

УДК: 528.88

DOI: 10.35595/2414-9179-2025-2-31-21-35

Д. А. Богатырев¹, А. А. Бурдин², С. В. Пьянков³

**ОЦЕНКА ХАРАКТЕРИСТИК
ЛЕСНЫХ НАСАЖДЕНИЙ ПЕРМСКОГО КРАЯ
ПО ДАННЫМ ВОЗДУШНОГО ЛАЗЕРНОГО СКАНИРОВАНИЯ
И МУЛЬТИСПЕКТРАЛЬНОЙ СЪЕМКИ**

АННОТАЦИЯ

Оценка характеристик лесных насаждений методами дистанционного зондирования представляет собой важную и актуальную задачу в современном лесном хозяйстве, экологическом мониторинге и управлении природными ресурсами. В настоящей работе на примере участка смешанного леса на территории Березовского р-на Пермского края выполнена оценка таксационных характеристик насаждений (породы деревьев, высоты, диаметра и запаса древесины) на основе данных аэрофотосъемки и воздушного лазерного сканирования (ВЛС) с беспилотного летательного аппарата (БПЛА). Для валидации результатов проведен наземный сплошной перечет деревьев с определением перечисленных таксационных характеристик. Съемка и ВЛС выполнена с высоты 150 м с помощью БПЛА «DJI Matrice 350». Выполнена сегментация данных ВЛС для выделения отдельных деревьев, в результате которой верно распознано 962 дерева на тестовом участке, что на 10 % меньше их реального количества. На основе данных ВЛС и мультиспектральной съемки выполнена оценка таксационных характеристик (высоты, диаметра и запасов древесины) методами множественной линейной регрессии и случайного леса «Random Forest». Показано, что высота деревьев по данным ВЛС оценивается с удовлетворительной точностью ($RMSE = 1,16$ м), тогда как оценка диаметра стволов ($RMSE = 5,40$ см) и запасов древесины требует совершенствования, в частности расширения обучающей выборки. При расчете запаса древесины ошибка модели составила $0,348\text{ м}^3$, а относительная ошибка на весь участок — 4,18 %. Для тестового участка выполнена также классификация пород деревьев на основе ортофотоплана, которая дала удовлетворительный результат.

КЛЮЧЕВЫЕ СЛОВА: воздушное лазерное сканирование, мультиспектральная съемка, таксационные характеристики, машинное обучение, случайный лес

¹ Пермский государственный национальный исследовательский университет, ул. Букирева, д. 15, Пермь, Россия, 614990, e-mail: gis@psu.ru

² Пермский государственный национальный исследовательский университет, ул. Букирева, д. 15, Пермь, Россия, 614990, e-mail: gis@psu.ru

³ Пермский государственный национальный исследовательский университет, ул. Букирева, д. 15, Пермь, Россия, 614990, e-mail: gis@psu.ru

Danil A. Bogatyrev¹, Artem A. Burdin², Sergey V. Pyankov³

ALGORITHMS FOR AUTOMATION OF SUB TREE TAXATION BASED ON AIRBORNE LASER SCANNING AND MULTISPECTRAL IMAGERY DATA

ABSTRACT

Assessment of forest stands characteristics by remote sensing methods is an important and urgent task in modern forestry, environmental monitoring and natural resource management. In the present study, an assessment of the taxation characteristics of plantings (tree species, height, diameter and growing stock) was carried out based on aerial photography and airborne laser scanning (ALS) data from an unmanned aerial vehicle (UAV) on the example of a mixed forest area in the Berezovsky district of Perm Krai. To validate the results, a ground-based survey of each tree with the definition of the listed taxation characteristics was carried out. The ALS was performed from an altitude of 150 m using the “DJI Matrice 350 UAV”. Segmentation of the ALS data was performed to identify individual trees, resulting in the correct recognition of 962 trees in the test area, which is 10% less than their actual number. Based on the ALS and multispectral survey data, the assessment of taxation characteristics (height, diameter, and wood stocks) was performed using multiple linear regression and random Forest methods. It is shown that the height of trees according to the VLS data is estimated with satisfactory accuracy (RMSE = 1.16 m), while the estimation of trunk diameter (RMSE = 5.40 cm) and wood reserves requires improvement, in particular, the expansion of the training sample. When calculating the wood stock, the model error was 0.348 m³, and the relative error for the entire site was 4.18 %. Also, a classification of tree species based on an orthophotoplane was performed for the test site, which gave a satisfactory result.

KEYWORDS: airborne laser scanning, multispectral imagery, forest stands characteristics, machine learning, random forest

ВВЕДЕНИЕ

Оценка таксационных параметров древостоев проводится при отводе лесных участков, предназначенных под рубку. Традиционная методика таксации лесосек регламентируется приказом Министерства природных ресурсов и экологии РФ⁴. Данный способ сводится к получению качественных и количественных характеристик насаждений на участке глазомерным или инструментальным путем, при этом конечной целью является определение объема заготавливаемой древесины. При площади лесосеки менее 3 га производится сплошной перечет всех деревьев, при большей площади закладывается ленточный перечет или реласкопические площадки. Определяются диаметры ствола на высоте 1,3 м, высота деревьев, породный состав, полнота древостоя и прочие характеристики, указываемые в таксационном описании лесосеки. Погрешность при определении объема древесины и породного состава не должна превышать 10 %⁵. Неточности при таксации лесосек могут повлечь дополнительные расходы для предприятий, занимающихся лесозаготовкой. В связи с этим разрабатываются альтернативные методы таксации лесосек, в частности на основе съемки с БПЛА мультиспектральными камерами и воздушными лазерными сканерами (лидарами).

¹ Perm State University, 15, Bukireva str., Perm, 614990, Russia, e-mail: gis@psu.ru

² Perm State University, 15, Bukireva str., Perm, 614990, Russia, e-mail: gis@psu.ru

³ Perm State University, 15, Bukireva str., Perm, 614990, Russia, e-mail: gis@psu.ru

⁴ Приказ Министерства природных ресурсов и экологии РФ от 17 октября 2022 г. № 688. М., 2022. 50 с.

⁵ Лесной кодекс РФ от 04.12.2006 № 200-ФЗ. М., 2024. 79 с.

В России в нормативных документах не описано использование данных дистанционного зондирования Земли (ДЗЗ) при отводе лесосек. Между тем в мировой практике при решении этой задачи широко применяется как мультиспектральная съемка с БПЛА, так и воздушное лазерное сканирование (ВЛС). Основным преимуществом последнего является высокая плотность и проникающая способность лазерных импульсов, что позволяет построить высокодетальную трехмерную модель рельефа и насаждений в виде облака точек, определить с приемлемой точностью основные таксационные характеристики (высоту, размеры кроны дерева, полноту насаждений) и рассчитать разными способами запас древесины.

