

Drone data for decision making in regeneration forests: from raw data to actionable insights¹

Stefano Puliti and Aksel Granhus

Abstract: In this study, we aim at developing ways to directly translate raw drone data into actionable insights, thus enabling us to make management decisions directly from drone data. Drone photogrammetric data and data analytics were used to model stand-level immediate tending need and cost in regeneration forests. Field reference data were used to train and validate a logistic model for the binary classification of immediate tending need and a multiple linear regression model to predict the cost to perform the tending operation. The performance of the models derived from drone data was compared to models utilizing the following alternative data sources: airborne laser scanning data (ALS), prior information from forest management plans (Prior) and the combination of drone + Prior and ALS + Prior. The use of drone data and prior information outperformed the remaining alternatives in terms of classification of tending needs, whereas drone data alone resulted in the most accurate cost models. Our results are encouraging for further use of drones in the operational management of regeneration forests and show that drone data and data analytics are useful for deriving actionable insights.

Key words: UAV, DAP, forest inventory, photogrammetry, precommercial thinning, airborne laser scanning.

Résumé : Dans cette étude, nous cherchons à développer des moyens d'utiliser directement les données brutes de drones en tant qu'informations exploitables, nous permettant ainsi de prendre des décisions de gestion directement à partir de données provenant de drones. Les données photogrammétriques et l'analyse de données de drones ont été utilisées dans le but de modéliser les besoins immédiats et les coûts de régénération des forêts. Les données de référence sur le terrain ont été utilisées pour former et valider un modèle logistique pour la classification binaire du besoin en soins sylvicoles immédiat et un modèle de régression linéaire multiple pour prédire le coût d'exécution de soins sylvicoles. La performance des modèles dérivés des données de drones a été comparée à celle des modèles utilisant diverses sources de données comme suit : données de balayage laser aéroporté (BLA), données antérieures de plans de gestion forestière (« Prior »), et la combinaison de drone + Prior et de drone + BLA. L'utilisation de données de drones et de données antérieures a surpassé les autres diverses solutions en termes de classification des soins sylvicoles, tandis que les données de drones à elles seules ont permis d'obtenir les modèles de coûts les plus précis. Nos résultats sont encourageants en matière d'utilisation ultérieure de drones dans la gestion opérationnelle des forêts de régénération et montrent que les données de drones et l'analyse des données sont utiles pour obtenir des informations exploitables. [Traduit par la Rédaction]

Received 22 September 2020. Accepted 24 November 2020.

S. Puliti and A. Granhus. Division of Forest and Forest Resources, National Forest Inventory, Norwegian Institute for Bioeconomy Research (NIBIO), Høgskoleveien 8, 1433 Ås, Norway.

Corresponding author: Stefano Puliti (e-mail: stefano.puliti@nibio.no).

Footnote: This article is part of a virtual issue on innovative applications of unmanned aerial vehicles in forest science and management jointly published by the *Journal of Unmanned Vehicle Systems* and the *Canadian Journal of Forest Research*.

This work is licensed under a [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Mots-clés : véhicule aérien sans pilote (UAV), photogrammétrie aérienne numérique (« DAP »), inventaire forestier, photogrammétrie, éclaircie précommerciale, balayage laser aéroporté (BLA).

Introduction

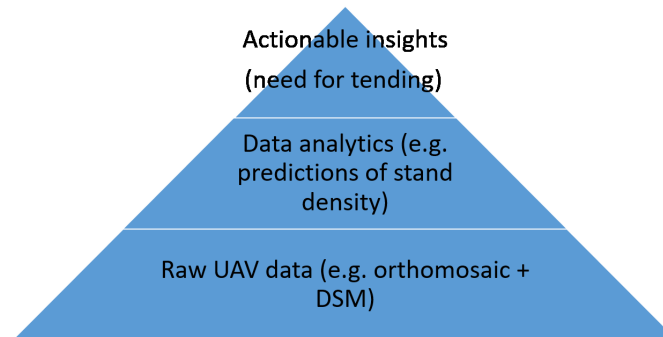
Drone data have proven valuable in modeling and providing accurate data analytics on key forest biophysical variables (Puliti et al. 2015; Guerra-Hernández et al. 2017; Iglhaut et al. 2019; Mulverhill et al. 2020). However, there is still a gap between the quantitative knowledge of forest biophysical variables and the possibility to make silvicultural decisions. This gap reflects an increasing need for evaluating the quality of drone data based on the quality of the decisions that these data enable rather than only on the accuracy of predictions of biophysical variables. At the core of any operational future use of drones in forestry lies the advantage of reducing the human input in gathering information for decision-making. Thus, drone data must provide actionable insights or information that allows decision-making without the need for performing a costly field visit. Therefore, there is a need to develop methods to summarize raw drone data (e.g., an orthomosaic or a digital surface model) into data analytics (e.g., predictions of tree density) and finally in actionable insights intended as a decision on whether to perform stand-specific silvicultural treatments (Fig. 1).

Amongst the many drone applications in forestry, their use for regeneration forest surveys has been suggested as a promising niche application (Goodbody et al. 2017). Puliti et al. (2019) demonstrated that using drones in regeneration forest surveys could reduce the costs compared with field surveys by half while maintaining high accuracy levels. In addition to being costly, field surveys in regeneration forests are also potentially biased due to the difficult navigation and poor visibility in the dense vegetation. Thus, management decisions may vary based on the path that the surveyor walks and terrain-related variables (e.g., slope and ruggedness). On the contrary, drone surveys provide a birds-eye view of the forest stand of interest, potentially allowing for more informed, comprehensive, and objective decisions. To date, an increasing body of literature has demonstrated that drone data can be used to accurately model regeneration forest biophysical variables and produce drone data analytics in the form of maps of tree density, canopy height, and tree species (Goodbody et al. 2017; Feduck et al. 2018; Fromm et al. 2019; Imangholiloo et al. 2019; Puliti et al. 2019; Castilla et al. 2020; Pearse et al. 2020). Although the results of these were encouraging for the production of accurate data analytics, the possibility to make decisions based on these data remains unknown.

