

**Методика определения
местоположений
и характеристик насаждений
на больших участках леса
с использованием данных LiDAR**

Актуальность и постановка задач



Портативный лазерный сканер

- Автоматизация принятия решений в задачах с большой долей ручного труда
- Построение цифровой модели местности
- Использование технологий лазерного сканирования
- Выделение и идентификация всех объектов
 - Нахождение местоположений позволяет связать различные данные (облака точек, фото, таксационная информация)
 - Описание особенностей местности, возможность контролируемо применять алгоритмы сегментации
 - Отсутствие обширных баз данных, надежных методов обработки данных, подходящих к лесам Центральной России

Данные LiDAR

Фрагмент файла PCD:

```
# .PCD v.7 - Point Cloud Data file format
...
FIELDS x y z intensity
...
POINTS 213863
DATA ascii
0.93773 0.33763 0 8.9129e+03
0.90805 0.35641 0 6.6056e+03
0.81915 0.32865 0 6.2492e+03
0.97192 0.27881 0 7.6208e+03
...
```

В общем виде 3D сцену с использованием лидарных данных можно представить в виде матрицы векторов, каждый вектор представляет собой выражение:

$$\mathbf{p}_i = \{x_i, y_i, z_i, \theta_i^1, \theta_i^2, \dots, \theta_i^M\}, i = 1 \dots N$$

где первые три параметра – координаты точки в евклидовом пространстве, остальные – дополнительные параметры (каналы цвета, интенсивность и прочее).

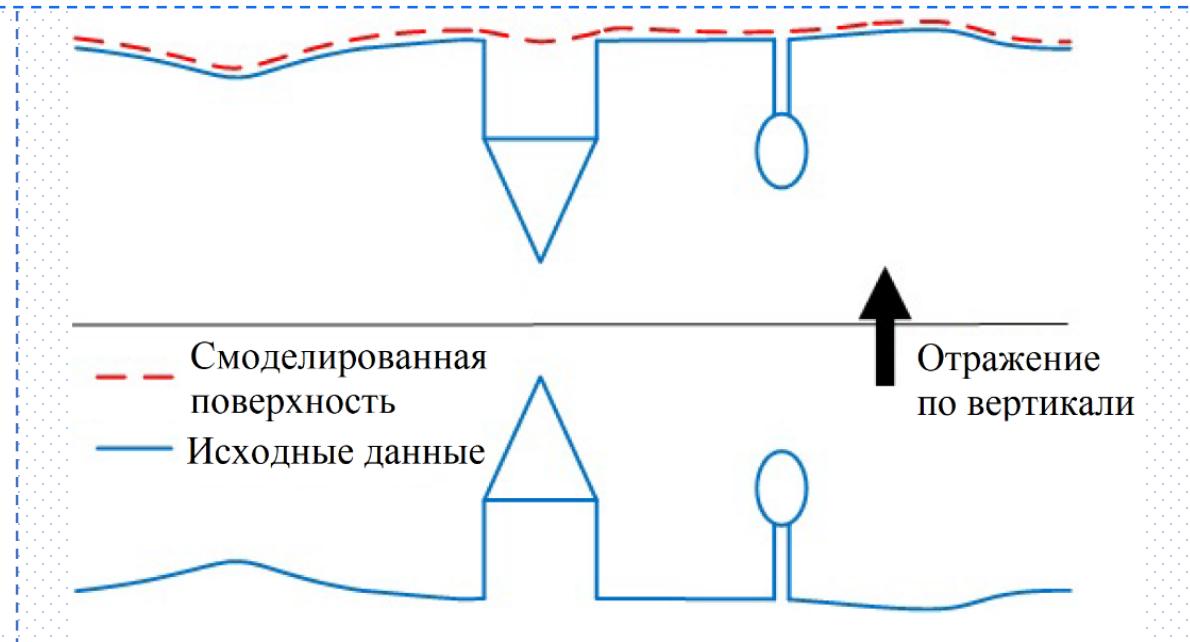
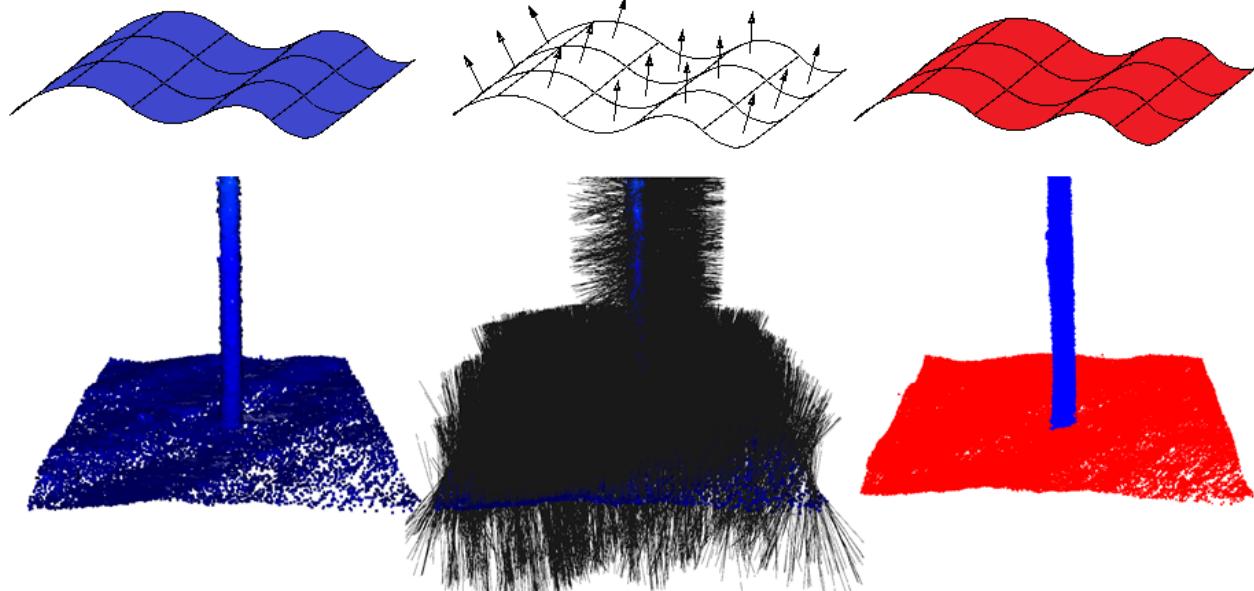
N – количество точек в 3D сцене,

