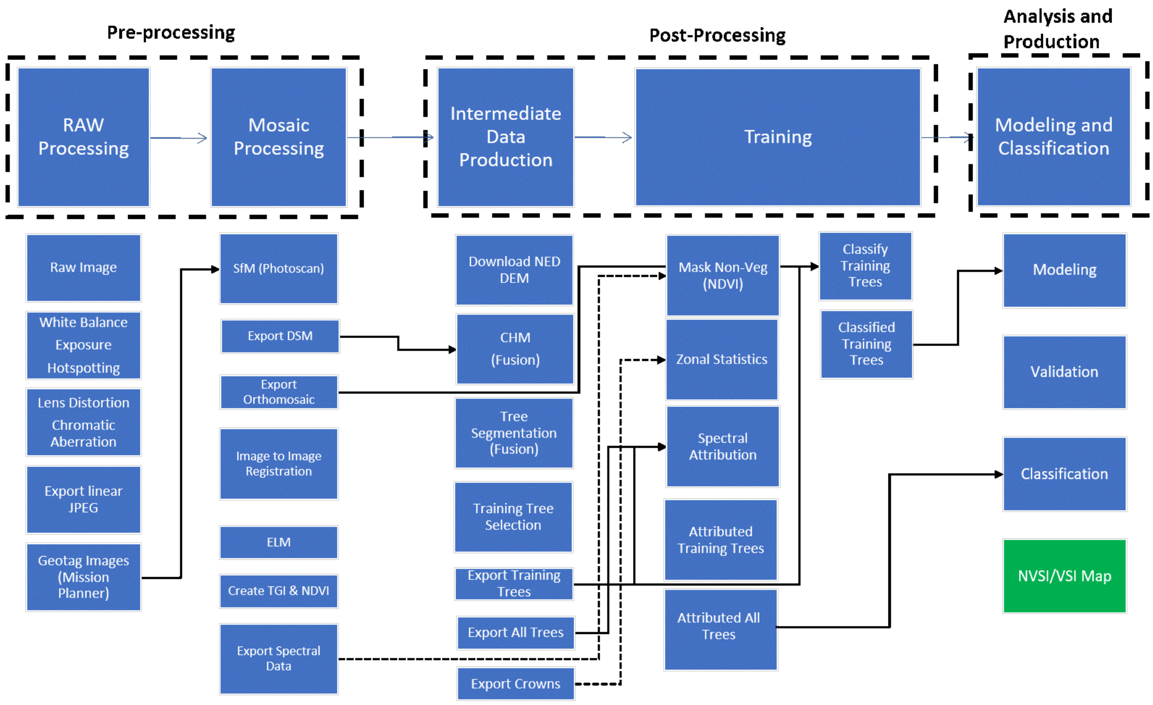


Figure A.2: Processing workflow diagram depicting three general data processing stages (i.e., pre-processing, post-processing, and analysis and production), steps and pertinent sub-routines. Arrows depict functional relationships between outputs and inputs. The three stages are broken down into steps that contain subroutines represented by the smaller rectangles within a column. The general flow is down the column and left to right.