An important decision to be made for regeneration forest stands concerns whether precommercial thinning, hereafter referred to as “tending” for simplicity, is needed or not. Such treatment aims to facilitate the growth of the most commercially valuable trees in the stand by removing undesired stems that compete with the future crop trees for light, water, and nutrients.

Tending consists of reducing the tree density in areas with high competition to facilitate the growth of the most commercially valuable tree species. In boreal forests in Norway, such tending consists of the removal of deciduous species in favor of the more economically valuable coniferous species such as Norway spruce (*Picea abies* (L.) Karst) and Scots pine (*Pinus sylvestris* L.), while at the same time considering an even spacing of the trees that will make up the future stand. After clear-cutting, fast-growing deciduous species like birches (*Betula* spp.), aspen (*Populus tremula* L.), willow (*Salix caprea* L.), and alder (*Alnus* spp.) often establish dense thickets, especially in fertile sites, that may severely retard the growth of the planted coniferous saplings (Braathe 1988; Brække and Granhus 2001). In naturally

Fig. 1. Schematic representation of the hierarchical structure of information obtainable from drone.



regenerated stands of low to medium site fertility, the deciduous species' competition is often less severe. Still, a tending is often needed to obtain the desired density of future crop trees and to create an even spacing in the stand. In this respect, tending operations are critical to maintaining future timber supply and assortment quality (Korhonen et al. 2013). Despite the long tradition of airborne laser scanning (ALS)-based forest inventories in Scandinavian countries (Næsset et al. 2004; Kangas et al. 2018), the accuracy obtainable for predicting regeneration forest biophysical variables (e.g., tree density) or tending need has been unsatisfactory for operational application (Korpela et al. 2008; Korhonen et al. 2013; Ørka et al. 2016). In Scandinavia, ALS forest management inventories are carried out infrequently at intervals of 10–20 years. Hence, because of the rapid growth in regeneration forests, the information is rapidly outdated and thus unreliable for decision-making. In Norway, forest managers often have access to prior inventory information derived either from aerial photo interpretation or from field visits. The use of prior information is known to be beneficial for forest inventories based on remotely sensed data (Kangas et al. 2020). However, there is no specific understanding of the importance of these data for decision-making in the regeneration forest.

During the 10 years 2010–2019, the annual tended area of forest land in Norway was on average 27 000 ha (Statistics Norway 2020). The current operational standard includes field visits to each of the stands to assess the tending needs and costs through eyeball estimates of stand density, species composition, and canopy height. The considerable cost savings and increase in predictive accuracy when using drones in regeneration forests compared to alternative remotely sensed data (Puliti et al. 2019) could provide forest managers with an attractive alternative solution to field surveys. Nevertheless, drones' future operational use for assessing the tending needs relies on the quality of the decisions that drone data enable.

This study aimed at using drone photogrammetric data and derived data analytics to predict the two variables that represent the current information need for managing regeneration forests, i.e., stand-level tending needs and costs. To contextualize the study in the current information panorama for regeneration forests, we tested five different scenarios in terms of auxiliary data availability: drone, ALS, prior information from forest management plans (Prior), and their combination (i.e., drone + Prior and ALS + Prior).

Materials

Study area

The study area is located within the 35 000 ha property of Romeal and Stange Almenninger (RASA), in the Stange municipality in southeastern Norway (approximately

60.63°N; 11.41°E). The terrain is hilly with altitudes ranging from 250 m above sea level (a.s.l.) to 600 m a.s.l., and the climate is moderately continental. The forests in this area are dominated by Norway spruce and Scots pine and are actively managed for timber production. Soils are mostly podzolic, and the majority of the stands are of intermediate site productivity, except for some areas in the lowermost parts of the property where higher site indices occur.

Field data

According to operational practices, we initially selected all forest stands that required information on the need for tending, i.e., from a list of stands belonging to maturity class 2 (see definitions in [Breidenbach et al. 2020](#)) provided by RASA. Out of the entire set of stands, we purposively sampled 27 stands to cover a range of site productivity, species composition, and stand size. For further details, we refer the reader to the study by [Puliti et al. \(2019\)](#), where the same data as in this study was used.

According to current operational standards, the local forest owner's association collected the field data, consisting of visiting each stand and subjectively assessing the need for tending. For each stand, the variables recorded consisted of eyeball estimates of tree density (n trees ha^{-1}), mean height (m), and the number of future crop trees (n trees ha^{-1}). Based on these values and forestry personnel's expertise from RASA, a decision was made on whether a tending operation was immediately needed. Out of 27 stands, six were identified as having an immediate tending need. Furthermore, the costs (NOK ha^{-1}) for performing the tending operation were estimated according to the forest owners' association's operational conditions. The costs were converted to Euros based on the exchange rate at the experiment's time (1 EUR = 9.5 NOK). On average, the estimated cost for stands with immediate tending needs was 186 EUR ha^{-1} . For some stands, even though they were not classified as having immediate tending needs, a cost estimate was still provided. The average estimated cost when including these stands was 94 EUR ha^{-1} . The variables mentioned above summarize the current information needs in Norway's operational forestry for regeneration forests' decision-making. Furthermore, the time required to perform each field check was recorded and subdivided into (i) time to approach the stand from the roadside (total of 5.1 h) and (ii) time required to assess the described variables for the stand (total of 10 h).

Although this study does not directly use field plot data, it is important to keep in mind that field plot data were used to fit the models by [Puliti et al. \(2019\)](#). To avoid confusion, we refer the reader to [Puliti et al. \(2019\)](#) for further details on the field plot data used to model the regeneration forest biophysical variables.

