



Article

Influence of Structure from Motion Algorithm Parameters on Metrics for Individual Tree Detection Accuracy and Precision

Wade T. Tinkham ^{1,*} and George A. Woolsey ^{2,†}¹ USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO 80526, USA² Department of Forest and Rangeland Stewardship, Colorado State University, Fort Collins, CO 80523, USA

* Correspondence: wade.tinkham@usda.gov

† These authors contributed equally to this work.

Abstract: Uncrewed aerial system (UAS) structure from motion (SfM) monitoring strategies for individual trees has rapidly expanded in the early 21st century. It has become common for studies to report accuracies for individual tree heights and DBH, along with stand density metrics. This study evaluates individual tree detection and stand basal area accuracy and precision in five ponderosa pine sites against the range of SfM parameters in the Agisoft Metashape, Pix4DMapper, and OpenDroneMap algorithms. The study is designed to frame UAS-SfM individual tree monitoring accuracy in the context of data processing and storage demands as a function of SfM algorithm parameter levels. Results show that when SfM algorithms are properly tuned, differences between software types are negligible, with Metashape providing a median F-score improvement over OpenDroneMap of 0.02 and PIX4DMapper of 0.06. However, tree extraction performance varied greatly across algorithm parameters, with the greatest extraction rates typically coming from parameters causing increased density in dense point clouds and minimal point cloud filtering. Transferring UAS-SfM forest monitoring into management will require tradeoffs between accuracy and efficiency. Our analysis shows that a one-step reduction in dense point cloud quality saves 77–86% in point cloud processing time without decreasing tree extraction (F-score) or basal area precision using Metashape and PIX4DMapper but the same parameter change for OpenDroneMap caused a ~5% loss in precision. Providing reproducible processing strategies is a vital step in successfully transferring these technologies into usage as management tools.

Keywords: uncrewed aerial system; UAS; drone; Metashape; Pix4D; OpenDroneMap; ponderosa pine; F-score; forest monitoring



Citation: Tinkham, W.T.; Woolsey, G.A. Influence of Structure from Motion Algorithm Parameters on Metrics for Individual Tree Detection Accuracy and Precision. *Remote Sens.* **2024**, *16*, 3844. <https://doi.org/10.3390/rs16203844>

Academic Editor: Brenden E. McNeil

Received: 1 September 2024

Revised: 10 October 2024

Accepted: 14 October 2024

Published: 16 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Strategies for uncrewed aerial system (UAS) monitoring of forest structure, function, and condition have rapidly expanded over the last twenty years [1]. Practical methods have been developed that allow UAS to identify individual forest features as individual trees or ‘tree-approximate objects’ [2], characterize the height of vegetation [3], describe tree crown size and cover [4], and model forest’s aboveground biomass [5] and metrics of forest health [6,7]. While traditional field sampling with fixed area or variable radius plots can provide a sample for individual tree diameter at breast height (DBH; 1.37 m above ground) and total height that can be extrapolated to estimate metrics like average stand-level tree density and basal area (or the area occupied by all trees at DBH), these UAS forest monitoring strategies make substantial leaps forward by providing spatially continuous estimates of forest density [8] and spatial patterns [9]. While numerous strategies and sensors exist for characterizing forest attributes from UAS platforms, structure from motion (SfM) photogrammetry is among the most widely tested and affordable remote sensing approaches to describe three-dimensional forest attributes [10]. Common data collection standards for acquiring very-high-resolution UAS imagery for SfM processing

and point cloud generation are starting to be presented in the scientific literature. For SfM forest monitoring, research shows that maximizing forward image overlap while maintaining >80% side image overlap helps maximize canopy height model-based individual tree detection [11]. Generally, within the range of altitude above ground level allowed by regulating groups (e.g., US Federal Aviation Administration's 122 m maximum), flight altitude has had minimal impact on individual tree detection [12] and area-based aboveground biomass modeling [13]. However, the proportion and accuracy of individual tree DBH values extracted from point clouds experience small reductions in quantity and accuracy as altitude increases [12]. Comprehensive documentation on how data acquisition, processing, and modeling impact data fidelity is necessary for the reliable adoption of new technology by practitioners. While common standards for data acquisition are beginning to emerge, less effort has focused on parameterizing subsequent SfM photogrammetry processing algorithms.

Numerous SfM algorithms have been implemented in both proprietary and open-source software packages. While most of these software packages use similar overall workflows, there are small differences in their strategy that may make them better suited for modeling the built environment versus natural terrain and vegetation. While some comparisons of data processing efficiency have been made between software packages [14], most comparisons have not evaluated how data processing assumptions impact the end-user metrics that managers rely on. Additionally, while some evaluations have quantified how software parameterization impacts metrics like individual tree detection [15], this has not been comprehensively carried out across software packages. Such testing is necessary to optimize the data processing time and storage capacity needed for land management organizations to operationalize these technologies.

This study compares the accuracy and precision of individual tree detection against the range of SfM parameters available in two commercial and one open-source algorithm. Specifically, using five ponderosa pine (*Pinus ponderosa* Lawson & C. Lawson) dominated sites with 100% census field stem-mapped inventories, we evaluate how SfM algorithm parameters impact individual tree detection and resulting stand basal area estimates compared to field-based measurements. Additionally, we discuss these results in the context of data processing and storage demands at the different SfM algorithm parameter levels.

2. Materials and Methods

2.1. Study Ecosystem

Ponderosa pine-dominated forests occupy semi-arid lower montane regions of western North America [16]; Figure 1. These forest systems are characterized by low to moderate canopy cover (i.e., 20–60%) that is maintained by frequent, low-severity fires with return intervals varying between 7 and 40 years [17]. Forest structures within these systems historically exhibited varying degrees of spatial aggregation, often described as a matrix of individual trees, clumps of trees, and canopy openings [18,19]. Suppression of fire and historical forest management and grazing practices of the 1900s within ponderosa pine-dominated forests has resulted in homogenization of forest structures, increases in fuel loading, and the establishment of less fire-resistant tree species like Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. *glauca* (Beissn.) Franco), white fir (*Abies concolor* (Gord. & Glend.) Lindl. Ex Hildebr.), Rocky Mountain juniper (*Juniperus scopulorum* Sarg.), and Engelmann spruce (*Picea engelmannii* Parry ex Engelm.; [20,21]). Recent management of these forests has emphasized either fire hazard reduction through traditional space-based thinning from below to reduce fuel continuity [22] or treatments that balance fire hazard objectives with other ecological objectives through restoration of spatially aggregated forest structures [17,23].

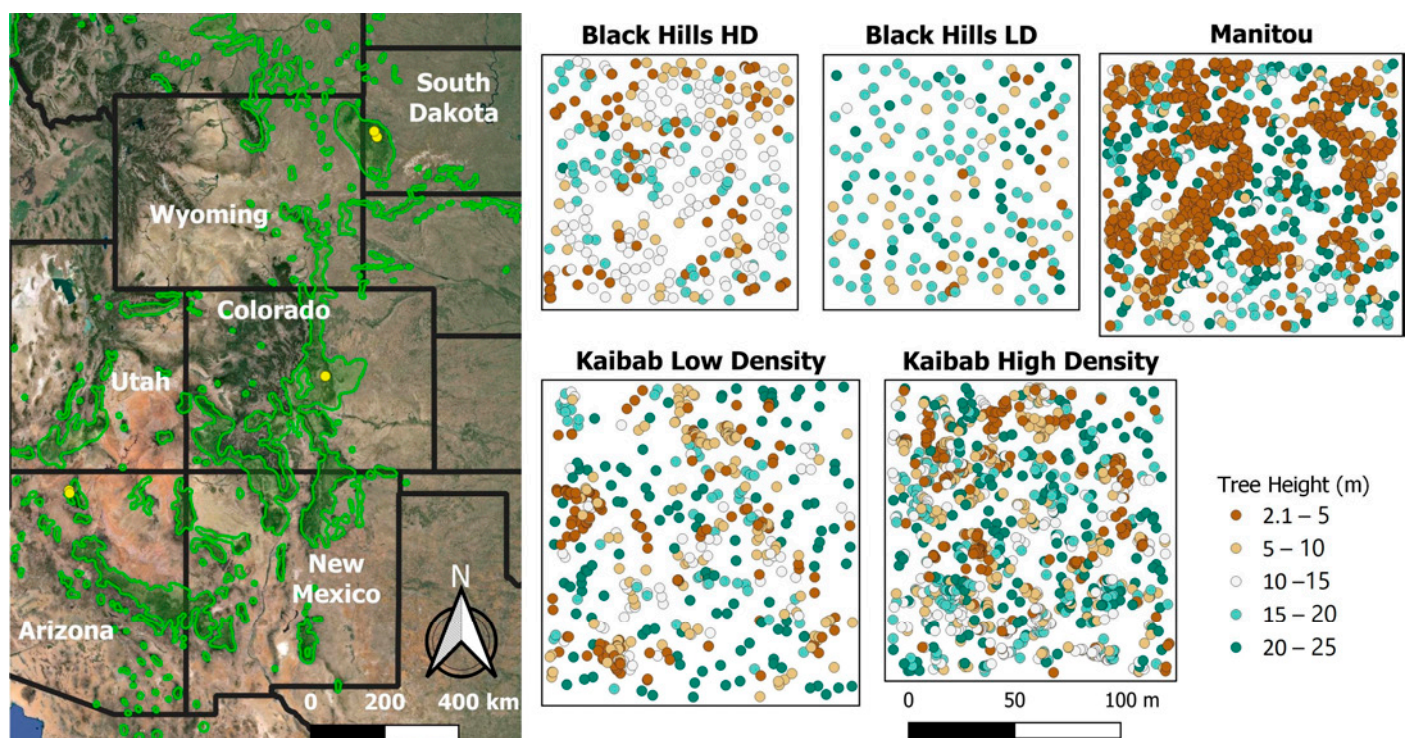


Figure 1. Study map showing (left) the distribution of ponderosa pine within the Central Rocky Mountains (green polygons; [24]) with yellow points indicating the five study locations. Small maps show the field validation data for each site colored by tree height.

2.2. Field Data Collection

Study sites were distributed across the central and southern Rocky Mountains within ponderosa pine-dominated stands. Sites include two stands in the Kaibab National Forest (AZ, USA), one at Manitou Experimental Forest within the Pike National Forest (CO, USA), and two at the Black Hills Experimental Forest as part of the Black Hills National Forest (SD, USA; Figure 1). All sites had greater than 99% ponderosa pine basal area ($\text{m}^2 \text{ha}^{-1}$) dominance but had contrasting past management histories that produced a range of forest densities and tree size variability (Table 1).

Table 1. Description of each study site's extent and forest structure density and size variation.