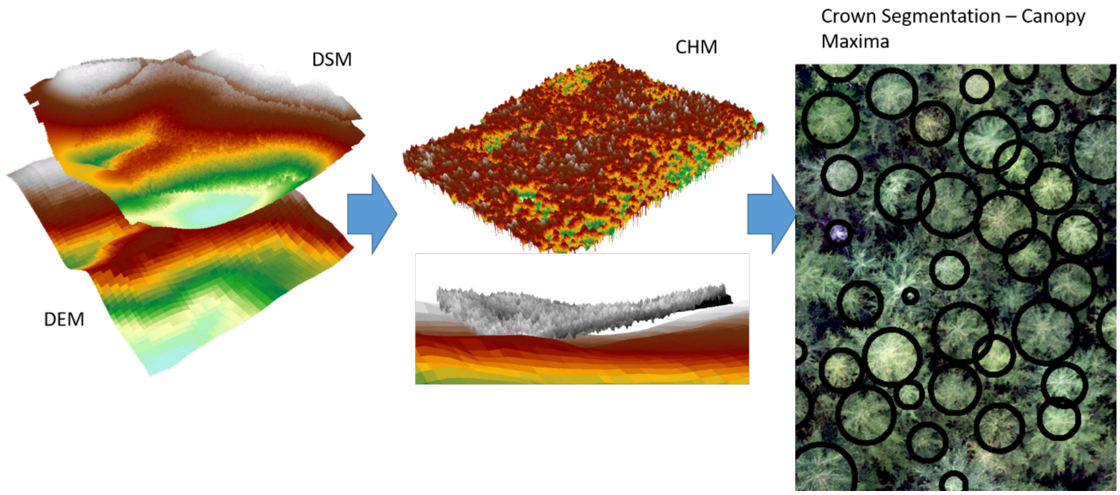


Figure A.3: Tree and crown segmentation workflow diagram depicts the differences between DSM and DEM data that allow relative-height CHM creation and subsequent crown segmentation with the CanopyMaxima function implemented in USFS FUSION. Site N02 data was used for this example. Note that the majority of the trees were segmented and crown areas are reasonable for most trees although some crown areas are vastly under or over estimated.