Prior inventory information

Information from previous inventories herein referred to as prior is frequently available in areas characterized by forest management tradition. In Norway, such information is obtained either through manual photo interpretation or updates based on field visits. The former is an integral part of current operational forest management inventories ([Næsset et al. 2004](#)) and is used to stratify the area according to different forest developmental stages, tree species, and site indices. Forest management inventories are typically updated at intervals of around 10–15 years, and for selected stands, the information can be updated by the local forest owners upon field visits. Thus, it is clear that these data are characterized by a degree of subjectivity, and concerning their applicability for decision making in regeneration forests, the data quality can be highly heterogeneous.

The explanatory variables from the prior information (see [Table 1](#)) included information on site index, tree species composition, volume, tree density, mean height, and indications for planned future harvest activities.

Table 1. Description of the variables available in the prior inventory information.

Variable	Description	Range values
SI_{h40}	Height site index or height (m) at 40 years of age	6, 8, 11, 14, 17, 20, 23
$SI_{species}$	Site index according to tree species (spruce, pine, and deciduous)	(1–2)
AGE	Age of the forest (years)	10–20
V_{HA}	Timber volume per hectare ($m^3 ha^{-1}$)	(0–140)
$V_{SPRUCE\%}$	% of spruce volume	(0–80)
$V_{PINE\%}$	% of pine volume	(0–100)
$V_{DECIDUOUS\%}$	% of deciduous volume	(0–70)
N_{before}	Tree density before tending ($n ha^{-1}$)	(1300–5000)
$N\%_{crop}$	% of the N_{before} remaining as crop trees	—
$N\%_{PINE_{crop}}$	% number of future-crop pine trees	(0–100)
$N\%_{DECIDUOUS_{crop}}$	% tree density of future-crop deciduous trees	(0–20)
$N\%_{SPRUCE_{before}}$	% tree density spruce before tending	(0–100)
$N\%_{PINE_{before}}$	% tree density pine before tending	(0–70)
$N\%_{DECIDUOUS_{before}}$	% tree density deciduous before tending	(0–50)
$H_{conifer}$	Mean height coniferous (m)	(0.5–7)
$H_{deciduous}$	Mean height deciduous (m)	(0–6)
POT_PROD	Stand potential productivity (m^3)	(0–52)
HARVEST_TYPE	Planned logging form (full, partial, or no harvest)	1, 2, 3
YEAR_HARVEST	Year of harvest maturity (year)	(2086–2116)

Drone data

We acquired the drone data during August 2018, using a DJI Phantom 4 PRO (DJI 2018) integrated with a 1 in. 20 megapixels red, green, and blue (RGB) sensor with a mechanical shutter and focal length of 8.8 mm. As for the field data, we recorded the time required to walk from the roadside to the drone take-off and landing point (total of 2.5 h) and to execute the flight (total of 5.3 h). The drone images were processed using Agisoft Photoscan version 1.4.1 (Agisoft 2018), and we generated three raw data products for each stand: (i) an orthomosaic with a pixel size of 3 cm \times 3 cm, (ii) a point cloud with a point density of approximately 300 points m^{-2} classified into ground and nonground points, and (iii) a digital surface model (DSM) with a pixel size of 6 cm \times 6 cm.

The entire stand area was tessellated into 50 m^2 square grid cells, and the 92 drone predictor variables described by Puliti et al. (2019) were computed for the area corresponding to each field plot and each grid cell. The drone predictor variables were grouped into three main categories: (i) spectral variables (i.e., RGB band values, band ratios, and textural variables), (ii) DTM-independent variables (DSM summary statistics, topographic variables, and textural variables), and (iii) normalized point cloud variables (height percentiles, density variables, and local maxima counts). In addition to these variables, we now included an additional one, Hartigan's dip statistic, as preliminary tests suggested this is an important variable for regeneration forests. This statistic is obtained by performing the Hartigan's dip test for unimodality (Hartigan and Hartigan 1985) on the points height's distribution within a plot or a grid cell. A value of one was assigned to each grid cell for which the null hypothesis was confirmed (i.e., unimodal distribution) and a value of two on the contrary case (i.e., multi-modal distribution). The test was performed using the *dip*test R package (Maechler 2016). The rationale behind this variable is that it is meant to discriminate between uniformly distributed areas (i.e., open ground or uniform forest cover) and more heterogeneous forest structures.

The drone-derived predictor variables were then used by Puliti et al. (2019) to fit random forest models for mean tree height (H_m ; m) and tree density (N ; trees ha^{-1}) using 580 plots of 50 m^2 each with in-situ measurements of the dependent variables. A summary of the models' performance in terms of predictive accuracy (i.e., root-mean square error; RMSE) and systematic errors (mean difference; MD) is presented in Table 2. A thorough evaluation of

Table 2. Summary of the models devised by Puliti et al. (2019) for mean height (\widehat{H}_m) and tree density (\widehat{N}) in terms of predictive accuracy (RMSE and RMSE%) and presence of systematic errors (MD and MD%) and using either drone or airborne laser scanning (ALS) data as auxiliary data.

Predicted variable	Auxiliary data	RMSE	RMSE%	MD	MD%
\widehat{H}_m (m)	Drone	0.56	23.6	0.1	3.7
	ALS	0.76	31.6	0.3	12.5
\widehat{N} (n trees ha ⁻¹)	Drone	1175	21.8	-62	-1.1
	ALS	2355	43.7	-481.5	-8.9

the models and the drone predictor variables is provided by Puliti et al. (2019). According to an area-based approach, the models for H_m and N were applied to each grid cell to predict H_m and N . The grid-level drone predictor variables, as well as the data analytics in the form of model predictions, were then aggregated at stand-level by calculating the arithmetic mean and the standard deviation for each of the variables within all of the grids included in a stand.