Site	Hectares	Trees ha^{-1}	Basal Area ($\text{m}^2 \text{ha}^{-1}$)	Height (m) *	DBH (cm) *
Kaibab High Density	1.7	574.1	39.6	12.8 (7.2)	24.0 (17.4)
Kaibab Low Density	2.1	246.5	22.5	12.5 (8.6)	27.3 (20.5)
Manitou	1.6	639.5	24.8	8.5 (7.3)	15.4 (16.0)
Black Hills High Density	1.0	308.3	11.2	11.1 (4.8)	19.7 (8.8)
Black Hills Low Density	1.0	171.6	14.9	15.7 (6.0)	30.7 (12.8)

* values reported as mean (standard deviation).

The two Kaibab stands are ~65 km southeast of Kanab, UT, USA, at 2400 m elevation with slopes <10% and were inventoried in 2018. These neighboring stands differ in that Kaibab Low Density (LD) experienced a restoration style thinning in 1993, while Kaibab High Density (HD) was reserved as a control for comparison. The Manitou stand is ~40 km northwest of Colorado Springs, CO, USA, at 2500 m elevation with slopes <5% and was inventoried in 2018. This site was selectively logged in the early- to mid-1880s and experienced minor mountain pine beetle damage in the late 1970s, resulting in a multi-aged forest structure following successive regeneration pulses [25]. The two Black Hills stands

are ~32 km northwest of Rapid City, SD, USA, at 1650 m elevation with slopes of 10–15%, and were inventoried in 2017. One stand received a commercial thinning in 2014, while the other was thinned using free selection in a way designed to promote horizontal and vertical heterogeneity in 2012.

At each study site, a grid of survey points was established using a Pentax PCS-515 (TI Asahi Co., Saitama, Japan) total station and an Emlid Reach-RS2 (Emlid Tech Kft, Budapest, Hungary) real-time kinetic GPS. Using the grid of survey points, all trees >2.0 m tall were stem-mapped within each of the study sites. Trees were located based on their distance and direction from their nearest survey point. Each tree was characterized for diameter at breast height (DBH; 1.37 m above ground) and total height, with height observations taken using a TruPulse 200 laser hypsometer (Laser Technology, Inc., Centennial, CO, USA) considered capable of providing ± 5 –10% accuracy.

2.3. UAS Image Acquisitions

All UAS acquisitions were completed with a DJI Phantom 4 Pro equipped with a 20-megapixel (5472 pixels $\times$ 3648 pixels) metal oxide semiconductor (CMOS) red–green–blue sensor with a fixed 8.8 mm focal length. Each acquisition was pre-programmed and conducted using Altizure (version 4.6.8.193; Shenzhen, China) for Apple iOS. While all flight plans followed a standard serpentine path with nadir camera orientation, there were small variations in flight altitude (80–99 m), forward (all at 90%) and side (85 or 90%) overlap, and flight speed (3 or 4 m s^{−1}). Specific flight plan parameters for each site are provided in Table S1. Even with added buffers around each acquisition area, all UAS flights were completed in a single battery, resulting in 132 to 157 photos for each of the study sites and averaging ~2.2 min per hectare of core collection area.

2.4. UAS Image Processing

Complete factorial combinations of each SfM algorithm's parameters were evaluated for the five UAS acquisitions. Across all SfM algorithms, imagery location was based on the x-, y-, and z-coordinate information stored in the EXIF (Exchangeable Image File Format) metadata on each photo without ground control point correction. Each of these algorithms integrate structure from motion to generate a sparse point cloud during the image alignment phase that then provides the necessary inputs to implement the multi-view stereo step to generate the dense point cloud. For each data set, the full SfM image processing time for dense point cloud generation and the point cloud density were recorded. The set of processing parameters used for image alignment in each SfM software package is available in Table S2, while the process used to generate the dense point cloud data sets used for analysis is described in the software-specific sections below. Because of the use of different computing hardware for each software algorithm, we report relative processing time by dividing the processing time for each parameter combination by the maximum time for that software algorithm. These relative processing times are available in Figure S1 and point cloud density summary statistics are reported in Figure S2.

2.4.1. Agisoft Metashape

Processing completed in Agisoft Metashape version 1.6.4 (Agisoft LLC., St. Petersburg, Russia), hereafter referred to as Metashape, utilized a computer with an Intel i7-9700 8-core central processor unit, a NVIDIA GeForce RTX 2060 graphics processing unit, and 64 gigabytes of random-access memory. The image alignment process is the initial SfM processing step which calculates the camera position and orientation using key point detection and matching across images. The image alignment process is outlined in Table S2 and was held constant to investigate how dense cloud generation parameters (quality and filtering) impact the resulting UAS point clouds and tree detection rates and accuracy. The Metashape SfM algorithm has two primary settings used during dense cloud generation, which is performed after image alignment. First, quality, which impacts the imagery resolution and has five levels from the original image resolution to a downscaling of

resolution by a factor of 256. Second, depth filtering, removes outlier points that could not be observed from enough independent images and has four levels ranging from no filtering (“Disabled”) to an “Aggressive” level that has been shown to significantly reduce point cloud density [15]. The combination of Metashape parameters generated 20 point clouds (the pairwise combination of five quality settings and four filtering modes) for each of the five sites for 100 total point clouds.

2.4.2. Pix4Dmapper

Processing completed in Pix4Dmapper, hereafter referred to as Pix4D, version 4.8.4 (Pix4D S.A., Prilly, Switzerland) utilized a computer with an Intel i9-13950HX 24-core computer processor unit, an NVIDIA GeForce RTX 4090 graphics processing unit, and 64 gigabytes of random-access memory. However, each data set was only given access to 12 central processor unit cores and 24 gigabytes of random-access memory. The initial processing step of the Pix4D SfM algorithm allows users to define the image size to be used to identify the key points for matching across images during image alignment. The image alignment process is detailed in Table S2 and was held constant at the default original image scale to investigate how dense cloud generation parameters impact the resulting UAS point clouds and tree detection rates and accuracy. After image alignment, the Pix4D SfM algorithm has two primary settings that users can adjust for generating the dense point cloud. First, image scale, which defines the scale of the images at which points are computed and has four levels from the original image resolution to a downscaling of one-eighth the original image size with a default of half the image size. Second, point density defines the density of the densified point cloud and has three levels ranging from high to low with a default setting of “Optimal” that is described to represent a balance between processing time and point density. The combination of Pix4D parameters generated 12 point clouds (the pairwise combination of four image scale settings and three point density modes) for each of the five sites, for 60 total point clouds.

2.4.3. OpenDroneMap

Processing, completed using OpenDroneMap, an open-source public domain code repository, utilized a server interface with Amazon Web Services with access to an Intel Xeon Platinum 8175M 4-core central processor unit and 16 gigabytes of random-access memory. The OpenDroneMap SfM algorithm has two primary settings used during dense cloud generation. First, feature quality, which scales image resolution during feature extraction for aligning the images to generate the point cloud with five levels from “Ultra” to “Lowest” with a default of “High”. Second, point cloud quality setting determines the density of the dense point cloud, with five levels from “Ultra”, which generates denser point clouds, to “Lowest”, with a default of “Medium”. The web server was not able to run the “Ultra” setting for point cloud quality, leaving the available OpenDroneMap parameter combinations to generate 20 point clouds (the pairwise combination of five feature quality settings and four point cloud quality settings) for each of the five sites for 100 total point clouds.

2.4.4. Alignment of Software Parameters

To allow for comparison of tree detection rates and accuracy across the software packages, we mapped the Pix4D and OpenDroneMap dense cloud parameter levels to the Metashape quality and filtering mode levels (Table S3). Aligning these parameter levels enables a broader understanding of the trends in tree extraction performance. While the tested parameters function similarly between the software packages, our aligning of the parameter levels is not meant to signify that the settings are identical.

2.5. UAS Point Cloud Processing

2.5.1. Height Normalization and CHM Generation

All processing of the point clouds derived from the algorithm and parameter combinations was identical for each of the 260 data sets (100 from both Metashape and OpenDroneMap and 60 from Pix4D). Point cloud processing was completed in the R statistical program language [26] using the lasR [27] package. All point cloud processing was completed on a computer with an Intel i7-10750H 6-core computer processor unit and 32 gigabytes of random-access memory. Because all point cloud processing and tree extraction occurred using the same computer hardware, we report summary statistics of total point cloud processing time in Figure S2. Initial processing removed points with duplicate coordinates. Each data set was processed using the Cloth Simulation Filter ground segmentation algorithm [28] to classify points as being part of the ground surface as implemented by the RCSF package's [29] 'CSF' command with default settings. After classification, points that had zero or only a few other surrounding points were identified as noise and removed from further processing using the 'classify_isolated_points' command in the lasR package [27] with default settings. The point cloud was then height-normalized based on the ground classified points filtered to include a maximum of 20 points per m² using the Delaunay triangulation approach implemented in the lasR 'normalize' command [27]. Next, the height-normalized non-ground points were used to generate a canopy height model (CHM) at 0.25 m resolution using the lasR 'rasterize' command with the points-to-raster method using a local maximum approach of point heights filtered between 2 and 60 m and processed using the algorithm from St-Onge [30] to fill pits and spikes in the raster.

2.5.2. Tree Detection and DBH Modeling

Each canopy height model raster was processed using a local maximum filter with a variable window size [31] via the lidR 'locate_trees' command [32] to identify the position and height of individual trees with a minimum height of 2 m. We defined the variable window size using the *window search radius* (m) function in Equation (1) with parameterization based on Creasy et al. [3] with a lower bound of 1 m and an upper bound of 5 m.

$$\text{window search radius}(m) = g(\text{CHM}) = \begin{cases} 1 & 0 \leq \text{CHM} < 2 \\ 0.75 + (\text{CHM} \times 0.14) & 2 \leq \text{CHM} \leq 30 \\ 5 & \text{CHM} > 30 \end{cases} \quad (1)$$

where the *CHM* represents the individual pixel values in meters. This approach to individual tree detection works by iterating over every pixel in the CHM, defining the *window search radius*, and evaluating if the focal pixel is the tallest value within the window.

We developed site-specific allometric equations using data from the USDA Forest Service's Forest Inventory Analysis (FIA) program [33] to estimate individual tree DBH based on the UAS-detected tree height [34]. Relationships of height predicting DBH were based on the methods from Swayze and Tinkham [35] and fit using Equation (2) as second-order polynomials via the Bayesian modeling package brms [36] with the strictly positive response variable of DBH modeled using the Gamma likelihood with a log link. For each study site, the model was run with 4 chains, each chain had 4000 samples, with a warmup of 2000 samples and 8000 total post-warmup samples. Traceplots and R-hat values were assessed for proper mixing and model convergence [37]. Models were fit to FIA plot data within a 100 m square buffer of the respective UAS flight boundary. Representative FIA plots for each study site were identified using TreeMap [38], a model of FIA plot locations imputed throughout forested areas of the conterminous United States at 30 m spatial resolution. The resulting equation (Equation (2)) for each of the five study sites was used to predict a *DBH* value for each UAS-detected tree.

$$\text{DBH}(cm) = b_0 \times \text{Height}(m) + \text{Height}(m)^{b_1} \quad (2)$$

We then used the UAS-derived tree list with individual tree location, height, and DBH to estimate the stand basal area ($\text{m}^2 \text{ ha}^{-1}$) of trees within a study site's extent for comparison to field-estimated basal area based on the stem-mapped tree list filtered using a minimum tree height of 2 m (see Section 2.2 above). For each of the 260 data sets (100 from both Metashape and OpenDroneMap and 60 from Pix4D) we calculated the percentage error in basal area to represent bias (i.e., underestimation/overestimation of field-measured basal area by the UAS method) and the absolute percentage error in basal area to represent precision (i.e., how “close” the UAS value is to the field value).