Данные ВЛС являются информационной основой мониторинга и инвентаризации лесов во многих странах мира. Изначально для задач инвентаризации лесов использовался подход, основанный на выделении однородных площадей, близких по характеристикам к таксационным выделам «area-based approach». Опыт его применения рассматривается в ряде публикаций [Junttila et al., 2008; Kauranne et al., 2017]. В последующие годы совершенствование лидаров и повышение плотности точек лазерного сканирования позволило развивать методы, основанные на идентификации отдельных деревьев по данным ВЛС. Такие алгоритмы предполагают на первом этапе группировку точек, которые соответствуют отдельным деревьям или их группам — т. н. вокселизацию. Затем вершины деревьев выявляются на основе идентификации локальных максимумов высоты или выпуклости [Anandakumar et al., 2019] с помощью методов подгонки формы «shape fitting» [Qian et al., 2023] или с помощью кластеризации [Li et al., 2023]. В работе [Pu et al., 2023] предлагаются различные стратегии обработки данных ВЛС для идентификации отдельных деревьев в сложных лиственных насаждениях, включая методы на основе облака точек и методы уточнения с помощью пространственной кластеризации. Более современные алгоритмы используют сверхточные нейронные сети для работы с облаками точек, такие как, например, «PointNet» [Kim et al., 2023]. Характерные значения точности детектирования отдельных деревьев в лесном массиве составляют порядка 0,6–0,75 (по показателю F-меры). Обычно число пропусков превышает число ложных детектирований, т. к. пропускаются относительно низкие и тонкие деревья, вклад которых в запас насаждения невелик. Точность измерения высоты деревьев находится на уровне десятков см [Sparks et al., 2022]. В работе [Chen et al., 2021] была достигнута точность определения высоты в пределах 0,2 м, и количества деревьев на уровне 90 %. В целом данное направление исследований в мире активно развивается на фоне повышения доступности как самих данных ВЛС, так и развития новых алгоритмов их обработки. Однако перспективы применения данных ВЛС для оценки характеристик сложных насаждений (смешанных, разновозрастных, с разновозрастными вырубками) требуют дополнительного изучения. Целью настоящей работы является оценка точности определения таксационных характеристик смешанных лесов на основе данных ВЛС и мультиспектральной съемки с БПЛА.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Проведение мультиспектральной аэрофотосъемки с БПЛА и создание ортофотопланов

Мультиспектральная аэрофотосъемка проведена с помощью камер «MicaSense RedEdge MX», установленной на БПЛА «DJI Matrice 300 RTK» и «DJI Mavic 3 Multispectral». Камера «MicaSense RedEdge MX» фиксирует изображение в пяти спектральных каналах: Red (668 nm), Green (560 nm), Blue (475 nm), RedEdge (717 nm) и NIR (840 nm) с разрешением 1280×960 пикс. Камера «DJI Mavic 3 Multispectral» работает в четырех каналах (Red, Green, NIR, RedEdge) и имеет матрицу 2592×1944 пикс. Съемка выполнялась в надир, с пространственным разрешением 15–35 см/пикс, продольным перекрытием 80 %

и поперечным — 70 %. Для более точного построения ортофотопланов использовалась цифровая модель местности. Съемка велась в вегетационный период (май–сентябрь) при безоблачных или полностью облачных погодных условиях, чтобы минимизировать влияние теней. Для калибровки яркости использовалась отражательная панель «MicaSense RedEdge MX» и встроенный сенсор «DLS Mavic 3 Multispectral».

С помощью ПО «Agisoft Metashape» было создано 32 мультиспектральных ортофотоплана на общую площадь 1 653 га. В процессе обработки выполнялась коррекция яркости снимков и объединение каналов в синтезированное изображение. Окончательный результат представлен в виде геопривязанных ортофотопланов, позволяющих различать породы деревьев. Пример полученного ортофотоплана представлен на рис. 1.

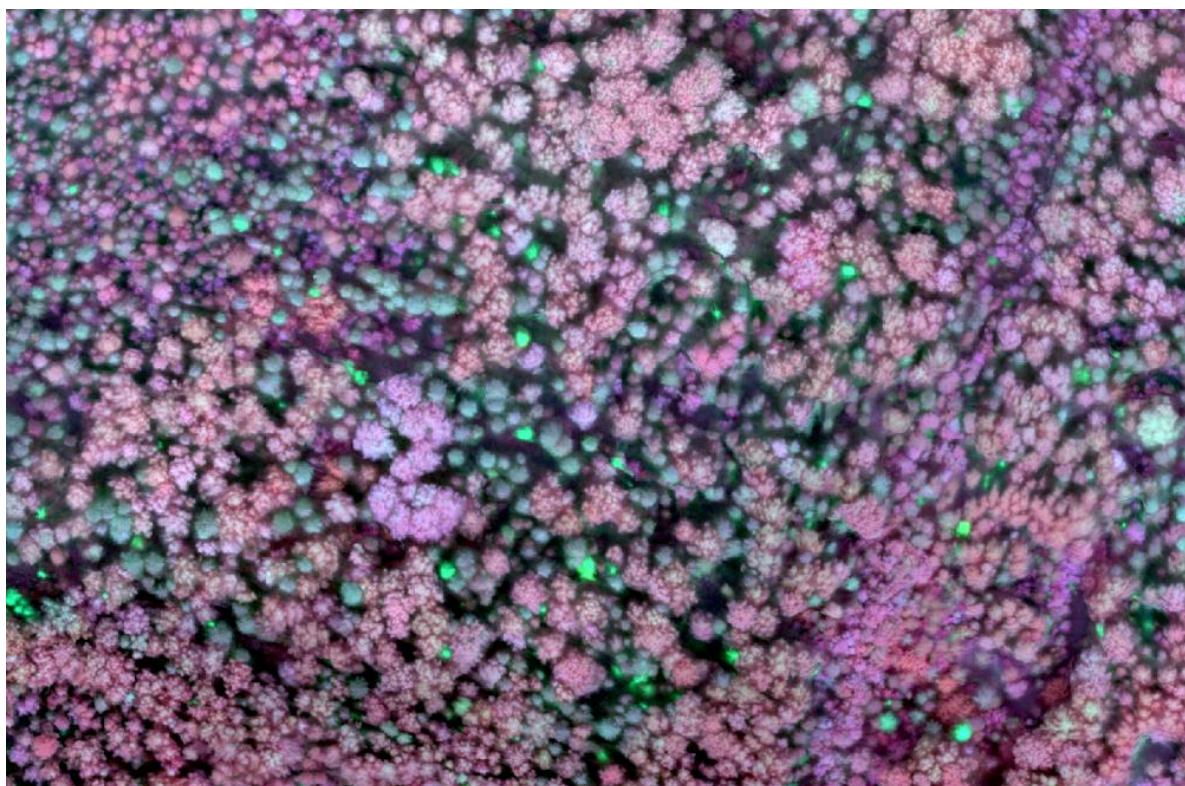


Рис. 1. Пример мультиспектрального ортофотоплана с размером пикселя 35 см, в синтезе каналов «ближний ИК, красный, зеленый»