M – количество дополнительных параметров.

Каждый объект представляется в виде выражения

$$\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}^T$$

Обнаружение точек земной поверхности



Метод на основе оценки нормалей

1. Разбиение на основе k-d-дерева
2. Вычисление главной оси соседних точек, используя ковариационный анализ
3. Фильтрация точек по допустимому углу наклона нормали и близости к земле

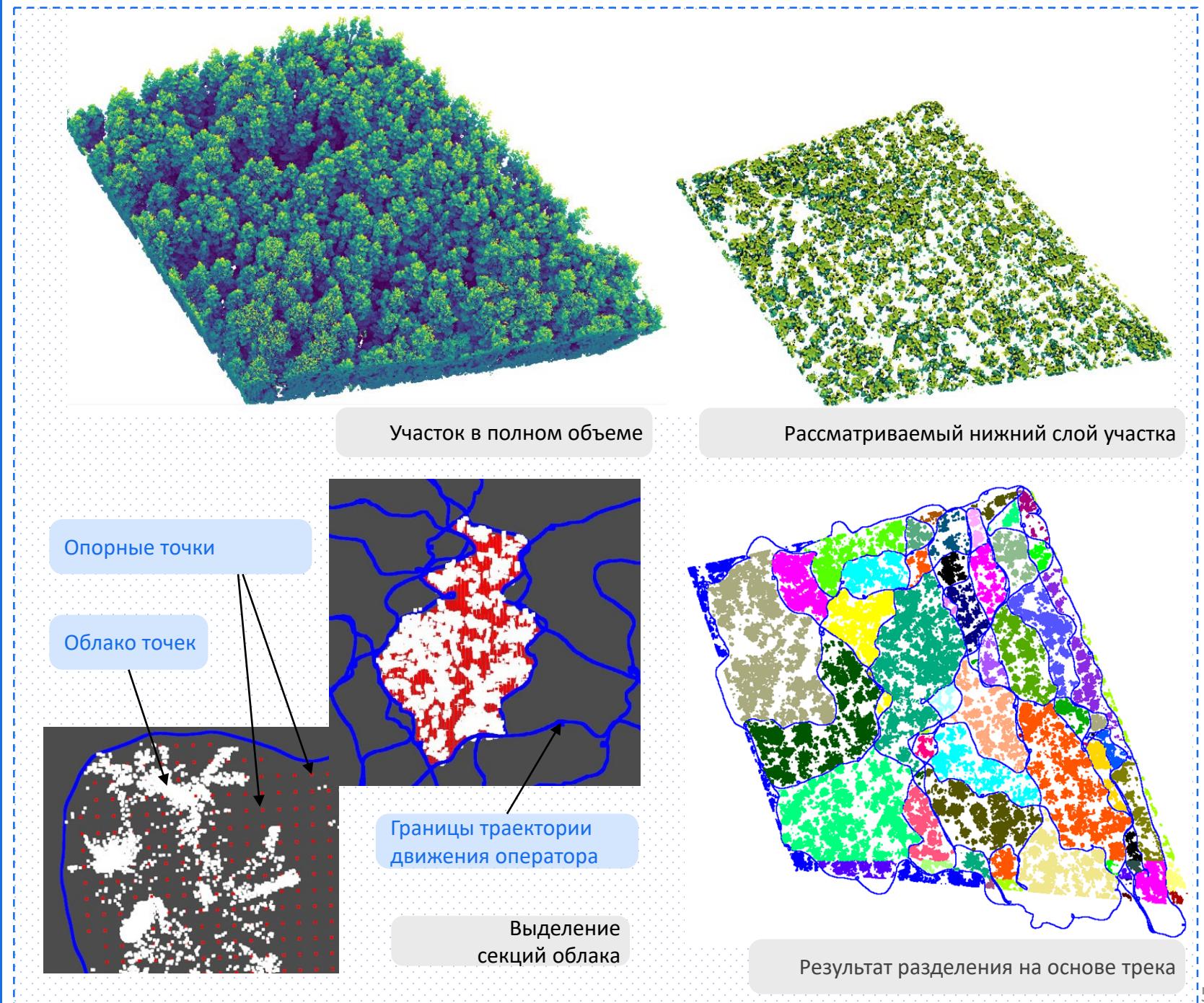
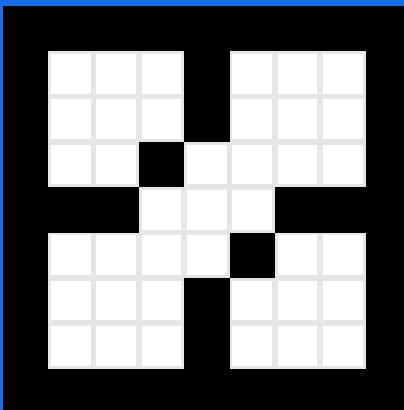
Метод на основе моделирования поверхности

1. Отражение по вертикали исходных данных
2. Моделирование поверхности в виде жесткой ткани
3. Вычисление смещения частиц ткани под действием силы тяжести
4. Изменение положение этих частиц в соответствии с внутренними силами

Методы разделения участка

Алгоритм заливки замкнутых областей (Flood Fill)

- Расставляются опорные точки с заданным интервалом
- Опорная точка определяет квадратную ячейку
- Ячейки закрашиваются с учетом границ траектории



Методы разделения участка

Разделение посредством диаграммы Вороного (Voronoi Tessellation)

- Прореживание точек
- Ядерная оценка плотности
- Построение диаграммы Вороного
- Сегментация ячеек Вороного