ALS data

Airborne laser scanning (ALS) data with an average point density of 5 points m⁻² were collected in June 2016 using an Optech ALTM Titan MV. In this study, ALS data were used compared with current state-of-the-art drone results. Similar to the drone data, a set of explanatory ALS variables were used together with the field plot data to fit random forest models for H_m and N . The model's performances are reported in Table 2. The models were then applied to the population of grid cells with ALS explanatory variables in the same way as with the models derived with drone data. The grid-cell explanatory variables and the predicted H_m and N values were then aggregated to the stand-level by computing the mean and the standard deviation of each variable.

Methods

The statistical analysis consisted of modeling and validating the tending need and costs using a logistic and regression model. The adoption of parametric approaches was deemed suitable due to the lack of an extensive training data set and because previous experience has shown logistic models to be suitable for modeling tending needs (Korhonen et al. 2013). The models for tending need and costs were fitted separately for the different scenarios in terms of data availability: drone, ALS, drone + Prior, ALS + Prior, and Prior.

Tending need

The logistic model was formulated as

$$(1) \quad \pi = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}$$

where π is the probability of the need for tending, x_1 to x_n are the predictor variables, β_0 is the intercept, and β_1 to β_n are parameters to be estimated or regression coefficients.

The explanatory variables were selected using the variable importance in terms of mean decreased Gini coefficient reported by the Random Forests model (Breiman 2001; Liaw and Wiener 2002). Such an ad-hoc method was implemented after a preliminary analysis had revealed that it resulted in the selection of the most accurate models compared with alternative variable selection methods (e.g., step-wise or branch-and-bound selection). To account for randomization in the random forests, we selected the three most important

variables over 200 iterations. Furthermore, we tested for the presence of multicollinearity by only accepting models that had a variance inflation factor <5.

The logistic model was validated through a leave-one-out cross-validation. Each stand was removed iteratively from the training data pool for a newly fitted logistic regression and used only to validate the model's predictions. The classifier's performance was evaluated using the summary information from the confusion matrices consisting of overall accuracy, and the class-wise user's accuracy, and producer's accuracy confusion matrix.

Tending cost

To model the tending cost, we opted for a multiple linear regression model. The regression model was formulated as

$$(2) \quad Y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon$$

where Y is the stand tending cost, and ε is the residual error.

The variable selection consisted of a branch-and-bound search for the best subset and model selection based on the Bayesian information criterion (Puliti et al. 2015). Furthermore, to ensure the absence of multicollinearity, the models with a variance inflation factor >5 were discarded. The models' performances were evaluated in terms of model fit (Adj. R^2), predictive accuracy (RMSE), and the presence of systematic errors (MD), and their respective values were compared with the average field-estimated tending cost for stands in need of immediate tending (RMSE% and MD%). The Adj. R^2 , RMSE, RMSE%, MD, and MD% were calculated.

Results and discussion

Tending need

The overall comparison between the selected drone and ALS variables for classifying tending needs revealed that, in general, drone variables were more strongly correlated (Pearson's correlation coefficient in the range 0.38–0.59) with the tending need and cost (Fig. 2) than ALS (Pearson's correlation coefficient 0.32–0.54) or prior inventory information (Pearson's correlation coefficient –0.33 to 0.57). The stronger explanatory power of drone variables was further confirmed by the fact that the drone model required one variable less than the ALS or the Prior ones. When including the information for prior inventory information, the N_{before} was included in both the drone + Prior and ALS + Prior, indicating the importance of the number of stems for discriminating the stands according to the tending needs. Interestingly, variables derived from the predictions of biophysical variables such as the tree density (\hat{N}) and the future crop tree density (\hat{N}_c) were selected in three out of the four fitted models, both for drone and ALS data. This finding highlights the important link between drone data analytics (maps) and actionable insights (management decisions). The importance of tree density information for determining tending needs was further strengthened by (i) the selection of N_{before} in all of the models including prior inventory information and (ii) by the fact that height-related variables were discarded from all of the tending need models, indicating that in such drone data-driven approach, the tree density plays a more important role in the determining the need and cost of tending.

For the drone models, the selection of a textural variable computed using the DSM further confirmed the usefulness of digital terrain model (DTM)-independent variable for describing forest biophysical properties (Giannetti et al. 2018; Puliti et al. 2020). Specifically, the $dsmENTR$ is a measure of the disorder in an image and achieves its largest value when all pixels in a DSM have a uniform height. Thus, low values correspond to open and sparse canopy structures and high values to uniform thickets of vegetation where the competition needs to be regulated (Fig. 2). Concerning the ALS model, in addition to the

Fig. 2. Pearson's correlation coefficient plots for the selected variables for drone or airborne laser scanning (ALS) data for the logistic model for classifying the tending need. The ellipses describe the strength and direction of the correlation. The variables included, textural variables ($\widehat{dsmENTR}_{mean}$), predictions based on previously existing models for number of trees (\widehat{N}_{sd}), number of future crop trees (\widehat{N}_{cSD}), the standard deviation of the ratios between the blue and red bands (B/R_{sd} mean), the mean or standard deviation of the Hartigan's dip statistics (Dip_{mean} and sd) as well as variables from previous inventories (Prior) such as the Stand potential productivity (POT_PROD), stand density (N) spruce percentage of tree density (N_{before}) and the proportion of pine trees in the future crop trees ($N\%_{PINEafter}$).

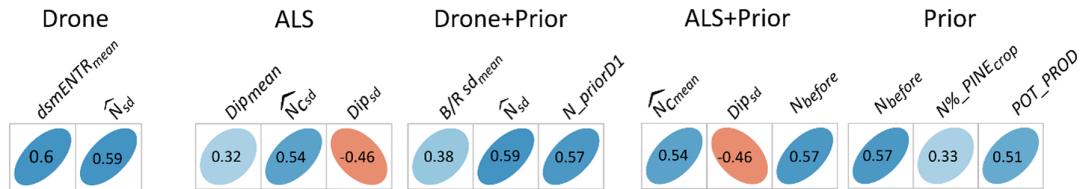


Table 3. Summary of the accuracy assessment for the logistic classification of the need for tending and regression modeling for the tending cost.