Fig. 1. An example of a multispectral orthophotoplane with a pixel size of 35 cm, in the synthesis of near-infrared, red and green channels

Съемка с помощью воздушного лазерного сканера (лидара)

Воздушное лазерное сканирование (ВЛС) также выполнено с носителя БПЛА «DJI Matrice 350»¹, на котором был установлен подвес «LiDAR Zenmuse L2». Данный подвес установлен на трехосевом стабилизаторе, что позволяет сканировать не только в надир, но и под разными углами. Запуск БПЛА выполняется с открытой площадки в лесу (вырубки, дороги, поля) по заранее записанному полетному заданию. Для высокоточного позиционирования использована станция «GNSS D-RTK-2». Съемка ВЛС выполнена с высоты 150 м, боковое перекрытие между пролетами составляло 60 %, съемка велась в надир,

¹ DJI Enterprise. Matrice 350 RTK. Specs. Электронный ресурс: <https://enterprise.dji.com/matrice-350-rtk/specs>. (дата обращения 09.03.2025)

скорость полета 10 м/с, плотность выходного облака точек 200 т/м². Стоит отметить и интервал съемки: в среднем полет длится не более 30 мин. Высокая точность геопозиционирования (1–2 см) определяется установленной базовой станцией «RTK», которая передает дифференциальные поправки на БПЛА. Аналогичным образом были построены и выполнены полетные задания для «Mavic 3 Multispectral». Съемка выполнена с размером пикселя 30 см, продольным перекрытием снимков 80 %, поперечным перекрытием 70 %, скоростью полета 15 м/с. Общая площадь покрытия данными ВЛС составляет порядка 1 650 га.

Данные в процессе съемки записаны на внутренний накопитель и содержат в себе файлы записи облака точек, траекторий, фотографий и координат базовой станции, которые в дальнейшем интерпретируются в соответствующем ПО.

Предварительная обработка данных ВЛС

Воздушное лазерное сканирование выполнено с использованием БПЛА «DJI Matrice 350 RTK», оснащенного «LiDAR Zenmuse L2». Подвес установлен на трехосевом стабилизаторе, обеспечивающем сканирование в надир и под разными углами. Для геопривязки использовалась базовая станция «D-RTK-2», передающая поправки в реальном времени. Съемка выполнена с высоты 150 м, с боковым перекрытием 60 % и плотностью выходного облака точек 200 т/м². Средняя продолжительность полета составляла 30 мин. Данные включали облака точек, траектории полета, изображения и координаты базовой станции. Обработка данных ВЛС выполнялась в «DJI Terra», где облако точек конвертировалось в формат «LAS». В ПО «Lidar360» выполнялись: классификация точек земли с помощью фильтров, удаление шумов и артефактов, нарезка облака точек с буферной зоной, фильтрация точек для построения цифровой модели местности (ЦМР) и модели лесного полога (ЦМЛП).

Формирование баз данных таксационных характеристик насаждений

Для закладки эталонов и сверки данных был выбран участок в Березовском районе Пермского края (57.783°N 57.593°E). Площадь участка 2,52 га, формула состава, исходя из таксации, проведенной в 2024 г. — 7ЛП1Б1Е1П. Более детально характеристики насаждений на участке приведены в табл. 1. Средний по участку запас древесины составляет 200 м³/га, полнота насаждений 0,5 — неравномерная.

*Табл. 1. Характеристики лесных насаждений на тестовом участке
Table 1. Characteristics of forest stands in the test area*

Порода дерева	Липа	Береза	Ель	Пихта	Ива
Возраст (лет)	65	65	80	80	—
Средняя высота (H, м)	22	20	21	18	—
Средний диаметр ствола (D, см)	24	20	24	20	—
Объем на участке (V, м ³)	352	50	51	50	—
Количество деревьев на участке	567	264	151	65	17
Количество сухостоя	0	6	7	31	0
H (м)	23	19	24	20	16
D (см)	26	19	35	35/24	15

Для данного участка был выбран метод сплошного перечета стоящих на корню деревьев. Помимо основных характеристик, таких как порода, высота, диаметр, были определены координаты ствола каждого дерева с целью дальнейшего сравнения сегментации, полученной по облаку точек ВЛС. В перечетную ведомость не включены деревья

диаметром менее 12 см, а также поваленные деревья. Сухостойные деревья внесены в перечетную ведомость отдельной строкой. С помощью сортиментных и товарных таблиц для равнинных лесов Урала были установлены разряды высот деревьев и в дальнейшем вычислены объемы каждого дерева на лесном участке (табл. 2).

*Табл. 2. Объем деревьев на лесном участке
Table 2. The volume of trees in the forest area*

Порода дерева	Разряд высот	Объем ствола в коре (м^3)
Липа	1	385
Береза	3	82
Ель	5	172
Пихта	3	31

Суммарный объем на участке — 670 $\text{м}^3/\text{га}$, запас — 265 $\text{м}^3/\text{га}$, полнота — 0,6, а формула породного состава — 6ЛП1БЗЕ+П. Сравнив данные характеристики с данными таксационных описаний лесоустройства, можно заметить превышение объема на 25 %, а также увеличение доли хвойных пород в составе. Стоит отметить, что сплошной перечет является более точным, чем таксация, т. к. расчет характеристик производится у каждого дерева, что в дальнейшем даст возможность более корректно проводить сравнение с данными ВЛС.

Предварительная обработка плотного облака точек

Для обработки плотного облака точек использован пакет «LidR» [Roussel et al., 2021], реализованный в языке «R». На первом этапе произведена загрузка и проверка облака точек ВЛС. Затем проведена классификация точек земли для построения ЦМР. Из протестированных методов классификации (PMF, CSF, MCC) метод «CSF» показал лучшие результаты, минимизируя ошибки при определении точек земли. ЦМР построена с использованием алгоритма обратного взвешивания расстояния (ОВР) с размером пикселя 0,5 м. Нормализация облака точек была выполнена относительно ЦМР и исходного облака точек для повышения точности. Для построения ЦМЛП использован алгоритм «point to raster» с размером пикселя 0,5 м. Недостаток метода — появление пустых пикселей, если разрешение сетки слишком велико для плотности точек. Один из вариантов устранения пустот — замена точек в облаке дисками известного радиуса (например, 15 см). Это моделирует лазерный след как круговую область, улучшая физический смысл модели. Проводится интерполяция пустых пикселей и сглаживание полученной ЦМЛП.