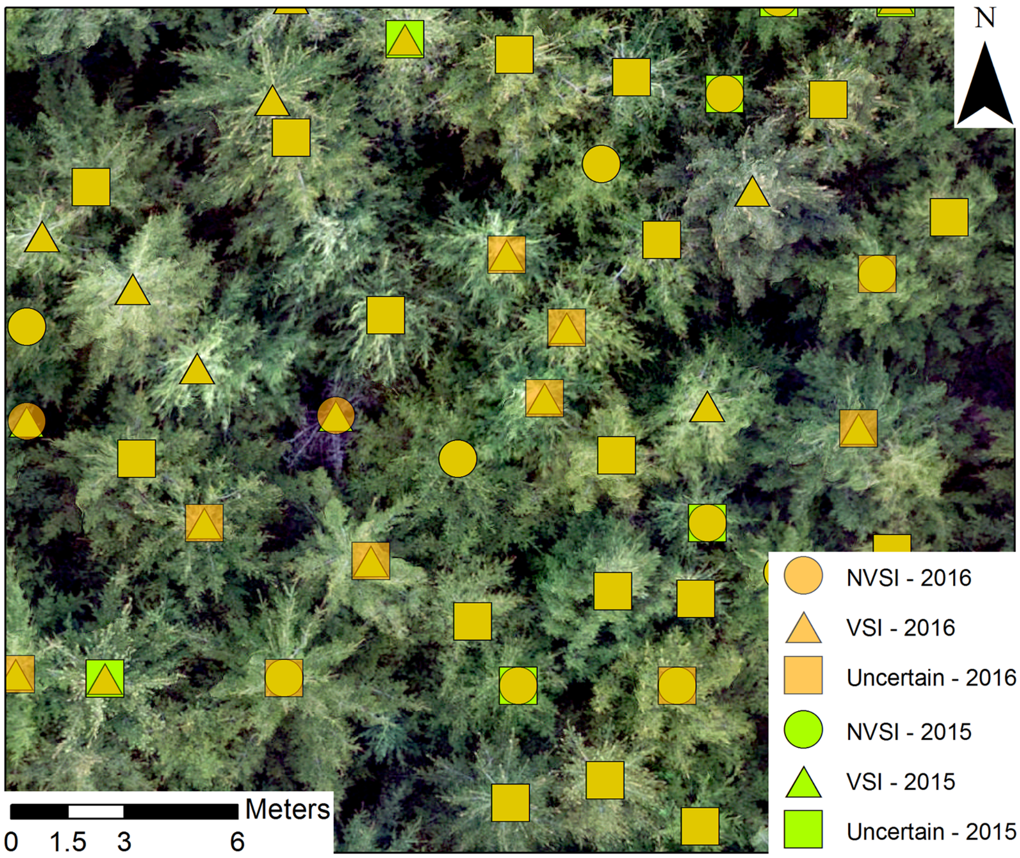


Figure A.4: Visualization of the detected status changes of individual trees at site S1W between 2015 and 2016.'Uncertain' is in reference to trees whose change status is uncertain due to the tuning of the classification probability algorithm. Trees with a high probability of exhibiting no visible sign of SNC infection (NVSI) are denoted with a circle and trees with high probability of exhibiting visible signs of SNC infection (VSI) are denoted as triangles. Colors refer to the years trees were surveyed. Overlapping symbols indicate a status change between years. Potentially undesirable status changes were indicated by overlapping circles and triangles. Squares indicate an uncertain status during a survey year.

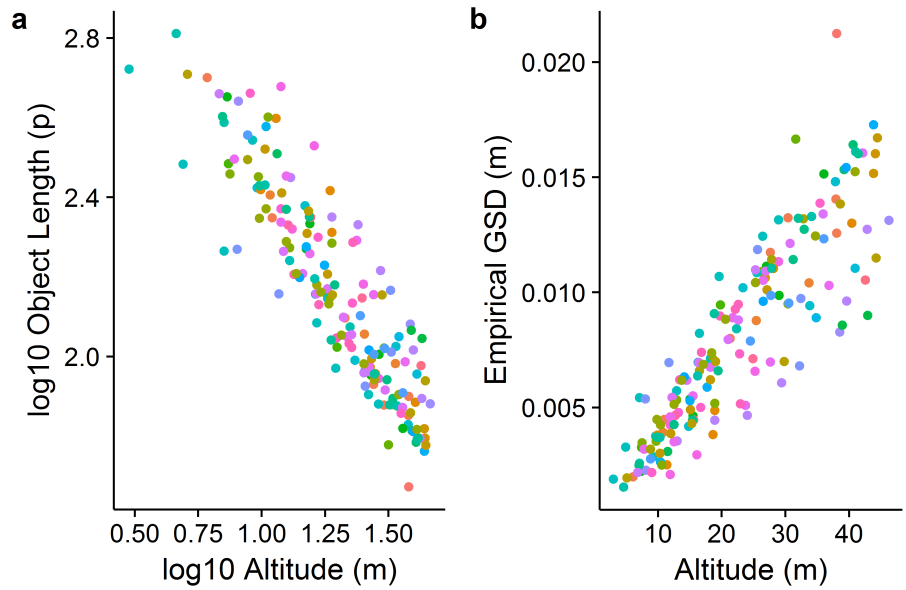


Figure A.5: The relationship between object length in pixels and altitude (a) depicting a wide variation in object length around a single altitude when altitude was not corrected for bias. The relationship between altitude and empirical ground sampling distance (eGSD) (b) shows the wide range in eGSDs observed from a single altitude when altitude bias is uncorrected.