2.6. Individual Tree Measurement Comparison

The UAS-SfM detected trees were matched with the field-inventoried trees through an iterative process. Prior to implementing the matching process, extracted trees from the highest density point clouds at each site were visually compared to the field-inventoried trees and revealed a less than 2 m horizontal miss-alignment for each site. Within a data set, a UAS-SfM target tree was selected and intersected with all field trees within a 3 m radius and within 2 m in height of the target UAS tree location and height. If there were multiple matched field trees, then the tree closest in height was considered the True Positive match and the matched UAS and field trees were removed from further matching. The remaining UAS and field trees were matched iteratively until no additional tree pairs could be identified using this process. If no field tree could be identified after this matching process, the UAS tree was considered a False Positive with any remaining unmatched field trees classified as False Negatives. Based on the *True Positives*, *False Positives*, and *False Negatives* in each data set, *F-score* was calculated using Equation (3) as an overall metric of individual tree detection quality.

$$F - score = 2 \times \frac{\left(\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \times \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \right)}{\left(\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} + \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \right)} \quad (3)$$

We used the True Positive trees to compare UAS-derived values to the field-measured values of tree height and DBH. For each of the 260 data sets, we calculated the mean error to represent UAS estimate bias and the root mean squared error (RMSE) to represent UAS estimate precision of tree height and DBH. Results from this analysis are presented in Supplementary S3.

2.7. Statistical Analysis

To evaluate how the SfM processing software and algorithm parameters impact overall tree detection based on the F-score, stand basal area quantification, and tree height and DBH measurement, we utilized a Bayesian hierarchical approach analogous to the traditional multifactor analysis of variance (ANOVA) implemented with the Bayesian modeling package brms [36] in the R statistical program [26]. Based on the observed data, our hierarchical Bayesian model allows us to estimate a baseline (i.e., the “overall mean”) for our dependent variable of interest, plus a deflection due to the level of another factor, plus a residual deflection due to the interaction of factors [39,40]. Models were run using four chains with 20,000 iterations each with a warmup of 10,000 samples and 40,000 total post-warmup samples. We modeled the F-score, which can take on continuous values restricted within the range of [0, 1], using a beta likelihood with a logit link function. We modeled the basal area percentage error using a Gaussian likelihood and we modeled the basal area absolute percentage error which is restricted to being positive using a Weibull likelihood function with a log link. We focus the results of this paper on the influence of SfM processing parameter settings on tree detection (F-score) and stand basal area quantification and report results of tree height and DBH measurement in Supplementary S3. Trace plots were utilized to visually assess model convergence, and sufficient convergence was checked with R-hat values near 1 [41]. Posterior predictive checks were used to evaluate model goodness-of-fit by comparing data simulated from the model with the observed

data used to estimate the model parameters [37]. We describe the mathematical details of these models, the computational procedures we used to fit them, and methods for model checking in Supplementary S4. We follow the guidelines outlined by Kruschke [42] for reporting Bayesian analysis details with the aim of improving the transparency and reproducibility of our analysis. To compare levels of a single SfM parameter setting (e.g., dense point cloud quality), we collapsed across the other factors in our model to perform main effect comparisons or contrasts [39,40]. To determine the probability (Pr) that a SfM parameter setting effect was positive (or negative), we calculated contrasts from the posterior predictive distribution as the proportion of the posterior distribution above (or below) zero [39] using the tidybayes package [43]. We summarized the uncertainty of the posterior distribution by reporting the 95% highest density interval (HDI) with the median value as the measure of central tendency.

3. Results

3.1. Point Cloud Comparison

Point cloud densities generated by the different SfM parameterizations ranged from around ten points per m^2 at the lowest dense cloud quality setting to over 3000 points per m^2 at the highest dense cloud quality setting. For each software, point cloud density generally increased at higher levels of dense point cloud generation quality except when increasing quality from lowest to medium within OpenDroneMap. Within each quality setting, the filtering setting had a small impact on point density variation within both Metashape and OpenDroneMap but greatly impacted the point density within Pix4D. Similarly, the time to produce the dense point cloud varies in a similar pattern across the algorithm parameters, with the dense cloud quality setting being the controlling parameter (Figure S1). Point cloud and tree detection processing time increased as point density increased and ranged from lows of less than one minute per ha when point density was 100 points per m^2 or less, to highs of over ten minutes per ha when point density was over 3000 points per m^2 . Complete summary statistics of point cloud density and point cloud processing time for performing tree detection and size estimation on the 260 data sets analyzed (100 Metashape, 100 OpenDroneMap, and 60 Pix4D data sets) are reported in Figure S2.

3.2. Overall Tree Detection Performance (F-Score)

We described overall tree detection performance using the F-score metric which can take on continuous values between 0 and 1 where a value of 1 indicates perfect matching of UAS-detected and field-surveyed trees. Across all SfM processing software and algorithm parameterizations tested on the five study sites, F-score ranged from a low of 0.0 to a high of 0.9 (Figure 2; Table S4). Across all software, tree detection performance generally increased at higher levels of dense point cloud generation quality and with less filtering of the dense point cloud (Figure 2).

3.2.1. Dense Point Cloud Generation Quality Setting

After collapsing across the filtering mode setting to compare the dense cloud quality setting effect within each software, there was a clear positive effect on tree detection of increasing quality using Metashape in successive quality steps from lowest to high, but there was weaker evidence that increasing quality from high to ultra high had a positive effect on tree detection (Pr. 63%; Figure 3A). Using Metashape, the highest estimated F-scores of 0.61 (95% highest density interval [HDI] = [0.40, 0.82]) and 0.58 ([0.38, 0.80]) were achieved using ultra high and high dense point cloud generation quality, respectively (Table S4). For OpenDroneMap, there was little evidence that increasing dense point cloud generation quality improved tree detection performance at the lower quality settings (lowest to medium) but increasing quality from medium to the higher quality settings had a strong positive effect on tree detection, and increasing quality from the medium to high (Pr. $\geq 99\%$) and from the high to ultra high (Pr. 94%) setting increased tree detection

performance (Figure 3A). Using OpenDroneMap, the highest estimated F-score of 0.58 ([0.34, 0.81]) was achieved using ultra high dense point cloud generation quality (Table S5). Increasing dense point cloud generation quality in Pix4D from the low to the medium setting improved tree detection performance (Pr. 87%) but there was little evidence that increasing quality to the higher settings had a positive effect (Figure 3A), with an F-score of 0.53 ([0.28, 0.77]) achieved at medium quality and 0.54 ([0.29, 0.78]) at ultra high quality (Table S5).

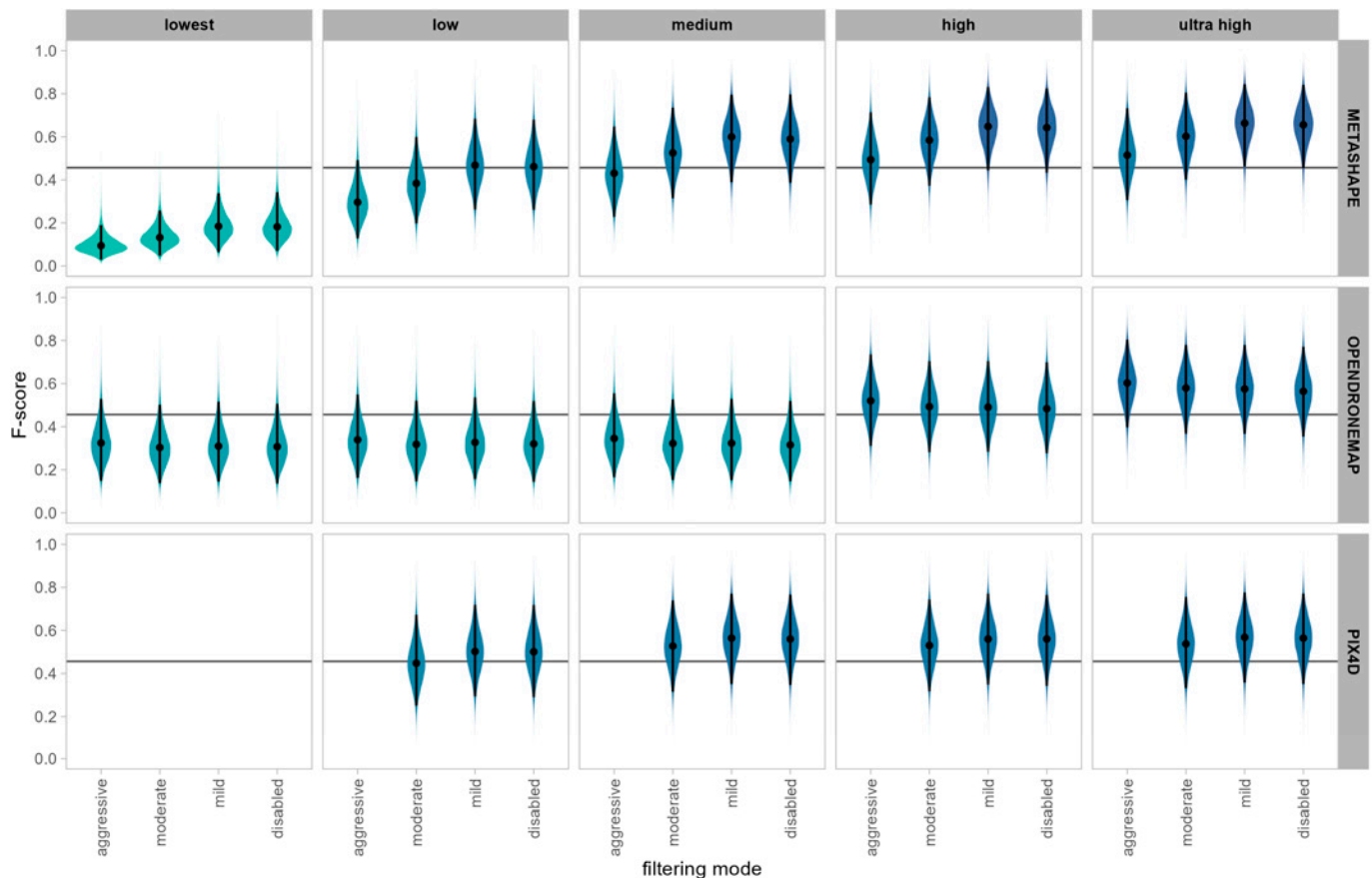


Figure 2. Overall UAS tree detection performance by software platforms (rows), dense point cloud quality (columns), and filtering mode (x-axis) settings of the SfM algorithms. Greater F-score values (darker blue) indicate better UAS tree detection. The overall mean across all software and SfM parameterizations is represented by the horizontal black line and each cell is represented by the posterior predictive distribution where the dot is the median posterior predicted value, and the vertical segment indicates the 95% highest density interval (HDI). Tabular data are available in Table S4.