Индивидуальное обнаружение и сегментация деревьев

Индивидуальное обнаружение деревьев — это процесс определения местоположения деревьев и извлечения метрик из данных ВЛС. Сегментация деревьев заключается в разграничении их по кронам. Для этого была написана функция фильтра локального максимума с переменным размером окна, адаптированного под высоту точек в облаке. Для точек высотой менее 10 м используется окно 2 м, более 20 м — 4 м, а для промежуточных значений размер окна определяется в соответствии с нелинейной зависимостью, которая учитывает высоту точек в плотном облаке. Этот подход обеспечил точное разделение крон деревьев. Алгоритм верно определил 962 дерева, что на 10 % меньше реального количества. Из-за недостаточной точности наземных данных привязка по деревьям не проводилась. Для сегментации выбран алгоритм [Dalponte, 2016], обеспечивающий точное определение крон деревьев. В каждом сегменте рассчитаны стандартные метрики высот (рис. 2).

Построение линейных моделей предсказания высоты и диаметра деревьев

При построении регрессионных моделей для оценки высоты деревьев (H) и диаметра деревьев на уровне 1,3 м (D) необходимо из набора метрик, полученных из данных ВЛС, идентифицировать предикторы, которые в наибольшей степени объясняют изменчивость H и D по проверочным данным (результатам наземной съемки деревьев). Метрики, имеющие высокую корреляцию с целевой переменной, могут быть использованы в модели. Однако следует избегать использования сильно коррелированных между собой метрик. Для этого был проведен корреляционный анализ ключевых переменных, связанных с высотой и диаметром деревьев.

Учитывая малое количество статистически значимых метрик, которые были отобраны по корреляционному анализу, было решено использовать метод пошагового регрессионного отбора «Stepwise selection». Он основан на итеративном добавлении и удалении переменных из модели с целью минимизации значения информационного критерия Акаике «AIC». Этот метод является эффективным инструментом для идентификации наиболее значимых переменных, особенно в ситуациях, когда данных много, а их интерпретация затруднена из-за мультиколлинеарности.

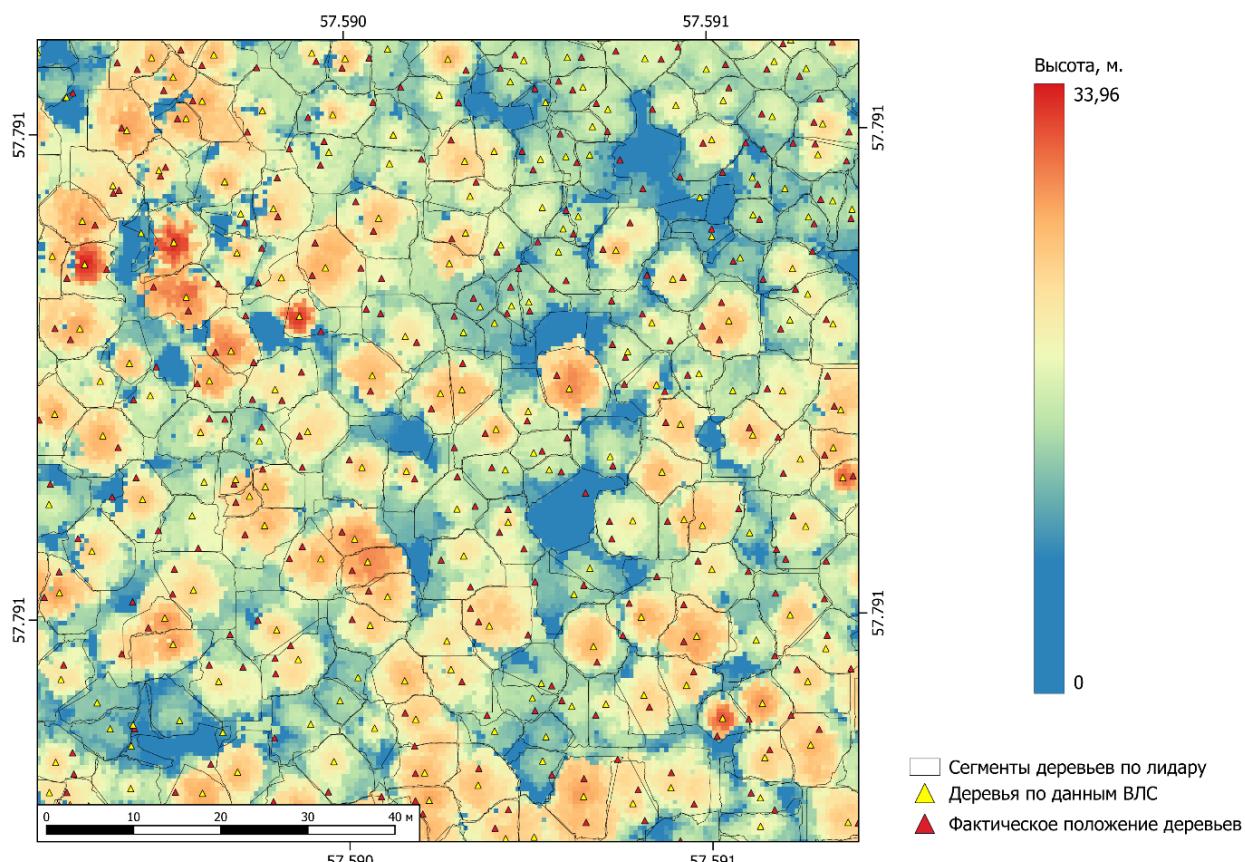


Рис. 2. Результат индивидуального определения и сегментации деревьев

Fig. 2. Result of individual tree identification and segmentation

В линейной модели предсказания высоты использовались восемь предикторов, среди которых наибольший вклад внесли: максимальная высота точек (Z_{max}), средняя высота (Z_{mean}) и стандартное отклонение высот (Z_{sd}). В модели диаметра ключевыми предикторами стали (Z_{max}), коэффициент эксцесса (Z_{kurt}) и процент точек выше 2 м ($Pzabov2$). Значимый вклад также внесли метрики, связанные с процентилями высот: $zq30$, $zq40$, $zq65$, $zq75$, $zq80$, $zq85$ и $zq90$. Для проверки точности модели данные разделялись на

обучающую (80 %) и тестовую (20 %) выборки. Оценка точности проводилась по коэффициенту детерминации (R^2) и среднеквадратичной ошибке (RMSE).