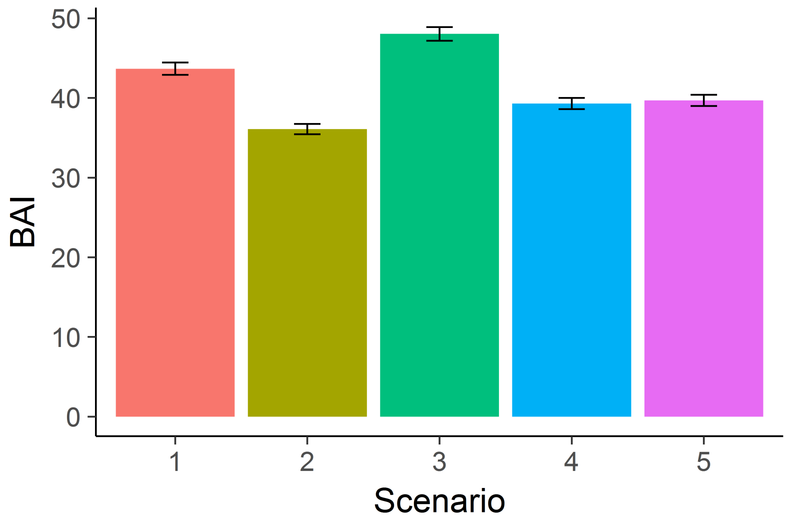


Figure A.6: Mean estimated body area index (BAI) for the entire population of gray whales ($n = 89$) in the context of each of the simulated whale length and surface area change scenarios specified in Table 4.1.

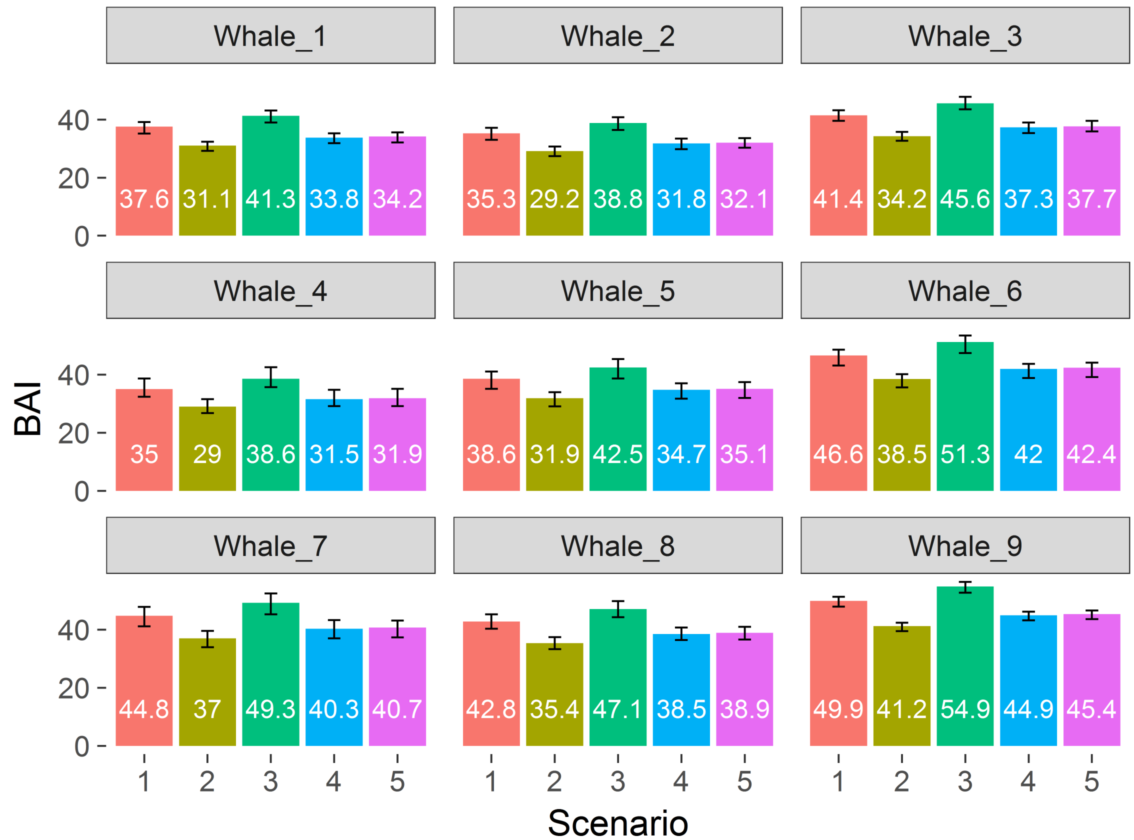


Figure A.7: Estimated body area index (BAI) for the five simulated change scenarios described in Table 4.1 for the nine gray whales examined in Figure 4.4a.