3.2.2. Dense Point Cloud Filtering Mode Setting

After collapsing across the dense point cloud generation quality setting to compare the dense cloud filtering mode effect within each software, there was a clear positive effect on tree detection when limiting the amount of filtering of the dense point cloud. For Metashape, the best F-score of 0.50 ([0.22, 0.79]) was achieved from mild filtering (Pr. 95%), with little evidence to support the disabled filtering level (Figure 3B; Table S6). Within OpenDroneMap, adjusting the filtering mode setting had limited impact on overall tree detection performance, with limited improvement in performance obtained by decreasing the filtering of the dense point cloud, and with the most aggressive filtering resulting in the best tree detection performance (Pr. ~70%; Figure 3B) with an F-score of 0.44 ([0.17, 0.72]; Table S6). Within Pix4D, adjusting the filtering mode also had limited impact on tree

detection performance, with the best F-score of 0.49 ([0.25, 0.75]; Table S6) achieved at the mild level of filtering (Pr. 79%; Figure 3B).

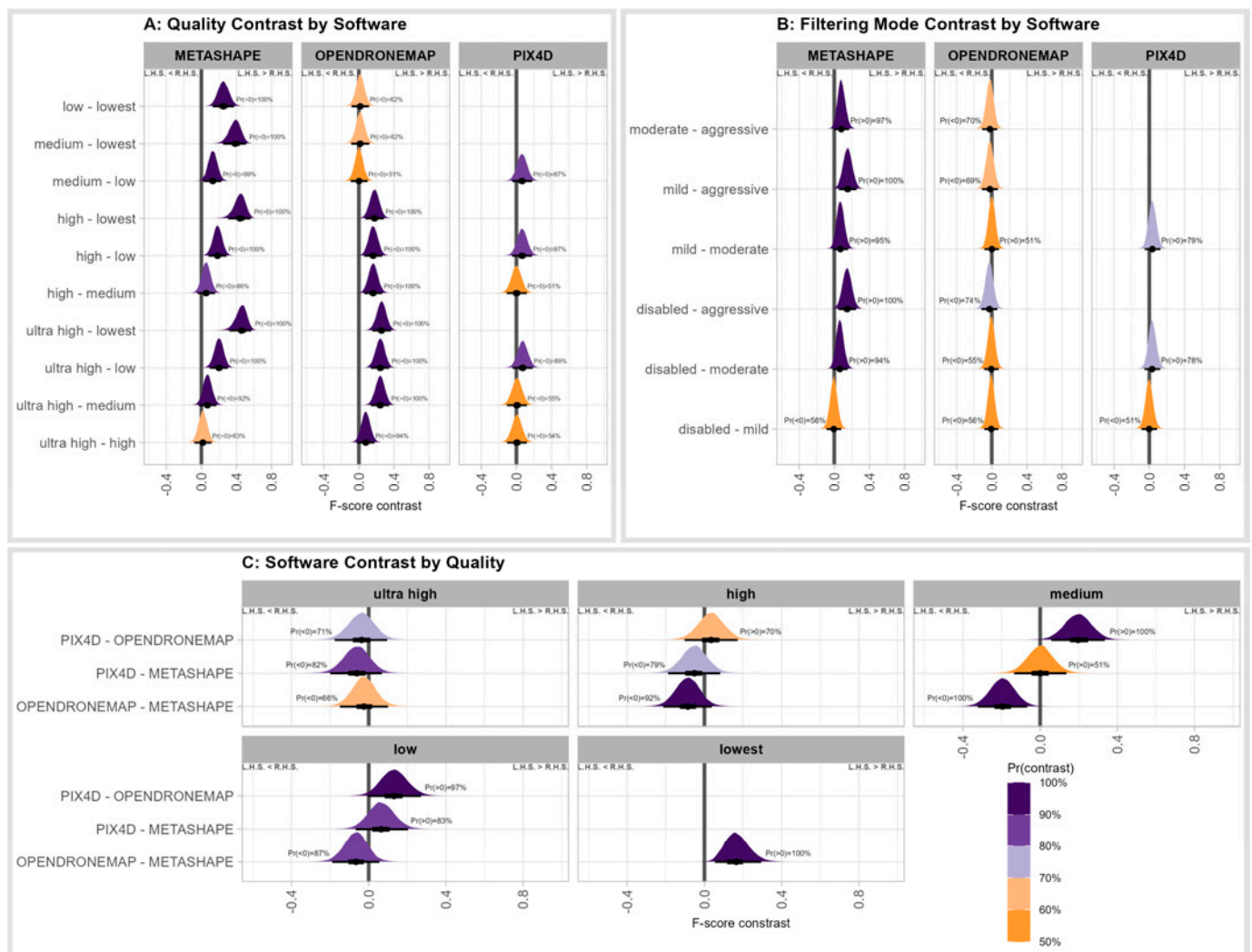


Figure 3. Change in UAS tree detection performance as influenced by (A) the dense point cloud generation quality setting within each software, (B) dense point cloud filtering mode within each software, and (C) software platform within each dense point cloud generation quality setting. Each contrast shows the posterior predictive distribution of the difference in F-score (i.e., contrast) where the dot is the median posterior predicted value, and the horizontal segment indicates the 95% highest density interval (HDI). Colors represent the probability of one setting or software outperforming another. High F-Score values (near one) indicate high UAS tree detection accuracy and, as such, a contrast (i.e., difference) in F-Score that is positive (difference > 0) implies that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. Tabular data available in Tables S7–S9.

3.2.3. Software

We compared overall tree detection performance between software by collapsing across the filtering mode setting to compare the effect of software within each dense point cloud quality level (Figure 3C). Across all levels of dense point cloud quality, Metashape and Pix4D performed similarly with respect to overall tree detection, with Metashape performing slightly better (Pr. ~80%) at higher quality settings and Pix4D performing slightly better (Pr. 84%) at lower quality (Figure 3C). OpenDroneMap tree detection performance was similar to Pix4D and Metashape at the ultra high dense point cloud generation quality

and to Pix4D at high quality, but there was strong evidence that OpenDroneMap performed worse than Metashape at intermediate (low, medium, high) quality levels (Pr. 87% to $\geq 99\%$; Figure 3C) and worse than Pix4D at medium and high quality (Pr. $\geq 95\%$; Figure 3C). OpenDroneMap produced F-score values averaging 0.09 points lower ($[-0.21, 0.04]$) than Metashape at high quality and 0.19 points lower ($[-0.32, -0.07]$) at medium quality (Figure 3C; Table S7).

3.3. Stand Basal Area

We calculated the percentage error of the UAS-derived basal area compared to field measurements to represent measurement bias and the absolute percentage error to represent precision. Percentage error in the basal area can take on continuous positive values (UAS > field measurement; overestimation) or negative values (UAS < field measurement; underestimation) with a lower limit of -100% (in cases with zero UAS trees detected and >0 field identified trees), with values close to zero indicating low measurement bias. Across all SfM processing software and algorithm parameterizations tested on the five study sites, basal area percentage error ranged from a low of -100% to a high of 52% (Figure 4A; Table S10). Basal area was generally underestimated by the UAS-SfM methodology at the low and lowest dense point cloud generation quality settings but shifted to either no bias or slight overestimation of basal area as the dense point cloud generation quality increased to the high and highest levels (Figure 4A; Table S10). Absolute percentage error in the basal area is strictly positive with a lower-bound of zero representing cases where UAS values perfectly match field measurements. Across the 260 data sets tested, absolute percentage error in the basal area ranged from a low of 0% to a high of 100% with each SfM software type achieving its most precise estimate within approximately 13% [$5\%, 24\%$] of the field-measured basal area (Figure 4B; Table S11).

3.3.1. Dense Point Cloud Generation Quality Setting

After collapsing across the filtering mode setting to compare the dense cloud quality setting's effect on UAS-derived stand basal area precision within each software, there was a clear improvement in precision (i.e., reduction in absolute percentage error) by increasing quality using Metashape. The most precise estimates of basal area using Metashape were achieved at the highest quality levels (Figure 4B; Table S12) with an absolute percentage error of 16% [$5\%, 31\%$]; Table S12) at the high quality level and only limited evidence that increasing to ultra high quality could improve precision (Pr. 66% ; Figure 5A) to 15% [$5\%, 29\%$]; Table S12). Using OpenDroneMap, the lowest absolute percentage error in basal area of 14% [$5\%, 27\%$]; Table S12) was realized at the ultra high dense point cloud quality setting, and there was a clear improvement in the precision of the estimate compared to lower quality settings (Pr. $> 90\%$; Figure 5A). There was some evidence that increasing dense point cloud generation quality to the high (Pr. 69%) and ultra-high (Pr. 73%) settings improved basal area precision when compared to the medium quality setting using Pix4D (Figure 5A). Using the ultra high quality setting in Pix4D yielded a basal area absolute percentage error of 14% [$4\%, 27\%$] compared to 14% [$4\%, 28\%$] and 16% [$5\%, 32\%$] using the high and medium setting, respectively (Table S12).

3.3.2. Dense Point Cloud Filtering Mode Setting

After collapsing across the dense point cloud generation quality setting to compare the dense cloud filtering mode effect within each software, the model and data suggest that the dense point cloud filtering mode setting had limited impact on UAS basal area absolute percentage error (Figures 4B and 5B; Table S13) in all three software platforms tested. In each software, the results suggest that improved basal area precision is achieved with less filtering of the dense point cloud (i.e., with mild or disabled filtering) but precision improvements were small ($<3\%$) compared to the moderate filtering setting (Figure 5B). There was modest evidence that the lowest absolute percentage error in basal area was achieved

using the disabled filtering setting in both Metashape (Pr. 67%) and OpenDroneMap (Pr. 75%) compared to the mild filtering setting (Figure 5B; Table S13).