Построение и проверка регрессионной модели Random Forest для предсказания высоты и диаметра деревьев

Метод машинного обучения «Random Forest» позволяет выявлять сложные взаимосвязи и взаимодействия благодаря использованию ансамблей деревьев решений, что делает их проверку особенно значимой в данной задаче [Breiman et al., 2001]. «Random Forest (RF)» строит множество независимых деревьев решений, каждое из которых обучается на случайной подвыборке данных и случайных признаках, что позволяет снизить переобучение и повысить устойчивость модели [Ho, 1998]. Этот метод обеспечивает высокую точность благодаря ансамблевому подходу, где результаты предсказания каждого дерева объединяются.

Была проведена настройка гиперпараметров с использованием сетки параметров модели. Метод оценивает все возможные комбинации гиперпараметров и выбирает ту, при которой ошибка минимальна. Для этого была создана сетка возможных значений параметра *mtry*, а также настроены параметры повторной кросс-валидации. Обучение модели с использованием функции «*train*» из пакета «*caret*» позволило автоматически определить оптимальное значение гиперпараметра *mtry*, минимизируя «RMSE».

Для проверки эффективности модели для оценки *H* была проведена оценка на той же тестовой выборке, что и в случае проверки линейных моделей. «RMSE» на тестовых данных составила 1,17, что свидетельствует о немного большей точности по сравнению с моделью множественной линейной регрессии. В модели предсказания высоты наибольшее влияние оказали «*Zmax*, *Zq85* и *Zmean*», в модели диаметра — «*Zmax*, *Zkurt* и *Pzabov2*». Оценка точности моделей представлена в табл. 4.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Результаты оценки высоты и диаметра деревьев на основе линейной модели и модели «Random Forest» приведены в табл. 3–4. Линейная модель на тестовой выборке показала точность предсказания высоты деревьев с RMSE = 1,20 и R^2 = 0,9233. Для диаметра модель показала RMSE = 5,40 и объяснила 69,21 %. Результаты могут указывать на то, что для предсказания диаметра могут потребоваться дополнительные метрики и параметры или более сложные нелинейные модели, такие как «Random Forest» или «Gradient Boosting». Очевидно, что в предсказании диаметра большое влияние будет играть характеристика породы дерева. Это говорит о необходимости разработки метода определения породы по данным ВЛС и проверки модели с учетом этой переменной.

Табл. 3. Результаты линейных моделей предсказания высоты (H) и диаметра (D) деревьев
Table 3. The results of linear models for predicting the height (H) and diameter (D) of trees

Параметр	Предикторы	RMSE	R^2	RMSE (тестовая выборка)
Высота (H, м)	<i>zmax</i> , <i>zmean</i> , <i>zsd</i> , <i>zkurt</i> , <i>pzabov2</i> , <i>zq30</i> , <i>zq65</i> , <i>zq85</i>	1,019	0,9233	1,20
Диаметр (D, см)	<i>zmax</i> , <i>zkurt</i> , <i>pzabov2</i> , <i>zq40</i> , <i>zq75</i> , <i>zq80</i> , <i>zq85</i> , <i>zq90</i>	5,56	0,6921	5,40

Табл. 4. Результаты моделей RF для предсказания высоты (H) и диаметра (D) деревьев
Table 4. The results of RF models for predicting the height (H) and diameter (D) of trees

Параметр	Предикторы	R ²	RMSE (тестовая выборка)	Кол-во деревьев (ntree)	Количество переменных (mtry)
Высота (H, м)	zmax, zmean, zsd, zkurt, pzbvzmn, zq30, zq65 и zq85.	0,9206	1,17	500	4
Диаметр (D, см)	zmax, zkurt, pzbvzmn, pzabov2, zq15, zq40, zq75, zq80, zq85 и zq90	0,6990	5,86		8

Модель RF, несмотря на свою большую гибкость, дает несколько меньшую точность, что может быть связано с особенностями самих данных, более простой зависимостью между признаками и целевой переменной, а также с возможными проблемами переобучения в модели RF. В отличие от случайного леса, который способен выявлять более сложные и нелинейные зависимости, линейная модель может дать более стабильные результаты при наличии выбросов в данных или при слабой корреляции между признаками, что и наблюдается в наших данных.

Определение высоты и диаметра деревьев на уровне ячеек сетки размером 10×10 м

Для анализа взаимосвязи между лидарными метриками и характеристиками в ячейках сетки (10×10 м) были рассчитаны стандартные и дополнительные метрики высоты. Полевые данные (средняя высота, средний диаметр, максимальная высота, сумма высоты и диаметра деревьев в пределах ячейки) были пространственно объединены с лидарными метриками в ArcGIS. Для исключения краевых эффектов применялся 10-метровый буфер облака точек. Итоговый набор включал 280 ячеек сетки. Отбор предикторов осуществлялся методом пошаговой регрессии, оставляя только наиболее значимые переменные. Данные делились на обучающую (80 %) и тестовую (20 %) выборки. Линейная модель предсказания средней высоты деревьев (H_mean) включила девять лидарных метрик, наиболее значимыми из которых оказались:

- соотношение плотности точек ниже заданной высоты 20 м (rat 20m);
- коэффициенты вариации высоты (zmax и zcv).

Модель предсказания среднего диаметра деревьев (D_mean) включила семь предикторов, среди которых логарифм соотношения точек земли к общему числу точек (lg_rt_g), (zmean) и (zmax) оказались наиболее значимыми. Результаты линейных моделей и моделей RF представлены в табл. 5.

Таким образом, результаты моделирования средней высоты (H_mean) и среднего диаметра (D_mean) деревьев на уровне ячеек сетки размером 10×10 м с использованием как линейных моделей, так и моделей «Random Forest» имеют ограниченную точность. Несмотря на тщательный отбор лидарных метрик и оптимизацию гиперпараметров, оба подхода продемонстрировали значительные отклонения предсказанных значений от реальных данных.

Табл. 5. Результаты моделей предсказания высоты (H_mean) и диаметра (D_mean) деревьев в ячейке 10×10 м
Table 5. Results of models predicting the height (H_mean) and diameter (D_mean) of trees in a 10×10 m cell

Параметр	Модель	Предикторы	R ²	RMSE (тестовая выборка)	Кол-во деревьев (ntree)	Количество переменных (mtry)
Средняя высота в ячейке (H, м)	Линейная	rat_20m, zmax, zcv, q15, q20, q25, q60, q65, q70	0,6804	2,40	—	—
	Random Forest	rat_20m, zmax, zcv, q15, q20, q25, q60, q65, q70	0,5809	2,49	500	2
Средний диаметр (D, см)	Линейная	lg_rt_g, zmean, zmax, q20, q35, q65	0,4506	5,64	—	—
	Random Forest	lg_rt_g, zmean, zmax, q20, q35, q65, q80	0,2652	—	500	4

Линейные модели для «H_mean» и «D_mean», хотя и обладают простотой интерпретации, показали относительно низкий уровень объясненной дисперсии и высокие значения RMSE. Модели RF, несмотря на их способность учитывать нелинейные зависимости, продемонстрировали еще более низкую точность. Эти результаты указывают на недостаточную информативность лидарных метрик для точного моделирования высоты и диаметра деревьев в среднем для ячейки регулярной сетки. С учетом полученных результатов дальнейшее применение моделей для определения объема на уровне ячеек сетки представляется нецелесообразным.