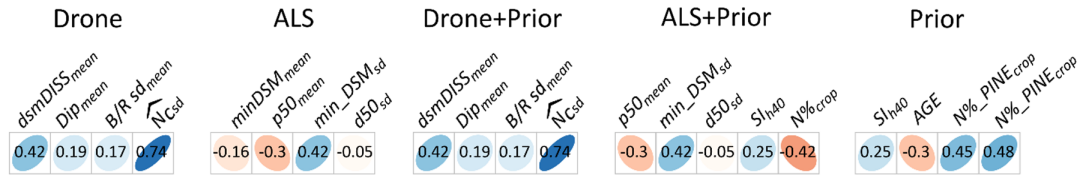
	Overall accuracy	Producer's accuracy (100-omission error)		User's accuracy (100-commission error)	
		No tending	Tending	No tending	Tending
Drone	84	90	67	90	67
Drone + Prior	89	90	83	95	71
ALS	81	90	50	86	60
ALS + Prior	81	85	67	90	43
Prior	85	95	50	83	75

Note: ALS, airborne laser scanning; Prior, prior information from forest management plans.

standard deviation of the predicted number of future crop trees (\widehat{N}_{cSD}), the mean (Dip_{mean}) and standard deviation (Dip_{SD}) of the Hartigan's dip statistic were selected, indicating the importance of this newly tested predictor variable. The variables selected for the models based on prior inventory information included the tree density (N_{before}), the percentage of future crop pine trees $N\%_{PINE_{crop}}$, and the stand potential productivity of (POT_PROD).

The evaluation of the predictive accuracy for the tending need model according to a leave-one-out cross-validation (see Table 3 and confusion matrices in Appendix A) revealed a larger overall accuracy for drone (84%–89%) compared with ALS data (81%), with the former yielding a larger producer's (90%) and user's accuracies (67%–83%). When including prior information, the overall accuracy of drone data was boosted to 89% thanks to an increase in the correctly detected stands in tending need resulting in the largest producer's accuracy (83%). Among the producer's and the user's accuracy, the former has a more severe impact on forest management. The omission errors (i.e., 100% producer's accuracy) correspond to stands for which the tending may not need to be done and thus lead to an overall economic loss due to the reduction in the value of the assortments obtainable from the future stand. Although limited by a small number of sampled stands, it was encouraging seeing the contribution from the drone in reducing the number of omitted stands in tending need. The reported OA values were larger than what was previously reported by Korhonen et al. (2013), who, using ALS data, found a stand-level OA of 71%. The larger overall accuracy was due to a larger percentage of correctly classified stand with no tending need in this study (90%) compared to the study by Korhonen et al. (2013) (70%). The relatively large OA of the logistic model based on ALS was a noteworthy finding, which can partly be

Fig. 3. Pearson's correlation coefficient plots for the selected variables for drone or airborne laser scanning (ALS) data for the regression model for predicting the tending cost. The ellipses describe the strength and direction of the correlation. The variables included, textural variables ($dsmDISS_{mean}$), predictions based on previously existing models for number of trees (\hat{N} sd), number of future crop trees (\hat{N}_{c} sd), the standard deviation of the ratios between the blue and red bands (B/R sd mean), the mean or standard deviation of the Hartigan's dip statistics (Dip mean), the mean and standard deviation of the digital surface model values (min_{DSM} mean and min_{DSM} sd), the median of the normalized height values ($p50$ mean), the density of ALS points below half of the top height ($d50$ sd) as well as variables from previous inventories (Prior) such as the site index (SI_{h40}) age (AGE), proportion of crop trees ($N\%$ crop) and the proportion of pine ($N\%_{PINE}$ crop).



explained by dataset-specific characteristics and perhaps by the use of the newly proposed variable, i.e., the Hartigan's dip statistic. The increase from 50% to 67% in the producer's accuracy when complementing ALS data with prior inventory information offers an appealing methodology for operational implementation. However, it is important to keep in mind that in this study, the ALS data were only two years old, and the use of older data is likely to have negative effects on the predictive accuracy of ALS models. Because our study was conducted using a limited number of stands, future research should further investigate the use of ALS + Prior models using larger sample sizes and across broader geographical scales. The use of prior inventory information alone yielded an OA of 85%, which, even though seemingly large, was driven by a large producer's accuracy (95%) for stands not needing tending while only 50% of the stands with tending needs were correctly identified. The relatively low ability for the prior information to correctly detect stands with tending needs justifies the current need for performing field visits.

Tending cost

The variable selection for the tending cost models (Fig. 3), in general, resulted in the selection of a larger number of variables (i.e., mostly four variables) than the tending need models, and these variables were, in general, less correlated to the field assessed tending costs than the explanatory variables for the tending need. The drone and drone + Prior models were identical, meaning that drone variables were a stronger predictor of costs than any variable from the prior inventory. In particular, the standard deviation of the predicted number of future crop trees (\hat{N}_{cSD}) from drone data was strongly correlated (Pearson's correlation coefficient = 0.74) with the tending cost. Interestingly, the number of future crop trees is a key variable to quantify the cost for the tending operations in current operational conditions. Thus, the inclusion of drone predicted \hat{N}_{cSD} in the cost model, further confirms the usefulness of drone data analytics to obtain actionable insights. Other explanatory variables included in the drone cost model included the textural variables from the DSM ($dsmDISS_{mean}$), the mean Hartigan's dip statistic (Dip_{mean}), and mean of the standard deviations of the band ratios calculated using the blue and red bands (B/RSD_{mean}). The first two variables confirm the usefulness of DTM-independent variables and of Hartigan's dip statistic even for drone data. The B/RSD_{mean} represents the only spectral variable in any of the fitted models in this study, and its inclusion in the tending cost model shows the synergy between spectral and three-dimensional information. An important drawback of using spectral variables is that they can vary substantially due to varying light and

Table 4. Summary of the independent validation for the cost model for each of the study scenarios and according to adjusted R^2 (Adj. R^2), root-mean square error (RMSE), mean difference (MD), and their relative values (RMSE% and MD%).