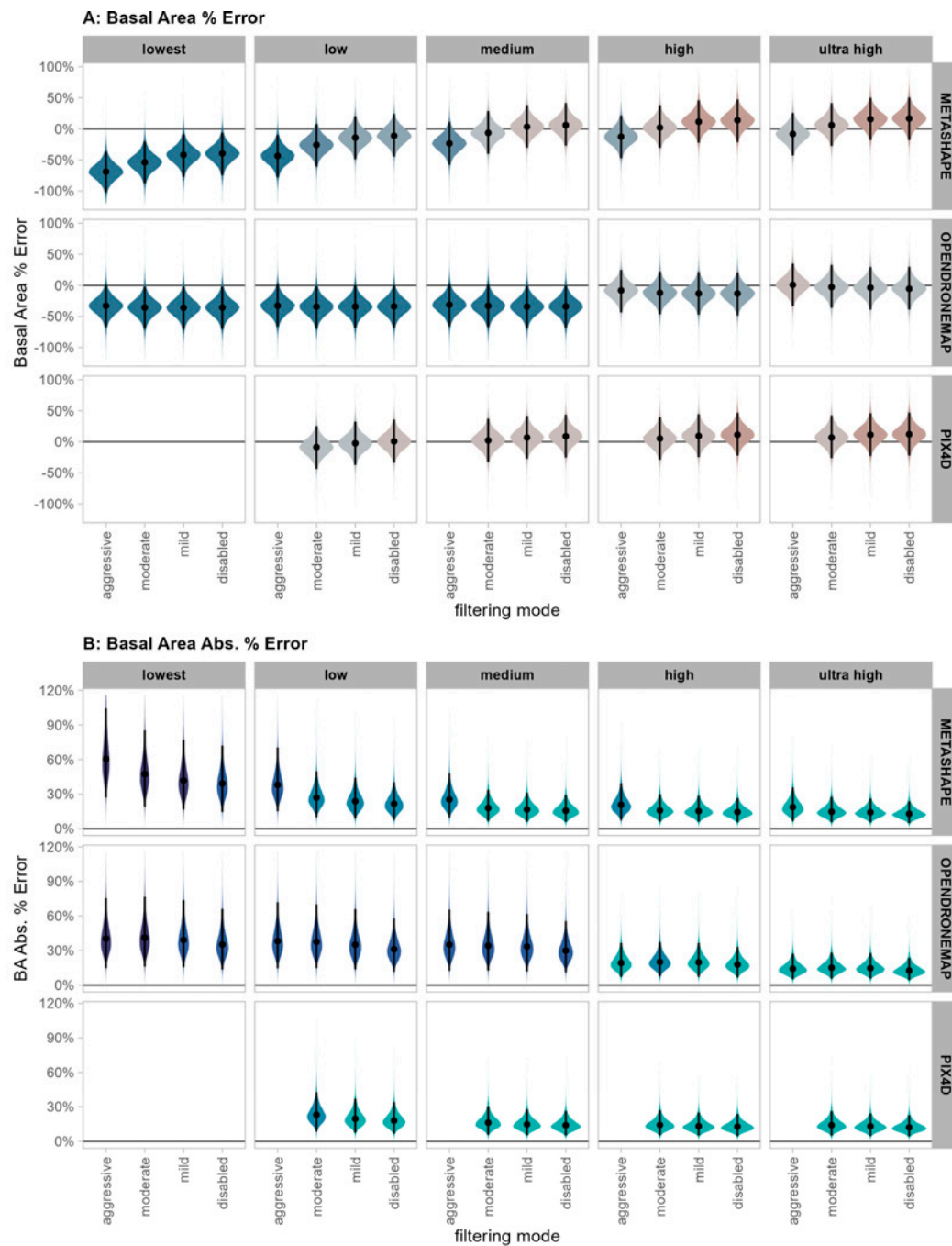


Figure 4. UAS-detected stand basal area (BA) bias (percentage error; top panel (A)) and precision (absolute percentage error; bottom panel (B)) by software platforms (rows), dense point cloud quality (columns), and filtering mode (x-axis) settings of the SfM algorithms. (A) Positive bias values (red) indicate UAS values overestimate field measurements, while negative bias values (blue) indicate UAS values underestimate field measured values. (B) Absolute percentage error is strictly positive with higher values (darker blue) implying less precise UAS estimates and a lower-bound of zero representing cases where UAS values perfectly match field measurements. Each cell represents the posterior predictive distribution where the dot is the median posterior predicted value, and the vertical segment indicates the 95% highest density interval (HDI). Tabular data are available in Tables S10 and S11.

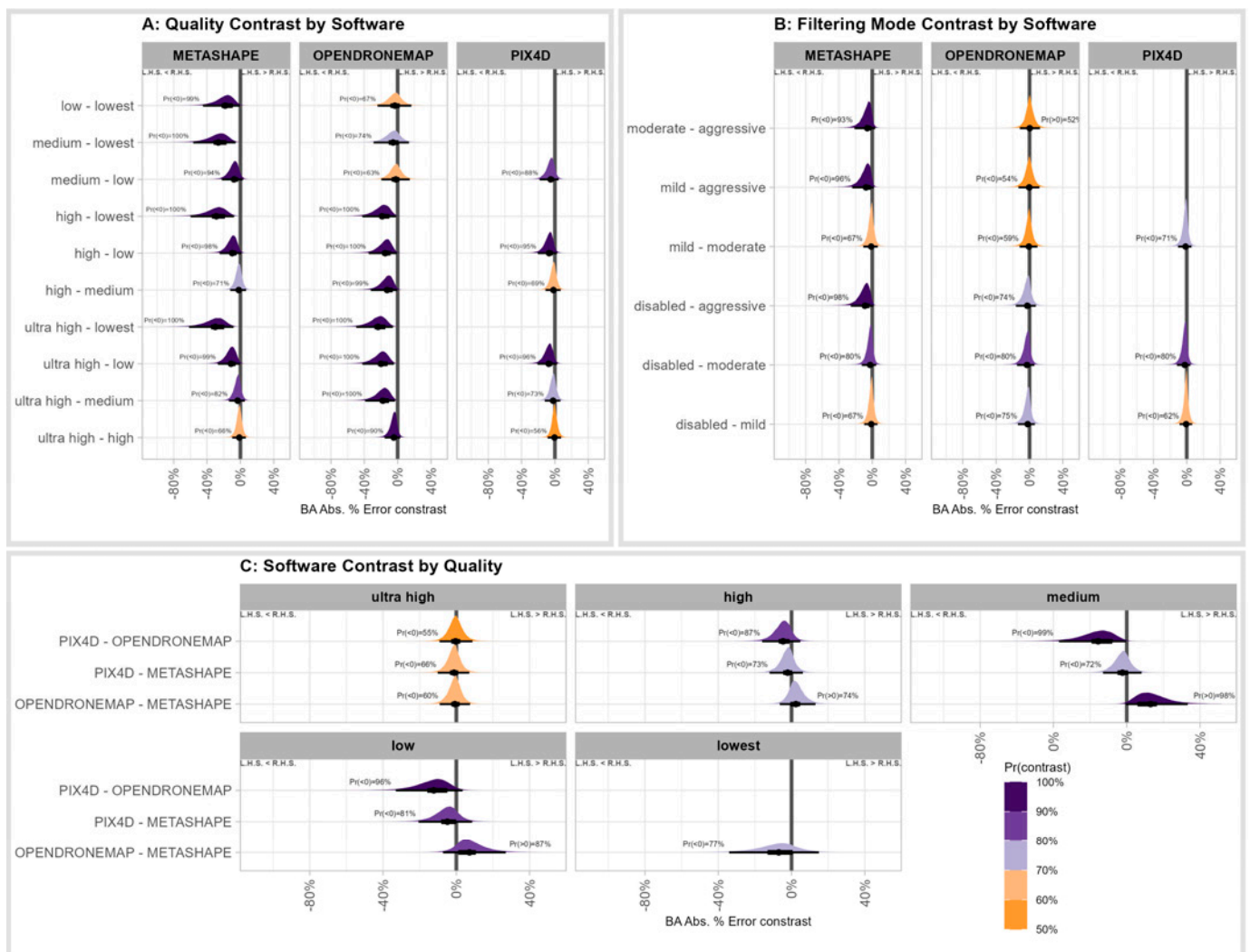


Figure 5. Change in UAS-derived stand basal area precision (absolute percentage error) as influenced by (A) dense point cloud generation quality setting within each software, (B) dense point cloud filtering mode within each software, and (C) software within each dense point cloud generation quality setting. Each contrast shows the posterior predictive distribution of the difference in stand basal area absolute percentage error (i.e., contrast) where the dot is the median posterior predicted value, and the horizontal segment indicates the 95% highest density interval (HDI). Colors represent the probability of one setting or software outperforming another. Lower absolute percentage error values mean that UAS values better match field measurements and, as such, a contrast (i.e., difference) in absolute percentage error that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. Tabulated data available in Tables S14–S16.

3.3.3. Software

We compared UAS-derived stand basal area accuracy between software by collapsing across the filtering mode setting to compare the effect of software within each dense point cloud quality level (Figure 5C; Table S14). There was no difference in basal area precision between the three software packages at the ultra high dense point cloud generation quality (Figure 5C). Across all other quality levels, Metashape and Pix4D performed similarly with respect to estimating stand basal area, with Pix4D estimates being slightly more precise (~2%) at higher quality settings (Figure 5C; Table S14). Using OpenDroneMap resulted in a small decrease in basal area precision at the high quality level compared to the other two

software packages, followed by significant decreases in precision (approximately 13–16%) at the medium quality level ($\text{Pr.} \geq 98\%$; Figure 5C; Table S14).

4. Discussion

4.1. Variation in Algorithm Performance

Comparison of the best F-scores achieved by each software shows that when algorithm parameters are ideally selected, each software can produce F-score values comparable to the best documented performance in open canopy conifer-dominated forests (i.e., 0.75 to 0.90 [8,12]), with the best tree extraction performance generally occurring at the highest level for dense point cloud generation quality. However, at each software's best settings, Metashape did provide a median improvement in F-score over OpenDroneMap of 0.02 and Pix4D of 0.06, respectively (Table S7).

Tree extraction performance varied greatly across the range of algorithm parameters, which is in line with the variation in point density and subsequently tree extraction processing time (Figure S2). While OpenDroneMap achieves its highest tree extraction rates using the ultra high dense point cloud quality setting (0.08 median F-score improvement from high), Metashape shows only a minor improvement of 0.02 median F-score improvement from high to ultra high, and Pix4D had only a 0.01 median F-score improvement from medium to ultra high. These relatively small improvements in tree extraction might suggest that using a lower point cloud quality setting may not sacrifice that much data quality, while providing large data processing and storage saving. These results generally align with previous studies that have suggested that maintaining image resolution during point cloud generation will retain canopy structure for differentiating individual trees [15,44].