Table A.1: Mean estimated calibration object length and supporting metrics resulting from each of the scaling error correction methods described in Table 4.2. mRMSE is the mean of the root mean squared errors Eq 4.4 across all the flights contributing to the estimate of the mean, CV% is the coefficient of variation in units of percentage. Mean.lwr and Mean.upr designate the lower and upper bounds of the 95% confidence intervals on the mean.

Method	Mean (m)	Mean.lwr	Mean.upr	mRMSE	CV (%)	Bias (m)
M1	1.25	1.11	1.39	0.34	14.91	0.25
M2	1.1	0.69	1.65	0.26	16.21	$9.6 * 10^{-2}$
M3	1.04	0.63	1.53	0.17	19.88	$3.5 * 10^{-2}$
M4	1.06	0.93	1.21	0.25	16	$6.3 * 10^{-2}$
M5	1.02	0.79	1.28	0.17	17.56	$1.9 * 10^{-2}$

Table A.2: Results of the five correction methods listed in Table 4.2 for analysis of the nine whales identified in Figure 4.4. Whale Length (WL) mean is the mean estimated WL and WL coefficient of variation (CV) is the CV of the individual observations that contributed to the mean.

Whale	Method	WL Mean	WL CV	Whale	Method	WL Mean	WL CV
1	1	11.66	9.03	5	3	12.37	5.8
2	1	12.47	5.99	6	3	9.03	5.1
3	1	8.34	3.39	7	3	10.8	1.75
4	1	12.53	3.26	8	3	7.72	0.61
5	1	10.94	6.26	9	3	9.85	1.69
6	1	11.09	4.73	1	4	9.6	9.14
7	1	11.34	1.57	2	4	10.3	6
8	1	9.51	0.97	3	4	6.87	3.59
9	1	11.31	1.67	4	4	10.31	3.49
1	2	9.85	9.16	5	4	9	6.17
2	2	10.58	6.01	6	4	9.15	4.73
3	2	7.05	3.61	7	4	9.34	1.66
4	2	10.59	3.52	8	4	7.84	0.99
5	2	9.24	6.16	9	4	9.34	1.7
6	2	9.39	4.73	1	5	11.92	9.47
7	2	9.6	1.67	2	5	10.97	5.97
8	2	8.05	0.99	3	5	7.31	3.57
9	2	9.59	1.7	4	5	12.16	3.95
1	3	14.55	9.85	5	5	10.6	5.99
2	3	11.3	6.02	6	5	9.02	5
3	3	7.56	3.46	7	5	10.33	1.75
5	3	14.21	4.34	8	5	7.75	0.99
				9	5	9.68	1.67

B Index of Acronyms

Acronym	Explanation
ADS	Aerial Detection Survey
AGL	Above Ground Level
ASL	Above Sea Level
B	Blue
BAI	Body Area Index
BCI	Body Condition Index
cGSD	Corrected Ground Sampling Distance
CIs	Confidence Intervals
COA	Certificate of Authorization
CV	Coefficient of Variation
DEM	Digital Elevation Model
DF	Douglas-fir
DSM	Digital Surface Model
eGSD	Empirical Ground Sampling Distance
G	Green
GCPs	Ground Control Points
GPS	Global Positioning System
GSD	Ground Sampling Distance
JPEG	Joint Photographic Experts Group
LOS	Line of Sight
MA	Mixed Alder
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared
NMC	Narrowband Multispectral Camera
NVSI	No Visible Signs of Infection
OM	Orthomosaic
PC	Principal Component
PSRI	Plant Senescence Reflectance Index
R	Red
RGB	Mosaiced Red, Green, and Blue
RMSE	Root Mean Squared Error
RTK	Real-time Kinematic
SNC	Swiss needle cast
sUAS	Small Unmanned Aircraft System
TGI	Triangular Greenness Index
VI	Vegetation Index
VSI	Visible Signs of Infection