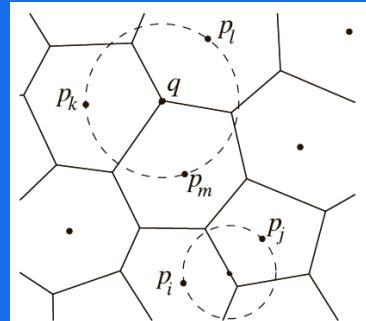
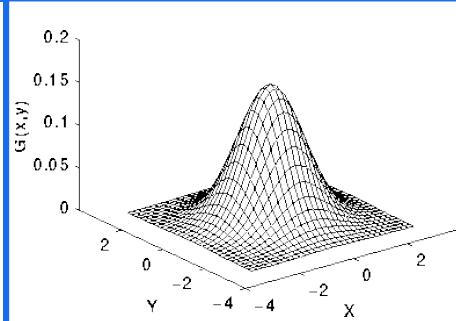
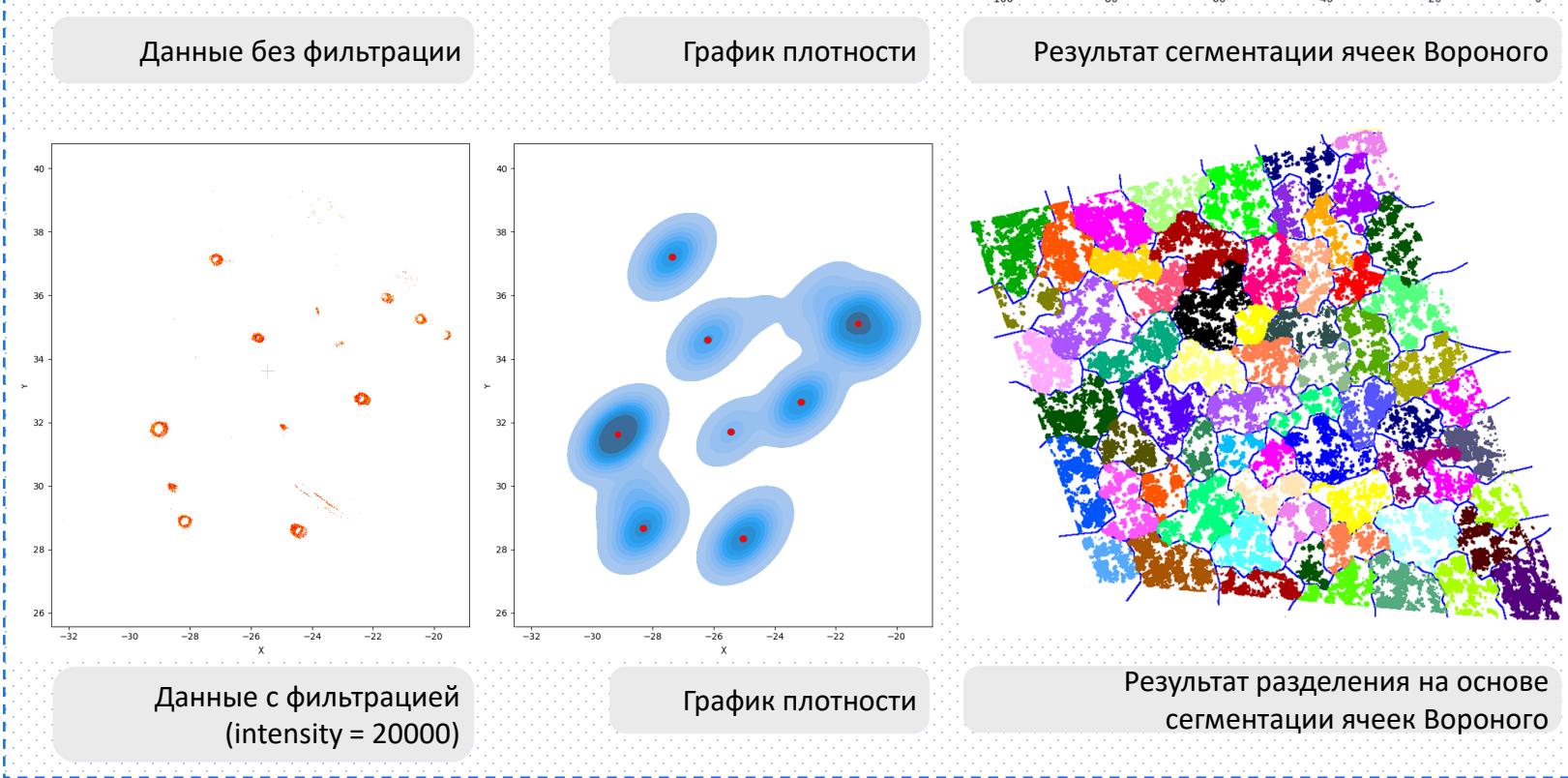
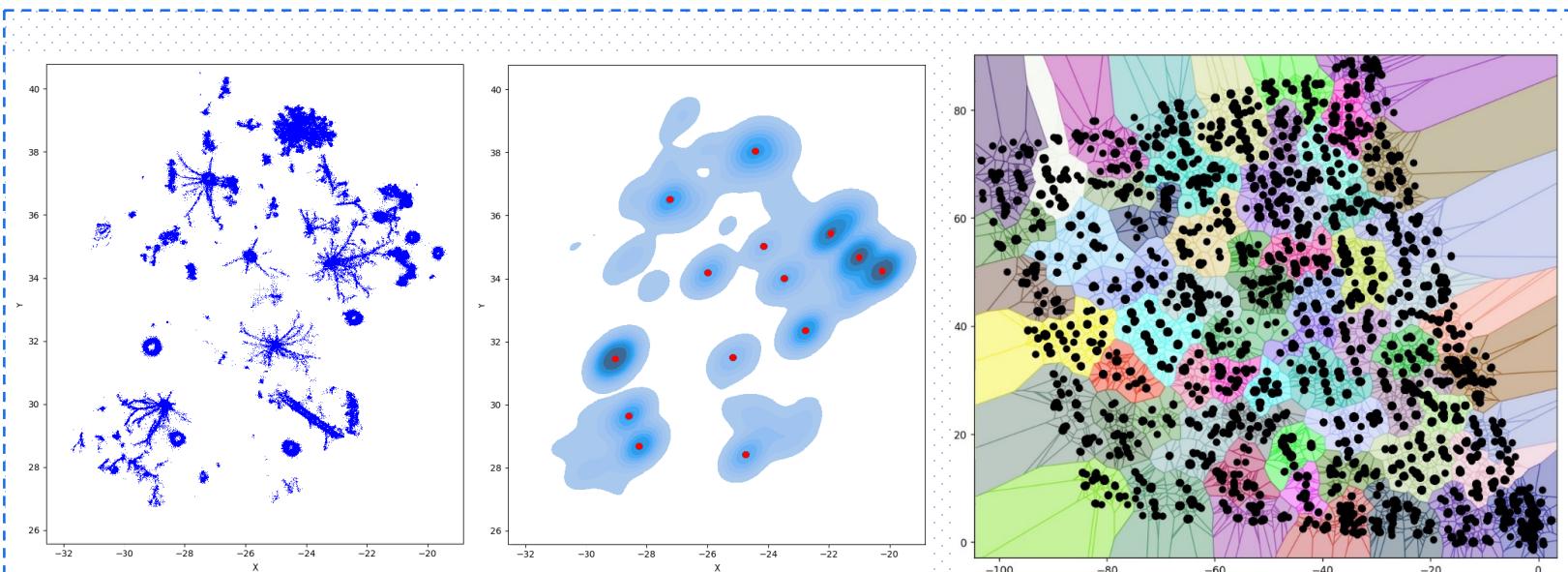