Using the highest dense point cloud generation quality level (Figure 5A) at which the probability of improvement in basal area precision compared to the prior level was 70% or greater at the disabled filtering mode, the most precise UAS-SfM basal area estimate achieved by each software in this study was as follows: 13% (95% HDI = [5%, 24%]) using Metashape, 13% ([5%, 24%]) using OpenDroneMap, and 12% ([5%, 23%]) using Pix4D (Table S11). These basal area precision estimates are comparable to basal area precisions obtained using UAS-SfM methods in previous research, including the following: 11% in mechanically thinned ponderosa pine forests [9]; 19% in planted subtropical forests [45]; 15% and 24% in even-aged, managed boreal forests [46,47]; and 16% in managed coniferous stands in even-aged to uneven-aged conditions in eastern Belgium [48]. Despite these achieved precisions for basal area, which mostly fall inside the common United States public land forest inventory design standard of $\pm 20\%$ allowable error [49], UAS-SfM methodologies for estimating stand-level metrics may need to be further developed to appropriately represent different forest types and developmental stages [10]. Fraser and Congalton [6], for example, found that UAS-SfM methods overestimated stand basal area by 15–42% compared to field-based measurements in complex mixed-species forests of the northeastern United States.

Our data and analysis suggest that the filtering mode setting had a limited impact on tree extraction and stand basal area estimation. Filtering mode by itself had the weakest effect on UAS-derived forest structure, and this comes with relatively high certainty (Figures S3 and 5B). While there was strong evidence that the basal area precision was optimized when filtering of the dense point cloud was minimized (i.e., set to disabled or mild), basal area precision only improved by 1–3% compared to more aggressive filtering modes (Table S16). These results agree with other studies that have found that the fidelity of CHMs is best maintained when applying SfM algorithms with minimal filtering being applied to realize minor improvements in forest structure representation [15,50].

As UAS monitoring of forest structures matures it will be important to understand how localized forest structures influence the reliability of different metrics. Our analysis included study sites with variable stand density conditions (Table 1) and this variability was reflected in the uncertainty in UAS-SfM tree extraction performance even at optimal settings (Figure 2). For Example, using Metashape, the best F-score ranged from 0.46 to

0.84 (Table S4), and basal area absolute percentage error ranged from 5% to 24% with 95% probability (Table S11). Inspection of the data showed that the quality of tree extraction declined (lower F-score) as the proportion of smaller, intermediate, and suppressed trees increased within a site. This common under-extraction of smaller size-class trees from individual tree detection methods, applied through top-down methods, has been widely reported in UAS structure from motion and LiDAR literature [9,51]. It has been suggested that while small trees are under-extracted, these methods likely capture within-stand spatial variation in small tree density [12], but further testing is necessary to better understand how reliably different tree size classes are represented. Within this study, the variable window function tree extraction was not tuned specifically for a single site; it is likely that extraction of the smallest trees could be improved, but these top-down methods will always underestimate the presence of small trees. The relatively narrow range of basal area precision compared to F-score values is attributed to the increased importance of the larger trees that these top-down tree extraction methods reliably identify. In the sparser SfM point clouds, with either lower point cloud quality generation settings or more aggressive filtering settings, the biases to under-detect trees and to underestimate their height likely compound to explain the significant basal area underestimation in these datasets. In the best tree extraction datasets, the remaining basal area error likely propagates through from the local height to DBH relationship failing to capture how local stand density can influence stem diameter. Similar effects have been reported in the past [52]. Future efforts should evaluate how integrating competition metrics into this relationship could improve DBH prediction and subsequent basal area estimation.

4.2. Interaction of Data Quality with Processing Time and Data Storage

Successful adoption of UAS forest inventory approaches will require that the time investment be balanced with any data improvements the system might yield. Reducing each software from ultra high to high point cloud quality setting achieved a point cloud generation time savings for Metashape, OpenDroneMap, and Pix4D of 77%, 86%, and 79%, respectively (Figure S1). Additionally, reducing the point cloud quality setting from ultra high to high for each software resulted in a ~75% reduction in point cloud processing and tree extraction time, along with ~75% savings in data storage need. These time and data storage savings were realized without sacrificing improvement in tree detection or basal area precision using Metashape and Pix4D, with results indicating a very low probability of improving F-score (Pr. 54–63%; Figure 3A) or basal area precision (Pr. 56–66%; Figure 5A) by using the ultra high quality setting over the high setting. However, for OpenDroneMap, it is highly probable that the change from high to ultra high quality resulted in improved tree detection performance (Pr. 94%; Figure 3A) and basal area precision (Pr. 90%; Table S15) by ~5%, but this setting change also increased processing time by ~86% (Figure S1). These findings point to the potential for real savings in data generation, storage, and processing time when operationalizing UAS-SfM forest monitoring strategies using Metashape and Pix4D. However, the larger loss in data accuracy for OpenDroneMap might indicate that the time and data storage savings by reducing from ultra high to high point cloud quality is not justified.

5. Conclusions

This study evaluated the influence of structure from motion algorithm settings for three popular software packages on tree extraction and stand basal area estimation in ponderosa pine dominated forests. Implementation of spatially informed forest management using UAS-SfM techniques requires an understanding of the tradeoffs between accuracy and time efficiency to promote reproducible workflows. While tree extraction performance varied across algorithm settings within a software, the best tree extraction for each software only varied by F-score = 0.06 when the algorithm parameters are ideally selected. Tree extraction and stand basal area accuracy and precision were maximized using a combination of settings with the highest point cloud density quality setting and the mildest point cloud

filtering setting. This analysis showed that reducing the point cloud quality setting one level resulted in an average time savings of 70–80% without reducing tree detection performance or basal area precision using Metashape and Pix4D, but the same reduction in point cloud generation quality using OpenDroneMap decreased forest structure measurement accuracy by ~5%.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/rs16203844/s1>. **Supplementary Material S1:** Table S1. Full UAS data collection parameters for the six study sites. Table S2. SfM image alignment algorithm processing parameterization. Table S3. Mapping of Pix4D and OpenDroneMap dense cloud generation parameterization options to the Metashape settings. Within Pix4D, the “Image Scale” setting most closely corresponds to the dense cloud quality while the “Point Density” setting represents the filtering mode. Within OpenDroneMap, the “Feature Quality” setting most closely corresponds to the dense cloud quality, while the “Point Cloud Quality” setting represents the filtering mode. Figure S1. Summary statistics of dense point cloud generation time reported as the relative time to generate the dense point cloud compared to the maximum for each software. Each boxplot summarizes the distribution of the data from five study sites for each software (rows), quality setting (columns), and filtering mode (x-axis) combination. The reported times were relativized as each type of software was available on computer systems with differing hardware, but still allow for assessment of how algorithm parameters influence time. Figure S2. Summary statistics of point cloud density measured in points per m² (top panel) and point cloud processing time measured in minutes per ha (bottom panel) to perform tree detection and DBH estimation. Note the logarithmic scale used on each y-axis. Each boxplot summarizes the distribution of the data, shown in the rug on the left-hand-side of each cell, from five study sites for each software (colors), quality setting (columns), and filtering mode (x-axis) combination. **Supplementary Material S2:** Table S4. Posterior predicted F-score and 95% HDI by dense point cloud generation quality, filtering mode, and software after collapsing across other model factors. Table S5. Posterior predicted F-score and 95% HDI by dense point cloud generation quality and software after collapsing across other model factors. Table S6. Posterior predicted F-score and 95% HDI by dense point cloud filtering mode and software after collapsing across other model factors. Table S7. Posterior predicted difference in F-score (i.e., contrast) and 95% HDI between software by dense point cloud generation quality after collapsing across other model factors. High F-Score values (approaching one) indicate high UAS tree detection accuracy and, as such, a contrast (i.e., difference) in F-Score that is positive; a (difference > 0) implies that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is positive was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was greater than zero. Table S8. Posterior predicted difference in F-score (i.e., contrast) and 95% HDI between dense point cloud generation quality by software after collapsing across other model factors. High F-Score values (approaching one) indicate high UAS tree detection accuracy, and, as such, a contrast (i.e., difference) in F-Score that is positive (difference > 0) implies that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is positive was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was greater than zero. Table S9. Posterior predicted difference in F-score (i.e., contrast) and 95% HDI between dense point cloud filtering modes by software after collapsing across other model factors. High F-Score values (approaching one) indicate high UAS tree detection accuracy, and, as such, a contrast (i.e., difference) in F-Score that is positive (difference > 0) implies that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is positive was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was greater than zero. Table S10. Posterior predicted basal area (BA) percentage error (bias) and 95% HDI by dense point cloud generation quality, filtering mode, and software after collapsing across other model factors. Table S11. Posterior predicted basal area (BA) absolute percentage error (precision) and 95% HDI by dense point cloud generation quality, filtering mode, and software after collapsing across other model factors. Table S12. Posterior predicted basal area (BA) absolute percentage error (precision) and 95% HDI by dense point cloud generation quality and software after collapsing across other model factors. Table S13. Posterior predicted basal area (BA) absolute percentage error (precision) and 95% HDI by dense point cloud

filtering mode and software after collapsing across other model factors. Table S14. Posterior predicted difference in basal area (BA) absolute percentage error (i.e., contrast) and 95% HDI between software by dense point cloud generation quality after collapsing across other model factors. Lower absolute percentage error values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in absolute percentage error that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is negative was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was less than zero. Table S15. Posterior predicted difference in basal area (BA) absolute percentage error (i.e., contrast) and 95% HDI between dense point cloud generation quality by software after collapsing across other model factors. Lower absolute percentage error values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in absolute percentage error that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is negative was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was less than zero. Table S16. Posterior predicted difference in basal area (BA) absolute percentage error (i.e., contrast) and 95% HDI between dense point cloud filtering mode by software after collapsing across other model factors. Lower absolute percentage error values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in absolute percentage error that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is negative was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was less than zero. Figure S3. UAS-detected tree height bias (mean error; top panel [A]) and precision (RMSE; bottom panel [B]) by software platforms (rows), dense point cloud quality (columns), and filtering mode (x-axis) settings of the SfM algorithms. (A) Positive bias values (red) indicate UAS values greater than field measurements and negative bias values (blue) represent UAS values less than field values. (B) RMSE is strictly positive with higher values (darker blue), implying less precise UAS values and a lower-bound of zero representing cases where UAS values perfectly match field measurements. Each cell represents posterior predictive distribution, where the dot is the median posterior predicted value, and the vertical segment indicates the 95% highest density interval (HDI). Figure S4. UAS-detected tree DBH bias (mean error; top panel [A]) and precision (RMSE; bottom panel [B]) for software platforms (rows), dense point cloud quality (columns), and filtering mode (x-axis) settings of the SfM algorithms. (A) Positive bias values (red) indicate UAS values greater than field measurements and negative bias values (blue) represent UAS values less than field values. (B) RMSE is strictly positive, with higher values (darker blue) implying less precise UAS values and a lower-bound of zero representing cases where UAS values perfectly match field measurements. Each cell represents the posterior predictive distribution where the dot is the median posterior predicted value and the vertical segment indicates the 95% highest density interval (HDI). Figure S5. Change in UAS-derived tree height precision (RMSE) as influenced by (A) dense point cloud generation quality setting within each software, (B) dense point cloud filtering mode within each software, and (C) software within each dense point cloud generation quality setting. Each contrast shows the posterior predictive distribution of the difference in tree height RMSE (i.e., contrast) where the dot is the median posterior predicted value and the horizontal segment indicates the 95% highest density interval (HDI). Colors represent the probability of one setting or software outperforming another. Lower RMSE values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in RMSE that is negative indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. Figure S6. Change in UAS-derived tree DBH precision (RMSE) as influenced by (A) dense point cloud generation quality setting within each software, (B) dense point cloud filtering mode within each software, and (C) software within each dense point cloud generation quality setting. Each contrast shows the posterior predictive distribution of the difference in tree height RMSE (i.e., contrast) where the dot plots the median posterior predicted value and the horizontal segment indicates the 95% highest density interval (HDI). Colors represent the probability of one setting or software outperforming another. Lower RMSE values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in RMSE that is negative indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. **Supplementary Material S3 [53]:** Table S17. Posterior