Tested scenario	Tending cost predictions				
	Adj. R^2	RMSE (EUR ha ⁻¹)	RMSE% (%)	MD (EUR ha ⁻¹)	MD% (%)
Drone	0.86	23.4	14.9	7.8	4.9
Drone + Prior	0.86	23.4	14.9	7.8	4.9
ALS	0.69	35.5	22.6	23.5	15.0
ALS + Prior	0.86	31.4	20	16.9	10
Prior	0.49	102.7	65.4	31.4	20.0

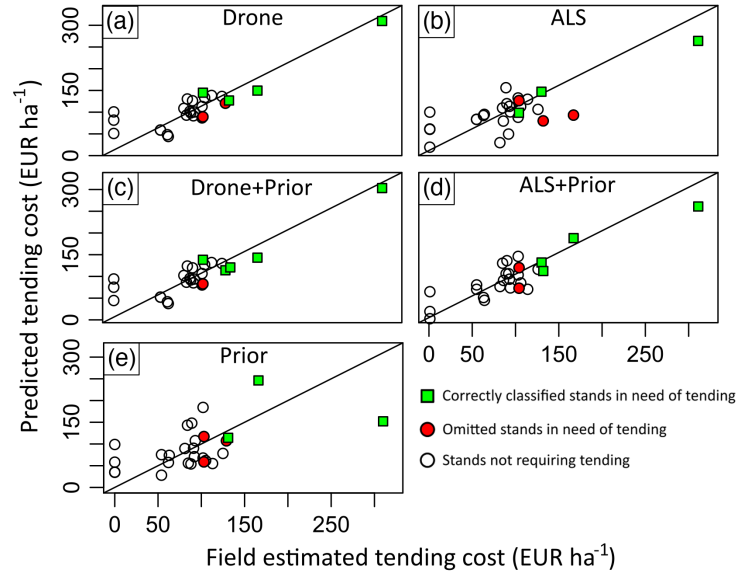
Note: ALS, airborne laser scanning; Prior, prior information from forest management plans.

atmospheric conditions, and thus their use may affect the transferability of a model to new data. In this study, the drone data were collected over several days and under varying light and atmospheric conditions; thus, the drone tending cost model can account for this variation to a certain degree. Furthermore, the selection of a band ratio was encouraging as ratios are typically more robust to variations in light conditions and topographic shading than the individual band digital number values. Further studies should, however, investigate the effect of the variation in light, atmospheric, and phenological conditions on the predictive accuracy of drone models.

The ALS cost model was characterized by the selection of the mean of two height percentiles the mean ($p0_{\text{mean}}$ and $p50_{\text{mean}}$), the standard deviation of the minimum value of the normalized point cloud ($p0_{\text{SD}}$), and a density variable ($d50_{\text{SD}}$). The minimum height ($p0$) represented the height of the lowest vegetation points and thus is useful to detect the presence of thick low vegetation and its variation through each stand. The costs increased when increasing $p0$ but decreased with its variability. The selection of the ALS height and density variables is consistent with previous literature on the use of ALS data for regeneration forests (Korhonen et al. 2013; Ørka et al. 2016) and shows that, differently from drone data, when high-quality height information is available it is indeed useful to model tending costs. When complementing ALS with prior information, the site index (SI_{h40}) and the percentage of the number of trees that are supposed to form the future population of crop trees ($N\%_{\text{crop}}$) were included in the model.

The leave-one-out validation of the tending cost model revealed that the use of drone data alone outperformed any of the other data sources, yielding the best model fit (Adj. $R^2 = 0.86$) and smallest RMSE (23.4 EUR ha⁻¹ or 15%) and MD (7.8) values (Table 4). While prior information availability did not affect the drone model, it had a substantial positive effect on the models relying on ALS data. These results are in line with the findings by Kangas et al. (2020), who found that models using both ALS explanatory variables and prior information were more accurate than using the two data sources separately. Even though the cost model using ALS + Prior had a similar model fit to the drone models, the respective residuals were not significantly different according to a paired t -test ($p = 0.6$). Furthermore, the RMSE and MD for ALS + Prior models were larger than drone data alone. This was mainly due to the underestimation of the stand's cost that had the largest field estimated costs (Fig. 4). Systematic errors can be particularly problematic in a forest enterprise as they can lead to substantial economic losses in the long term. A promising way to reduce the bias in the estimates using prior inventory information is to use growth models to update prior information to the present date (Kangas et al. 2020). However, such growth models are nonexistent in Norway; thus, it is currently not possible to update prior information on regeneration forests. Unsurprisingly, the use of prior information alone resulted

Fig. 4. Scatterplot of the field estimated cost versus the drone predicted costs. The diagonal line represents the 1:1 line. The green points represent the true positive according to the classification for tending need, while the red one represents the true negative or the stand that was wrongly classified as with no tending need.



in the poorer models, characterized by the largest RMSE (102 EUR ha⁻¹) and MD (31 EUR ha⁻¹). Such a poor ability to predict costs using prior information is one of the main reasons behind the current practice of conducting field checks. It is also true that a thorough assessment of prior inventory information can be difficult due to the large variability in terms of quality of these data, especially concerning regeneration forests. To the authors' best knowledge, this is the only study where remotely sensed data were used to estimate the costs for tending; thus, no comparison with previous literature was possible.