Оценка запаса древесины на тестовом участке

Учитывая, что высота деревьев при полевых обследованиях не была определена для всех деревьев, в учет шли только те, где она была известна — 474 деревьев из 1 064 на участке. Из-за недостаточной точности привязки при полевых обследованиях и возможных ошибок в определении сегментов по данным ВЛС, в некоторые сегменты деревья, определенные при полевых обследованиях, не попали. С учетом вышеперечисленных факторов для расчета запаса древесины было использовано 386 деревьев. Породы были разделены на хвойные и лиственные. Расчет производился на основе преобладающих пород: для хвойных по ели, для лиственных по липе.

Для хвойных пород, используя таблицу определения разряда высот деревьев для района предуральской равнины¹, где учитывается средняя высота (21,94 м) и средний диаметр (26,22 см) внутри участка, был определен разряд V. Для лиственных средняя высота составила 23,52 м и средний диаметр — 36,48 см, что соответствует разряду I. Учитывая разряд, были использованы соответствующие таблицы расчета объема древесины. На их основе был написан код на R для расчета объема для каждого выделенного дерева с учетом диаметра и принадлежности к лиственной или хвойной породе. Суммарный запас древесины на участке по фактическим данным для деревьев с известной высотой равен 311,4 м³.

¹ Сортиментные и товарные таблицы для равнинных лесов Урала. Федеральное агентство лесного хозяйства. Электронный ресурс: <https://rosleshoz.gov.ru/upload/iblock/31e/ipyel714kdmxxszmzglce3p9mfhspurc9.pdf> (дата обращения 28.03.2025)

Построение и проверка моделей для расчета запасов (объемов)

Для предсказания объема деревьев использовались стандартные лидарные метрики, включая высотные процентили ($zq80$, $zq85$, $zq90$), кумулятивные процентили ($zpcum1$, $zpcum6$, $zpcum8$) и процент точек выше заданного порога ($rzabov2$). Отбор предикторов выполнен методом пошаговой обратной регрессии.

С учетом того, что объемы деревьев в наших данных варьируются от 0,054 до 2,748 м³, ошибка модели на тестовых данных ($RMSE = 0,3486$ м³) является значительной, особенно для деревьев с меньшими объемами. Для деревьев с объемом ближе к верхнему диапазону ошибка составляет примерно 12,7 % от истинного значения. Таким образом, модель лучше справляется с предсказанием объемов для крупных деревьев, но менее точна для мелких.

Вместе с тем текущая модель не учитывает различия в породах деревьев, которые могут существенно влиять на их объемные характеристики. Добавление переменной «порода» (лиственные/хвойные) улучшило точность модели. R^2 увеличилось с 0,5859 до 0,6108, $RMSE$ снизилось с 0,348 м³ до 0,341 м³. Переменная «порода» оказалась статистически значимой ($p < 0,001$), что подтверждает важность учета различий в структуре древесины.

Модель «Random Forest» для предсказания объема использовала 500 деревьев ($ntrree = 500$) и 3 переменные на каждом разбиении ($mtry = 3$). На обучающей выборке объясненная дисперсия составила 61,09 %, но на тестовой выборке $RMSE$ оказалось выше (0,3618 м³), что говорит о слабой обобщающей способности модели. Данные результаты свидетельствуют о том, что линейная модель может быть предпочтительна для расчета объема на уровне участка, учитывая ее несколько лучшую обобщающую способность.

Сравнение фактического и расчетного запаса для всего участка

Для каждого из 386 деревьев, для которых были доступны все необходимые метрики, был рассчитан фактический и предсказанный объем. Линейная регрессионная модель была построена на обучающей выборке с исключением переменной порода. Сравнение суммарных значений объема в тестовой выборке демонстрирует хорошую степень соответствия между реальными данными и результатами модели. Суммарный фактический объем деревьев из тестовой выборки составляет 59,504 м³, в то время как суммарный предсказанный объем в той же выборке, рассчитанный с использованием модели, равен 61,99 м³. Абсолютная разница между этими значениями составляет 2,49 м³, или 4,18 %. Такое отклонение указывает на умеренную точность модели в оценке общего запаса древесины на уровне участка. Эти результаты подтверждают, что модель учитывает основные предикторы объема, включая структурные метрики, например, процентили высот и показатели плотности, без учета породы деревьев. Отклонения могут быть связаны с ограничениями модели, такими как нелинейные взаимодействия между переменными, погрешности в данных или недостаточное представление отдельных групп деревьев в обучающей выборке.

Определение древесной породы по ортофотоплану

Была проведена классификация ортофотоплана, полученного с использованием камеры «DJI Mavic 3 Multispectral». Исходные данные содержали 7 спектральных каналов: три канала RGB (синий, зеленый и красный) высокого разрешения, а также четыре спектральных канала (зеленый, красный, красный край и ближний ИК). Для обучения модели использовались полигоны, отрисованные на основе визуального анализа ортофотоплана с учетом полевых данных о породах. При этом полигоны создавались только для тех деревьев, где пространственная привязка полевых данных позволяла точно определить

породу. Тренировочные данные охватывали четыре класса: хвойные деревья, лиственные деревья, тени и земля. На основе обучающих полигонов и ортофотоплана была настроена модель «Random Forest» с параметрами: количество тренировочных выборок — 5 000, количество деревьев — 10.

Распределение числа пикселей по классам в итоговом слое классификации следующее: хвойные деревья — 23,99 %, лиственные деревья — 37,80 %, тени — 23,06 % и земля — 15,14 %. При этом фактическое распределение по породам выглядит так: хвойные деревья — 20,30 %, лиственные — 79,70 %. Различия в процентных соотношениях, вероятно, обусловлены различиями в размерах крон и занимаемой ими площади. Лиственные породы, как правило, имеют более развитые кроны по сравнению с хвойными, что приводит к увеличению количества пикселей, отнесенных к этому классу. Анализ значимости спектральных каналов показал, что наиболее важным оказался канал ближнего ИК-диапазона (канал 7) с наибольшим вкладом в точность модели. Меньшую значимость показали синий (канал 1) и красный мультиспектральный (канал 5) каналы. Результат классификации представлен на рис. 3.

Визуальный анализ получившейся схемы свидетельствует о соответствии полевых данных с теми, что были определены с помощью классификации моделью машинного обучения «Random Forest». Попадание некоторых хвойных деревьев в класс теней, а также попадание лиственных в хвойный класс можно отнести к ошибке, связанной с точностью привязки деревьев по ГНСС.