predicted tree height (m) mean error (bias) and 95% HDI by dense point cloud generation quality, filtering mode, and software after collapsing across other model factors. Table S18. Posterior predicted tree height (m) RMSE (precision) and 95% HDI by dense point cloud generation quality, filtering mode, and software after collapsing across other model factors. Table S19. Posterior predicted tree height (m) RMSE (precision) and 95% HDI by dense point cloud generation quality and software after collapsing across other model factors. Table S20. Posterior predicted tree height (m) RMSE (precision) and 95% HDI by dense point cloud filtering mode and software after collapsing across other model factors. Table S21. Posterior predicted difference in tree height (m) RMSE (i.e., contrast) and 95% HDI between software by dense point cloud generation quality after collapsing across other model factors. Lower RMSE values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in RMSE that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is negative was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was less than zero. Table S22. Posterior predicted difference in tree height (m) RMSE (i.e., contrast) and 95% HDI between dense point cloud generation quality by software after collapsing across other model factors. Lower RMSE values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in RMSE that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is negative was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was less than zero. Table S23. Posterior predicted difference in tree height (m) RMSE (i.e., contrast) and 95% HDI between dense point cloud filtering mode by software after collapsing across other model factors. Lower RMSE values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in RMSE that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is negative was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was less than zero. Table S24. Posterior predicted tree DBH (cm) mean error (bias) and 95% HDI by dense point cloud generation quality, filtering mode, and software after collapsing across other model factors. Table S25. Posterior predicted tree DBH (cm) RMSE (precision) and 95% HDI by dense point cloud generation quality, filtering mode, and software after collapsing across other model factors. Table S26. Posterior predicted tree DBH (cm) RMSE (precision) and 95% HDI by dense point cloud generation quality and software after collapsing across other model factors. Table S27. Posterior predicted tree DBH (cm) RMSE (precision) and 95% HDI by dense point cloud filtering mode and software after collapsing across other model factors. Table S28. Posterior predicted difference in tree DBH (cm) RMSE (i.e., contrast) and 95% HDI between software by dense point cloud generation quality after collapsing across other model factors. Lower RMSE values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in RMSE that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is negative was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was less than zero. Table S29. Posterior predicted difference in tree DBH (cm) RMSE (i.e., contrast) and 95% HDI between dense point cloud generation quality by software after collapsing across other model factors. Lower RMSE values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in RMSE that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is negative was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was less than zero. Table S30. Posterior predicted difference in tree DBH (cm) RMSE (i.e., contrast) and 95% HDI between dense point cloud filtering mode by software after collapsing across other model factors. Lower RMSE values mean that UAS values better match field measurements, and, as such, a contrast (i.e., difference) in RMSE that is negative (difference < 0) indicates that the left-hand side (L.H.S.) setting is more accurate than the right-hand side (R.H.S.) setting of the contrast and vice versa. The probability (Pr) that the contrast is negative was calculated from the posterior predictive distribution as the proportion of the posterior distribution where the contrast was less than zero. **Supplementary Material S4 [54]:** Table S31. Bayesian *p*-values for mean and standard deviation test statistics for each of our seven models of forest inventory metrics

derived by UAS-SfM methods. Lack of fit is indicated by a value close to 0 or 1, while a value of 0.5 indicates perfect fit [37]. Figure S7. Prior distributions in our Bayesian model of the F-score metric. Broad priors were selected so that the priors have minimal influence on the posterior [39]. Figure S8. MCMC chain convergence evidence using the R-hat convergence statistic for every model parameter in our Bayesian model of the F-score metric. Values near one indicate sufficient convergence of the Markov chains [40]. Figure S9. MCMC chain resolution evidence using the effective sample size indicator for the baseline parameter and group level deflection parameters in our Bayesian model of the F-score metric where values over or near 10,000 indicate stable parameter estimates from the Markov chains [39]. Figure S10. Graphical posterior predictive check providing evidence that our Bayesian model of the F-score metric usefully mimics the data. Where the test statistic of the observed data $T(y)$, shown in dark blue, is overlaid on the test statistic from the simulated data $T(yrep)$, shown in light blue. Figure S11. Prior distributions in our Bayesian models of the basal area percentage error metric and the basal area absolute percentage error metric. Broad priors were selected so that the priors have minimal influence on the posterior [39]. Figure S12. MCMC chain convergence evidence using the R-hat convergence statistic for every model parameter in our Bayesian models of basal area percentage error and the basal area absolute percentage error metric. Values near one indicate sufficient convergence of the Markov chains [40]. Figure S13. MCMC chain resolution evidence using the effective sample size (ESS) indicator for the baseline parameter and group level deflection parameters in our Bayesian models of the basal area percentage error metric and the basal area absolute percentage error metric. Values over or near 10,000 indicate stable parameter estimates from the Markov chains [39]. Figure S14. Graphical posterior predictive check providing evidence that our Bayesian models of the basal area percentage error metric and the basal area absolute percentage error metric usefully mimic the data. Where the test statistic of the observed data $T(y)$ shown in dark blue is overlaid on the test statistic from the simulated data $T(yrep)$ shown in light blue. Figure S15. Parameter prior distributions set using the data as a proxy such that the priors have minimal influence on the posterior [39]. For each of the four models, delineated by the plot panel titles, we present the prior distribution for the intercept (β), the parameter standard deviation ($\sigma\beta_1 \dots \sigma\beta_2 \times 3$), and the overall model dispersion (σy). Figure S16. MCMC chain convergence evidence using the R-hat convergence statistic for every model parameter. Values near one indicate sufficient convergence of the Markov chains [40]. Figure S17. MCMC chain resolution evidence using the effective sample size (ESS) indicator for the baseline parameter and group level deflection parameters where values over or near 10,000 indicate stable parameter estimates from the Markov chains [39]. Figure S18. Graphical posterior predictive check providing evidence that the model usefully mimics the data. Where the test statistic of the observed data $T(y)$, shown in dark blue, is overlaid on the test statistic from the simulated data $T(yrep)$, shown in light blue.

Author Contributions: Conceptualization, W.T.T.; methodology, W.T.T. and G.A.W.; validation, W.T.T.; formal analysis, G.A.W. and W.T.T.; writing—original draft preparation, W.T.T. and G.A.W.; writing—review and editing, W.T.T. and G.A.W.; funding acquisition, W.T.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the US Department of Agriculture National Institute of Food and Agriculture (2022-67021-37857) and US Department of Agriculture, Forest Service with funding from the Bipartisan Infrastructure Law. This research was supported by the USDA Forest Service, Rocky Mountain Research Station. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or US Government determination or policy.

Data Availability Statement: Data and analysis source code for this project has been published to the public domain at: https://github.com/georgewoolsey/uas_sfm_tree_detection.