General considerations

Our study showed how drone data outperformed ALS data and produced results similar to the field-based estimates when coupling the results from the tending need and cost models. This study also highlighted the existence of important synergies between the drone and prior inventory data. While not included in the cost model, using prior information in the drone tending model resulted in a reduction by half of the omission errors compared with the model based on drone data alone. Thus, it is evident that drone data has the potential for operational use for the management of regeneration forests.

In agreement with the previous literature (Korhonen et al. 2013; Ørka et al. 2016), the use of ALS data alone was not promising for operational application as only 50% of the stands in tending need were correctly detected, and systematic errors characterized the cost predictions. On the other hand, our study found the combination of ALS and prior information to be an appealing alternative to drone data. Even though ALS + Prior resulted in larger omission errors than drone + Prior, the costs per hectare for acquiring ALS over large areas are substantially smaller than those for acquiring drone data. Our results for ALS + Prior models encourage further investigation of the synergy between ALS and prior information, as this may provide an efficient method for managing regeneration forests on a large scale. However, a major disadvantage is that ALS data and prior information may not be available at all or, if available, they may be outdated and thus unreliable for decision-making.

Concerning the operational application of the proposed method, we showed that the drone data acquisition's total cost was approximately 50% of the cost for the field survey (Puliti et al. 2019). This was mainly due to the drastic reduction in the time required to walk to, within, and from each stand. Further work should focus on determining whether the smaller costs associated with drone data collection outweigh the costs because of sub-optimal decisions due to the presence of small errors in drone predictions. This will provide stronger evidence on the costs and benefits of using drone data in regeneration forests.

Conclusion

This study addressed the use of drone data to provide actionable insights for the management of regeneration forests. Given the available data, the main findings of this study can be summarized as follows.

- The drone data outperformed ALS data for predicting tending need and cost.
- The study demonstrated the usefulness of drone data analytics to convert raw drone data into actionable insights.
- The study showed the importance of tree density information (i.e., included in 9/10 tested models) for management of regeneration stands.
- This is the first study where the operational costs for silvicultural treatments are modeled using remotely sensed data.

Acknowledgements

This study was carried out within the DRONE-REG project, funded by grants from the Forestry Development fund (Utviklingsfondet), the Forest Trust Fund (Skogtiltaksfondet), and the Norwegian Agriculture Agency. We are grateful to RASA, Glommen Mjøsen Skog, and the County Governor of Innlandet County for the field data collection.

References

- Agisoft. 2018. Agisoft photoscan user manual professional edition, version 1.4. Available from http://www.agisoft.com/pdf/photoscan-pro_1_4_en.pdf.
- Braathe, P. 1988. Utviklingen av gjenvekst med ulike blandingsforhold mellom bartrær og løvtrær — II. Rapport fra Norsk institutt for skogforskning 8, 1–50. Available from https://nibio.brage.unit.no/nibio-xmlui/bitstream/handle/11250/2686505/NISK_rapport_8-88.pdf?sequence=1&isAllowed=y.
- Brække, F.H., and Granhus, A. 2001. Ungskogpleie i naturlig forynget gran på middels og høy bonitet. Rapport fra skogforskningen 10, 1–24. Available from <https://nibio.brage.unit.no/nibio-xmlui/bitstream/handle/11250/2558460/Skogforsk-Rapport-2004-10.pdf?sequence=2&isAllowed=y>.
- Breidenbach, J., Granhus, A., Hysten, G., Eriksen, R., and Astrup, R. 2020. A century of National Forest Inventory in Norway — Informing past, present, and future decisions. *Forest Ecosyst.* 7: 46.
- Breiman, L. 2001. Random forests. *Mach. Learn.* 45: 5–32.
- Castilla, G., Filiatrault, M., McDermid, G.J., and Gartrell, M. 2020. Estimating individual conifer seedling height using drone-based image point clouds. *Forests*, 11: 924.
- DJI. 2018. Phantom 4 PRO/PRO+ User Manual. Available from https://dl.djicdn.com/downloads/phantom_4_pro/Phantom+4+Pro+Pro+Plus+User+Manual+v1.0.pdf.
- Feduck, C., McDermid, G., and Castilla, G. 2018. Detection of coniferous seedlings in drone imagery. *Forests*, 9: 432.
- Fromm, M., Schubert, M., Castilla, G., Linke, J., and McDermid, G. 2019. Automated detection of conifer seedlings in drone imagery using convolutional neural networks. *Remote Sens.* 11: 2585.
- Giannetti, F., Chirici, G., Gobakken, T., Næsset, E., Travaglini, D., and Puliti, S. 2018. A new approach with DTM-independent metrics for forest growing stock prediction using drone photogrammetric data. *Remote Sens. Environ.* 213: 195–205. doi: [10.3390/rs11212585](https://doi.org/10.3390/rs11212585).
- Goodbody, T.R.H., Coops, N.C., Hermosilla, T., Tompalski, P., and Crawford, P. 2017. Assessing the status of forest regeneration using digital aerial photogrammetry and unmanned aerial systems. *Int. J. Remote Sens.* 1–19. doi: [10.1080/01431161.2017.1402387](https://doi.org/10.1080/01431161.2017.1402387).
- Guerra-Hernández, J., González-Ferreiro, E., Monleón, V.J., Faias, S.P., Tomé, M., and Díaz-Varela, R.A. 2017. Use of multi-temporal drone-derived imagery for estimating individual tree growth in *Pinus pinea* stands. *Forests*, 8: 300. doi: [10.3390/f8080300](https://doi.org/10.3390/f8080300).
- Hartigan, J.A., and Hartigan, P.M. 1985. The dip test of unimodality. *Ann. Statist.* 13: 70–84.