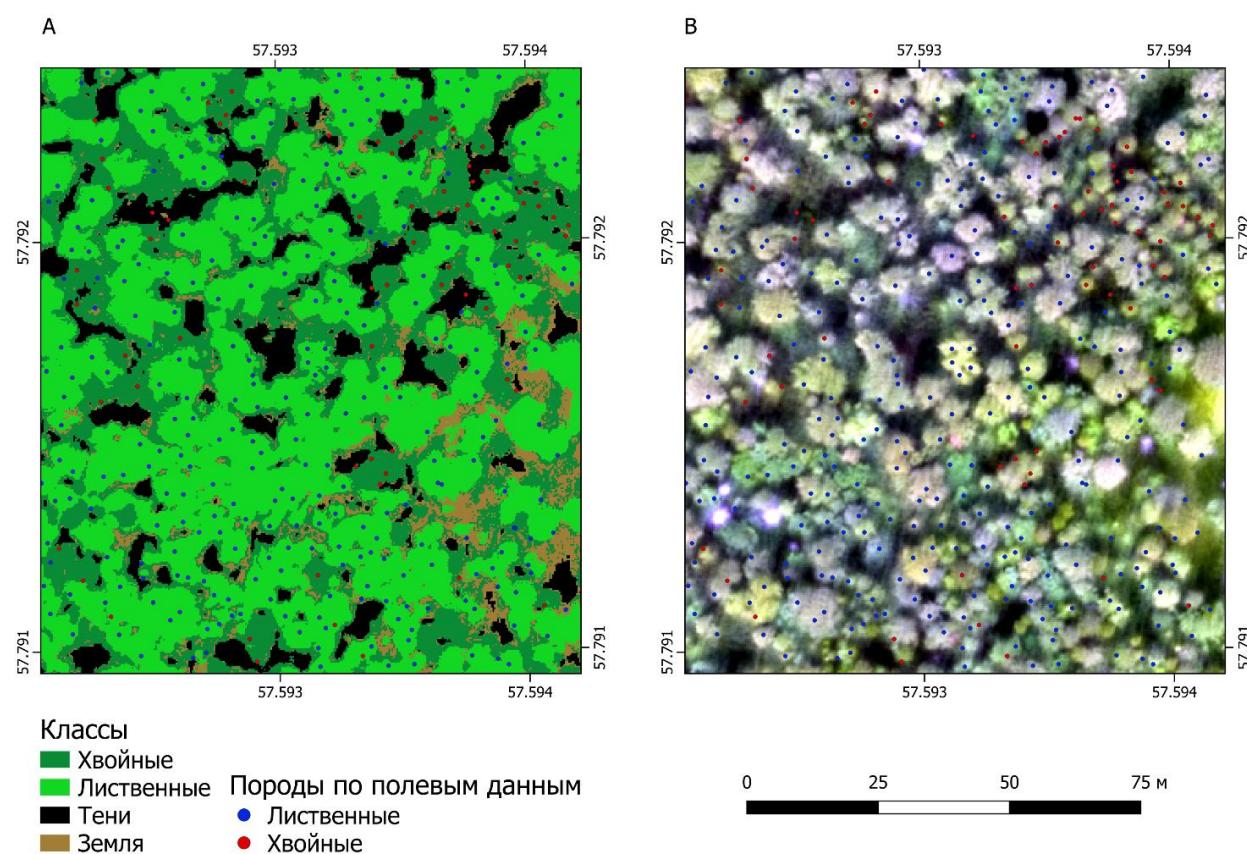


Рис. 3. Результат классификации Random Forest (A) и исходный ортофотоплан в искусственных цветах (B)

Fig. 3. The result of the Random Forest classification (A) and the initial orthophotoplane in artificial colors (B)

ВЫВОДЫ

Были выполнены задачи по обработке данных ВЛС и мультиспектральной съемки, анализу и выбору метрик, предсказанию высоты, диаметра и объема деревьев, а также классификации пород по данным ортофотоплана. Использованы как линейные модели регрессии, так и более сложные алгоритмы машинного обучения, такие как «Random Forest», для решения задач разной сложности.

Полученные результаты показали, что линейные модели обеспечивают высокую интерпретируемость и стабильность предсказаний для высоты деревьев на основе данных ВЛС, в то время как модели оценки диаметра деревьев требуют доработок. Для учета более сложных и нелинейных зависимостей в структурах данных были использованы модели «Random Forest», точность которых оказалась на уровне линейных моделей или несколько хуже. Это может быть обусловлено недостаточным размером обучающей выборки. Наилучшие результаты по определению высоты деревьев получены на основе модели случайного леса с ошибкой 1,16 м, что можно считать удовлетворительной точностью. Для диаметра наиболее точной оказалась линейная модель с ошибкой 5,40 см. При расчете запаса древесины ошибка линейной модели на тестовой выборке составила 0,348 м³, а относительная ошибка (разность фактического и расчетного запаса) на весь участок — 4,18 %. При классификации пород деревьев на основе мультиспектральных данных съемки с БПЛА модель RF продемонстрировала высокую общую точность (98,34 %). Наиболее значимыми являются ближний ИК и зеленый каналы.

Важным ограничением, выявлена в ходе работы, стало недостаточное количество полевых данных и их недостаточная пространственная точность, в связи с чем и был использован подход с агрегированием данных в прямоугольной сетке. Ограничения точности привязки деревьев и объема обучающих данных оказали существенное влияние на результаты. Для повышения качества обучения и тестирования моделей необходимо увеличить объем полевых данных, уделив особое внимание равномерному покрытию различных структурных участков леса. Более точная пространственная привязка данных позволила бы минимизировать ошибки, связанные с несовпадением полевых и дистанционных данных.

Совершенствование полученных результатов в дальнейшем, прежде всего, предполагает значительное расширение обучающих и проверочных наборов данных. Это позволит получить более достоверные оценки таксационных параметров насаждений по данным ВЛС на основе моделей машинного обучения, в т. ч. протестировать методы, которые не использовались в данной работе.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке Министерства образования и науки Пермского края в рамках Соглашения от 29.02.2024 № С-26/815.1.

ACKNOWLEDGEMENTS

The study was carried out with the financial support of the Ministry of Education and Science of Perm Krai within the framework of Agreement No. C-26/815.1 dated 29.02.2024.

СПИСОК ЛИТЕРАТУРЫ

Anandakumar R., Nidamanuri R., Krishnan R. Individual Tree Detection from Airborne Laser Scanning Data Based on Supervoxels and Local Convexity. *Remote Sensing Applications: Society and Environment*, 2019. V. 15. No. 100242. DOI: 10.1016/j.rsase.2019.100242.