Acknowledgments: We would like to thank Steven Alton and Paula Fornwalt of the USDA Forest Service's Manitou Experimental Forest who maintain one of the sites used in this study. Also, we thank Mike Battaglia, who maintains the sites at the USDA Forest Service's Black Hills Experimental Forest and Kaibab National Forests used in the study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Dainelli, R.; Toscano, P.; Filippo Di Gennaro, S.; Matese, A. Recent Advances in Unmanned Aerial Vehicle Forest Remote Sensing—A Systematic Review. *Part I A Gen. Framework. Remote Sens.* **2021**, *12*, 327.
- Jeronimo, S.M.A.; Kane, V.R.; Churchill, D.J.; McGaughey, R.J.; Franklin, J.F. Applying LiDAR Individual Tree Detection to Management of Structurally Diverse Forest Landscapes. *J. For.* **2018**, *116*, 336–346. [\[CrossRef\]](#)
- Creasy, M.B.; Tinkham, W.T.; Hoffman, C.M.; Vogeler, J.C. Potential for individual tree monitoring in ponderosa pine-dominated forests using unmanned aerial system structure from motion point clouds. *Can. J. For. Res.* **2021**, *51*, 1093–1105. [\[CrossRef\]](#)
- Freudenberger, M.; Magdon, P.; Nölke, N. Individual Tree Crown Delineation in High-Resolution Remote Sensing Images Based on U-Net. *Neural Comput. Appl.* **2022**, *34*, 22197–22207. [\[CrossRef\]](#)
- Zhou, X.; Zhang, X. Individual tree parameters estimation for plantation forests based on UAV oblique photography. *IEEE Access* **2020**, *8*, 96184–96198. [\[CrossRef\]](#)
- Fraser, B.; Congalton, R.G. Monitoring fine-scale forest health using unmanned aerial systems (UAS) multispectral models. *Remote Sens.* **2021**, *13*, 4873. [\[CrossRef\]](#)
- Lad, L.E.; Tinkham, W.T.; Sparks, A.M.; Smith, A.M.S. Predictive Models of Tree Foliar Moisture Content using Multispectral UAS Data: A Laboratory Study. *Remote Sens.* **2023**, *15*, 5703. [\[CrossRef\]](#)
- Belmonte, A.; Sankey, T.; Biederman, J.A.; Bradford, J.; Goetz, S.J.; Kolb, T.; Woolley, T. UAV-derived estimates of forest structure to inform ponderosa pine forest restoration. *Remote Sens.* **2020**, *6*, 181–197. [\[CrossRef\]](#)
- Hanna, L.; Tinkham, W.T.; Battaglia, M.A.; Vogler, J.C.; Ritter, S.M.; Hoffman, C.M. Characterizing Heterogeneous Forest Structure in Ponderosa Pine Forests via UAS-Derived Structure from Motion. *Environ. Monit. Assess.* **2024**, *196*, 530. [\[CrossRef\]](#)
- Iglhaut, J.; Cabo, C.; Puliti, S.; Piermattei, L.; O'Connor, J.; Rosette, J. Structure from Motion Photogrammetry in Forestry: A Review. *Curr. For. Rep.* **2019**, *5*, 155–168. [\[CrossRef\]](#)
- Young, D.J.N.; Koontz, M.J.; Weeks, J.M. Optimizing Aerial Imagery Collection and Processing Parameters for Drone-Based Individual Tree Mapping in Structurally Complex Conifer Forests. *Methods Ecol. Evol.* **2021**, *13*, 1447–1463. [\[CrossRef\]](#)
- Swayze, N.C.; Tinkham, W.T.; Vogeler, J.C.; Hudak, A.T. Influence of flight parameters on UAS-based monitoring of tree height, diameter, and density. *Remote Sens. Environ.* **2021**, *263*, 112540. [\[CrossRef\]](#)
- Swayze, N.P.; Tinkham, W.T.; Creasy, M.B.; Vogeler, J.C.; Hudak, A.T.; Hoffman, C.M. Influence of UAS flight altitude and speed on aboveground biomass prediction. *Remote Sens.* **2022**, *14*, 1989. [\[CrossRef\]](#)
- Pell, T.; Li, J.Y.Q.; Joyce, K.E. Demystifying the differences between structure-from-motion software packages for pre-processing drone data. *Drones* **2022**, *6*, 24. [\[CrossRef\]](#)
- Tinkham, W.T.; Swayze, N.C. Influence of Agisoft Metashape parameters on individual tree detection using structure from motion canopy height models. *Forests* **2021**, *12*, 250. [\[CrossRef\]](#)
- Peet, R.K. Forest vegetation of the Colorado Front Range. *Plant Ecol.* **1981**, *45*, 3–75. [\[CrossRef\]](#)
- Addington, R.N.; Aplet, G.H.; Battaglia, M.A.; Briggs, J.S.; Brown, P.M.; Cheng, A.S.; Dickinson, Y.; Feinstein, J.A.; Pelz, K.A.; Regan, C.M.; et al. *Principles and Practices for the Restoration of Ponderosa Pine and Dry Mixed-Conifer Forests of the Colorado Front Range*; United States Department of Agriculture/Forest Service Rocky Mountain Research Station: Fort Collins, CO, USA, 2018.
- Larson, A.J.; Churchill, D.C. Tree spatial patterns in fire-frequent forests of western North America, including mechanisms of pattern formation and implications for designing fuel reduction and restoration treatments. *For. Econ. Manag.* **2012**, *267*, 74–92. [\[CrossRef\]](#)
- Tinkham, W.T.; Dickinson, Y.; Hoffman, C.M.; Battaglia, M.A.; Ex, S.; Underhill, J. *Visualization guide to Heterogeneous Forest Structures Following Treatment in the Southern Rocky Mountains*; United States Department of Agriculture/Forest Service Rocky Mountain Research Station: Fort Collins, CO, USA, 2017.
- Allen, S.R.; Savage, M.; Falk, D.A.; Suckling, K.F.; Swetnam, T.W.; Shulke, T.; Stacey, P.B.; Morgan, P.; Hoffman, M.; Klingel, J.T. Ecological restoration of southwestern ponderosa pine ecosystems: A broad perspective. *Econ. Appl.* **2002**, *12*, 1418–1433. [\[CrossRef\]](#)
- Covington, W.W.; Moore, M.M. Southwestern ponderosa forest structure: Changes since Euro-American settlement. *J. For.* **1994**, *92*, 39–47. [\[CrossRef\]](#)
- Agee, J.K.; Skinner, C.N. Basic principles of forest fuel reduction treatments. *For. Econ. Manag.* **2005**, *211*, 83–96. [\[CrossRef\]](#)
- Reynolds, R.T.; Sánchez Meador, A.J.; Youtz, J.A.; Nicolet, T.; Matonis, M.S.; Jackson, P.L.; DeLorenzo, D.G.; Graves, A.D. *Restoring Composition and Structure in Southwestern Frequent-Fire Forests: A Science-Based Framework for Improving Ecosystem Resiliency*; United States Department of Agriculture/Forest Service Rocky Mountain Research Station: Fort Collins, CO, USA, 2013.
- USGS. *GAP/LANDFIRE National Terrestrial Ecosystems 2011*; Version 3; U.S. Geological Survey, Gap Analysis Program: Boise, ID, USA, 2016. [\[CrossRef\]](#)
- Boyden, S.; Binkley, D.; Shepherd, W. Spatial and temporal patterns in structure, regeneration, and mortality of an old-growth ponderosa pine forest in the Colorado Front Range. *For. Ecol. Manag.* **2005**, *219*, 43–55. [\[CrossRef\]](#)
- R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: <https://www.R-project.org/> (accessed on 7 May 2024).
- Roussel, J.R. *lasR: Fast and Pipeable Airborne LiDAR Data Tools*. R Package Version 0.5.3. 2024. Available online: <https://github.com/r-lidar/lasR> (accessed on 7 May 2024).

28. Zhang, W.; Qi, J.; Wan, P.; Wang, H.; Xie, D.; Wang, X.; Yan, G. An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. *Remote Sens.* **2016**, *8*, 501. [CrossRef]
29. Roussel, J.; Qi, J. RCSF: Airborne LiDAR Filtering Method Based on Cloth Simulation. R Package Version 1.0.2. 2020. Available online: <https://CRAN.R-project.org/package=RCSF> (accessed on 7 May 2024).
30. St-Onge, B. Methods for improving the quality of a true orthomosaic of Vexcel UltraCam images created using a lidar digital surface model. In Proceedings of the Silvilaser 2008, Edinburgh, UK, 17–19 September 2008; pp. 555–562.
31. Popescu, S.; Wynne, R. Seeing the Trees in the Forest: Using Lidar and Multispectral Data Fusion with Local Filtering and Variable Window Size for Estimating Tree Height. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 589–604. [CrossRef]
32. Roussel, J.R.; Auty, D.; Coops, N.C.; Tompalski, P.; Goodbody, T.R.H.; Sánchez Meador, A.; Bourdon, J.F.; De Boissieu, F.; Achim, A. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sens. Environ.* **2020**, *251*, 112061. [CrossRef]
33. Tinkham, W.T.; Mahoney, P.R.; Hudak, A.T.; Domke, G.M.; Falkowski, M.J.; Woodall, C.W.; Smith, A.M.S. Applications of the United States Forest Service Forest Inventory and Analysis dataset: A review and future directions. *Can. J. For. Res.* **2018**, *48*, 1251–1268. [CrossRef]
34. Kane, V.R.; Bartl-Geller, B.N.; Cova, G.R.; Chamberlain, C.P.; van Wagtenonk, L.; North, M.P. Where are the large trees? A census of Sierra Nevada large trees to determine their frequency and spatial distribution across three large landscapes. *For. Ecol. Manag.* **2023**, *546*, 121351. [CrossRef]
35. Swayze, N.C.; Tinkham, W.T. Application of unmanned aerial system structure from motion point cloud detected heights and stem diameters to model missing stem diameters. *MethodsX* **2022**, *9*, 101729. [CrossRef]
36. Paul-Christian, B. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **2017**, *80*, 1–28.
37. Hobbs, N.T.; Hooten, M.B. *Bayesian Models: A Statistical Primer for Ecologists*; Princeton University Press: Princeton, NJ, USA, 2015.
38. Riley, K.L.; Grenfell, I.C.; Finney, M.A.; Shaw, J.D. *TreeMap 2016: A Tree-Level Model of the Forests of the Conterminous United States Circa 2016*; Forest Service Research Data Archive: Fort Collins, CO, USA, 2021. [CrossRef]
39. Kruschke, J.K. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2015.
40. Kurz, A.S. Doing Bayesian Data Analysis in BRMS and the Tidyverse [Version 1.1.0]. 2023. Available online: <https://bookdown.org/content/3686/> (accessed on 20 July 2024).
41. Brooks, S.P.; Gelman, A. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **1998**, *7*, 434–455. [CrossRef]
42. Kruschke, J.K. Bayesian analysis reporting guidelines. *Nat. Hum. Behav.* **2021**, *5*, 1282–1291. [CrossRef]
43. Kay, M. tidybayes: Tidy Data and Geoms for Bayesian Models. R Package Version 3.0.6. 2023. Available online: <https://zenodo.org/records/13770114> (accessed on 8 October 2024).
44. Lisein, J.; Pierrot-Deseilligny, M.; Bonnet, S.; Lejeune, P. A photogrammetric workflow for the creation of a forest canopy height model from small unmanned aerial system imagery. *Forests* **2013**, *4*, 922. [CrossRef]
45. Cao, L.; Liu, H.; Fu, X.; Zhang, Z.; Shen, X.; Ruan, H. Comparison of UAV LiDAR and digital aerial photogrammetry point clouds for estimating forest structural attributes in subtropical planted forests. *Forests* **2019**, *10*, 145. [CrossRef]
46. Puliti, S.; Ørka, H.O.; Gobakken, T.; Næsset, E. Inventory of small forest areas using an unmanned aerial system. *Remote Sens.* **2015**, *7*, 9632–9654. [CrossRef]
47. Tuominen, S.; Balazs, A.; Saari, H.; Pölönen, I.; Sarkeala, J.; Viitala, R. Unmanned aerial system imagery and photogrammetric canopy height data in area-based estimation of forest variables. *Silva Fennica* **2015**, *49*, 1348. [CrossRef]
48. Bonnet, S.; Lisein, J.; Lejeune, P. Comparison of UAS photogrammetric products for tree detection and characterization of coniferous stands. *Int. J. Remote Sens.* **2017**, *38*, 5310–5337. [CrossRef]
49. USDA Forest Service. FS Veg Common Stand Exam Users Guide; Chapter 2: Preparation and Design. Version 2.12.6. 2015. Available online: <https://www.fs.usda.gov/nrm/fsveg/index.shtml>. (accessed on 14 May 2024).
50. Fawcett, D.; Azlan, B.; Hill, T.C.; Kho, L.K.; Bennie, J.; Anderson, K. Unmanned aerial vehicle (UAV) derived structure-from-motion photogrammetry point clouds for oil palm (*Elaeis guineensis*) canopy segmentation and height estimation. *Int. J. Remote Sens.* **2019**, *40*, 7538–7560. [CrossRef]
51. Sparks, A.M.; Corrao, M.V.; Keefe, R.F.; Armstrong, R.; Smith, A.M.S. An accuracy assessment of field and airborne laser scanning-derived individual tree inventories using felled tree measurements and log scaling data in a mixed conifer forest. *For. Sci.* **2024**, *70*, 228–241. [CrossRef]
52. Tinkham, W.T.; Swayze, N.C.; Hoffman, C.M.; Lad, L.E.; Battaglia, M.A. Modeling the missing DBHs: Influence of model form on UAV DBH characterization. *Forests* **2022**, *13*, 2077. [CrossRef]
53. Vastaranta, M.; Melkas, T.; Holopainen, M.; Kaartinen, H.; Hyypä, J.; Hyypä, H. Laser-based field measurements in tree-level forest data acquisition. *Photogramm. J. Finl.* **2009**, *21*, 51–61.
54. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehhtari, A.; Rubin, D.B. *Bayesian Data Analysis*; Chapman and Hall/CRC: London, UK, 2013.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.