- Iglhaut, J., Cabo, C., Puliti, S., Piermattei, L., O'Connor, J., and Rosette, J. 2019. Structure from motion photogrammetry in forestry: A review. *Curr. For. Rep.* **5**: 155–168. doi: [10.1007/s40725-019-00094-3](https://doi.org/10.1007/s40725-019-00094-3).
- Imangholiloo, M., Saarinen, N., Markelin, L., Rosnell, T., Näsi, R., Hakala, T., et al. 2019. Characterizing seedling stands using leaf-off and leaf-on photogrammetric point clouds and hyperspectral imagery acquired from unmanned aerial vehicle. *Forests*, **10**: 415. doi: [10.3390/f10050415](https://doi.org/10.3390/f10050415).
- Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K.T., et al. 2018. Remote sensing and forest inventories in Nordic countries — Roadmap for the future. *Scand. J. Forest Res.* **33**: 397–412. doi: [10.1080/02827581.2017.1416666](https://doi.org/10.1080/02827581.2017.1416666).
- Kangas, A., Gobakken, T., and Næsset, E. 2020. Benefits of past inventory data as prior information for the current inventory. *Forest Ecosyst.* **7**: 20. doi: [10.1186/s40663-020-00231-6](https://doi.org/10.1186/s40663-020-00231-6).
- Korhonen, L., Pippuri, I., Packalén, P., Heikkinen, V., Maltamo, M., and Heikkilä, J. 2013. Detection of the need for seedling stand tending using high-resolution remote sensing data. *Silva Fennica*, **47**: 1–20. doi: [10.14214/sf.952](https://doi.org/10.14214/sf.952).
- Korpela, I., Tuomola, T., Tokola, T., and Dahlin, B. 2008. Appraisal of seedling stand vegetation with airborne imagery and discrete-return LiDAR — An exploratory analysis. *Silva Fennica*, **42**: 753–772. doi: [10.14214/sf.466](https://doi.org/10.14214/sf.466).
- Liaw, A., and Wiener, M. 2002. Classification and regression by random forest. *R News*, **2**: 18–22.
- Maechler, M. 2016. Hartigan's dip test statistic for unimodality — Corrected. Available from <https://cran.r-project.org/web/packages/diptest/diptest.pdf>.
- Mulverhill, C., Coops, N.C., Tompalski, P., and Bater, C.W. 2020. Digital terrestrial photogrammetry to enhance field-based forest inventory across stand conditions. *Can. J. Remote Sens.* 1–18.
- Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., et al. 2004. Laser scanning of forest resources: The Nordic experience. *Scand. J. Forest Res.* **19**: 482–499. doi: [10.1080/02827580410019553](https://doi.org/10.1080/02827580410019553).
- Ørka, H.O., Gobakken, T., and Næsset, E. 2016. Predicting attributes of regeneration forests using airborne laser scanning. *Can. J. Remote Sens.* **42**: 541–553. doi: [10.1080/07038992.2016.1199269](https://doi.org/10.1080/07038992.2016.1199269).
- Pearse, G.D., Tan, A.Y.S., Watt, M.S., Franz, M.O., and Dash, J.P. 2020. Detecting and mapping tree seedlings in drone imagery using convolutional neural networks and field-verified data. *ISPRS J. Photogramm. Remote Sens.* **168**: 156–169. doi: [10.1016/j.isprsjprs.2020.08.005](https://doi.org/10.1016/j.isprsjprs.2020.08.005).
- Puliti, S., Dash, J.P., Watt, M.S., Breidenbach, J., and Pearse, G.D. 2020. A comparison of drone laser scanning, photogrammetry and airborne laser scanning for precision inventory of small-forest properties. *For. Int. J. Forest Res.* **93**: 150–162. doi: [10.1093/forestry/cpz057](https://doi.org/10.1093/forestry/cpz057).
- Puliti, S., Ørka, H., Gobakken, T., and Næsset, E. 2015. Inventory of small forest areas using an unmanned aerial system. *Remote Sens.* **7**: 9632. doi: [10.3390/rs70809632](https://doi.org/10.3390/rs70809632).
- Puliti, S., Solberg, S., and Granhus, A. 2019. Use of drone photogrammetric data for estimation of biophysical properties in forest stands under regeneration. *Remote Sens.* **11**: 233.
- Statistics Norway. 2020. Silviculture. Available from <https://www.ssb.no/jord-skog-jakt-og-fiskeri/statistikker/skogkultur>.

Appendix A

Confusion matrices for the classification of tending need

Table 1A. Confusion matrix for the cross-validated predictions using unmanned aerial vehicle predictor variables.

Logistic model	Field estimate		Total	User's accuracy
	No tending	Tending		
No tending	19	2	21	90%
Tending	2	4	6	67%
Total	21	6	27	—
Producer's accuracy	90%	67%	Overall accuracy	85%

Table 2A. Confusion matrix for the cross-validated predictions using unmanned aerial vehicle predictor variables and prior inventory information.

Logistic model	Field estimate		Total	User's accuracy
	No tending	Tending		
No tending	19	1	20	95%
Tending	2	5	7	71%
Total	21	6	27	—
Producer's accuracy	90%	83%	Overall accuracy	89%

Table 3A. Confusion matrix for the cross-validated predictions using airborne laser scanning predictor variables.

Logistic model	Field estimate		Total	User's accuracy
	No tending	Tending		
No tending	19	3	22	86%
Tending	2	3	5	60%
Total	21	6	27	—
Producer's accuracy	90%	50%	Overall accuracy	81%

Table 4A. Confusion matrix for the cross validated predictions using airborne laser scanning predictor variables and prior inventory information.

Logistic model	Field estimate		Total	User's accuracy
	No tending	Tending		
No tending	18	2	20	90%
Tending	3	4	7	43%
Total	21	6	27	—
Producer's accuracy	85%	67%	Overall accuracy	81%

Table 5A. Confusion matrix for the cross validated predictions using only the prior inventory information as predictor variables.

Logistic model	Field estimate		Total	User's accuracy
	No tending	Tending		
No tending	20	3	23	83%
Tending	1	3	4	75%
Total	21	6	27	—
Producer's accuracy	95%	50%	Overall accuracy	85%