Breiman L. Random Forests. *Machine Learning*, 2001. V. 45. Iss. 1. P. 5–32. DOI: 10.1023/A:1010933404324.

Chen X., Jiang K., Zhu Y., Wang X., Yun T. Individual Tree Crown Segmentation Directly From UAV-borne LiDAR Data Using the PointNet of Deep Learning. *Forests*, 2021. V. 12. No. 131. DOI: 10.3390/f12020131.

Dalponte M., Coomes D. A. Tree-Centric Mapping of Forest Carbon Density from Airborne Laser Scanning and Hyperspectral Data. *Methods in Ecology and Evolution*, 2016. V. 7. P. 1236–1245. DOI: 10.1111/2041-210X.12575.

Ho T. The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998. V. 20. No. 8. P. 832–844. DOI: 10.1109/34.709601.

Junttila V., Maltamo M., Kauranne T. Sparse Bayesian Estimation of Forest Stand Characteristics from Airborne Laser Scanning. *Forest Science*, 2008. V. 54. P. 543–552. DOI: 10.1093/forestscience/54.5.543.

Kauranne T., Pyankov S., Junttila V., Kedrov A., Tarasov A., Kuzmin A., Peuhkurinen J., Villikka M., Vartio V.-M., Sirparanta S. Airborne Laser Scanning Based Forest Inventory: Comparison of Experimental Results for the Perm Region, Russia and Prior Results from Finland. *Forests*, 2017. V. 8. No. 72. DOI: 10.3390/f8030072.

Kim D. H., Ko C. U., Kim D. G., Kang J. T., Park J. M., Cho H. J. Automated Segmentation of Individual Tree Structures Using Deep Learning over LiDAR Point Cloud Data. *Forests*, 2023. V. 14(6). No. 1159. DOI: 10.3390/f14061159.

Li Y., Xie D., Wang Y., Jin S., Zhou K., Zhang Z., Li W., Zhang W., Mu X., Yan G. Individual Tree Segmentation of Airborne and UAV LiDAR Point Clouds Based on The Watershed and Optimized Connection Center Evolution Clustering. *Ecology and Evolution*, 2023. V. 13. No. e10297. DOI: 10.1002/ece3.10297.

Pu Y., Xu D., Wang H., Li X., Xu X. A New Strategy for Individual Tree Detection and Segmentation from Leaf-on and Leaf-off UAV-LiDAR Point Clouds Based on Automatic Detection of Seed Points. *Remote Sensing*, 2023. V. 15. No. 1619. DOI: 10.3390/rs15061619.

Qian C., Yao C., Ma H., Xu J., Wang J. Tree Species Classification Using Airborne LiDAR Data Based on Individual Tree Segmentation and Shape Fitting. *Remote Sensing*, 2023. V. 15. No. 406. DOI: 10.3390/rs15020406.

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 Sensing of Environment*, 2021. V. 251. No. 112061. DOI: 10.1016/j.rse.2020.112061.

Sparks A., Smith A. Accuracy of a LiDAR-Based Individual Tree Detection and Attribute Measurement Algorithm Developed to Inform Forest Products Supply Chain and Resource Management. *Forests*, 2022. V. 13. No. 3. DOI: 10.3390/f13010003.

REFERENCES

Anandakumar R., Nidamanuri R., Krishnan R. Individual Tree Detection from Airborne Laser Scanning Data Based on Supervoxels and Local Convexity. *Remote Sensing Applications: Society and Environment*, 2019. V. 15. No. 100242. DOI: 10.1016/j.rsase.2019.100242.

Breiman L. Random Forests. *Machine Learning*, 2001. V. 45. Iss. 1. P. 5–32. DOI: 10.1023/A:1010933404324.

Chen X., Jiang K., Zhu Y., Wang X., Yun T. Individual Tree Crown Segmentation Directly From UAV-borne LiDAR Data Using the PointNet of Deep Learning. *Forests*, 2021. V. 12. No. 131. DOI: 10.3390/f12020131.

Dalponte M., Coomes D. A. Tree-Centric Mapping of Forest Carbon Density from Airborne Laser Scanning and Hyperspectral Data. *Methods in Ecology and Evolution*, 2016. V. 7. P. 1236–1245. DOI: 10.1111/2041-210X.12575.

Ho T. The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998. V. 20. No. 8. P. 832–844. DOI: 10.1109/34.709601.

Junttila V., Maltamo M., Kauranne T. Sparse Bayesian Estimation of Forest Stand Characteristics from Airborne Laser Scanning. *Forest Science*, 2008. V. 54. P. 543–552. DOI: 10.1093/forestscience/54.5.543.

Kauranne T., Pyankov S., Junttila V., Kedrov A., Tarasov A., Kuzmin A., Peuhkurinen J., Villikka M., Vartio V.-M., Sirparanta S. Airborne Laser Scanning Based Forest Inventory: Comparison of Experimental Results for the Perm Region, Russia and Prior Results from Finland. *Forests*, 2017. V. 8. No. 72. DOI: 10.3390/f8030072.

Kim D. H., Ko C. U., Kim D. G., Kang J. T., Park J. M., Cho H. J. Automated Segmentation of Individual Tree Structures Using Deep Learning over LiDAR Point Cloud Data. *Forests*, 2023. V. 14(6). No. 1159. DOI: 10.3390/f14061159.

Li Y., Xie D., Wang Y., Jin S., Zhou K., Zhang Z., Li W., Zhang W., Mu X., Yan G. Individual Tree Segmentation of Airborne and UAV LiDAR Point Clouds Based on The Watershed and Optimized Connection Center Evolution Clustering. *Ecology and Evolution*, 2023. V. 13. No. e10297. DOI: 10.1002/ece3.10297.

Pu Y., Xu D., Wang H., Li X., Xu X. A New Strategy for Individual Tree Detection and Segmentation from Leaf-on and Leaf-off UAV-LiDAR Point Clouds Based on Automatic Detection of Seed Points. *Remote Sensing*, 2023. V. 15. No. 1619. DOI: 10.3390/rs15061619.

Qian C., Yao C., Ma H., Xu J., Wang J. Tree Species Classification Using Airborne LiDAR Data Based on Individual Tree Segmentation and Shape Fitting. *Remote Sensing*, 2023. V. 15. No. 406. DOI: 10.3390/rs15020406.

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 Sensing of Environment*, 2021. V. 251. No. 112061. DOI: 10.1016/j.rse.2020.112061.

Sparks A., Smith A. Accuracy of a LiDAR-Based Individual Tree Detection and Attribute Measurement Algorithm Developed to Inform Forest Products Supply Chain and Resource Management. *Forests*, 2022. V. 13. No. 3. DOI: 10.3390/f13010003.