Диаграмма Вороного



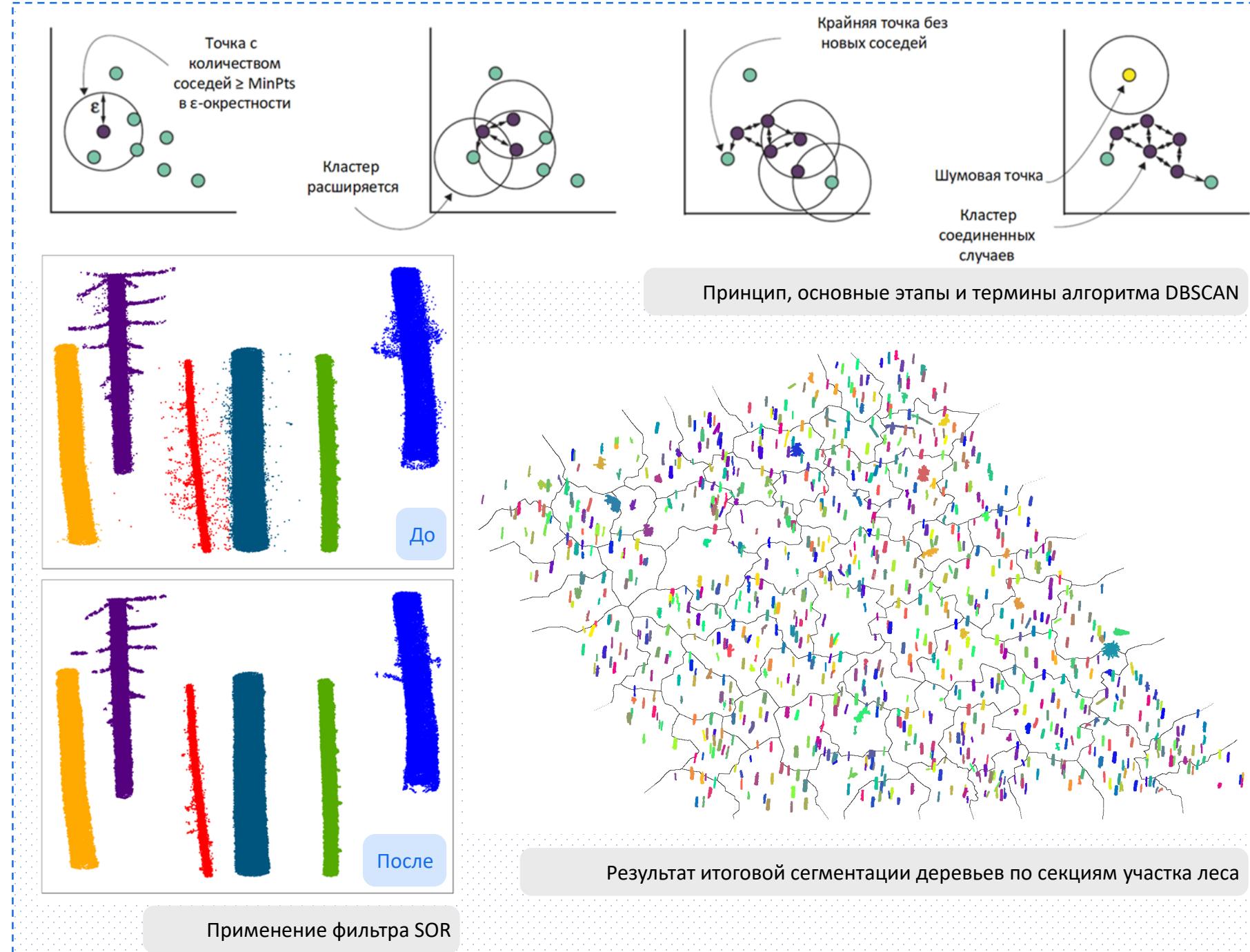
Ядро Гаусса



Процедура извлечения стволов

Последовательное применение алгоритмов кластеризации DBSCAN и HDBSCAN с применением алгоритма фильтрации SOR

1. Алгоритм DBSCAN основан на понятии связности по плотности данных
2. Алгоритм HDBSCAN – построение иерархического дерева, решение задачи поиска кластеров с наибольшей стабильностью
3. Алгоритм SOR – удаляет точки, находящиеся на удалении от своих соседей по сравнению со средним значением для облака точек



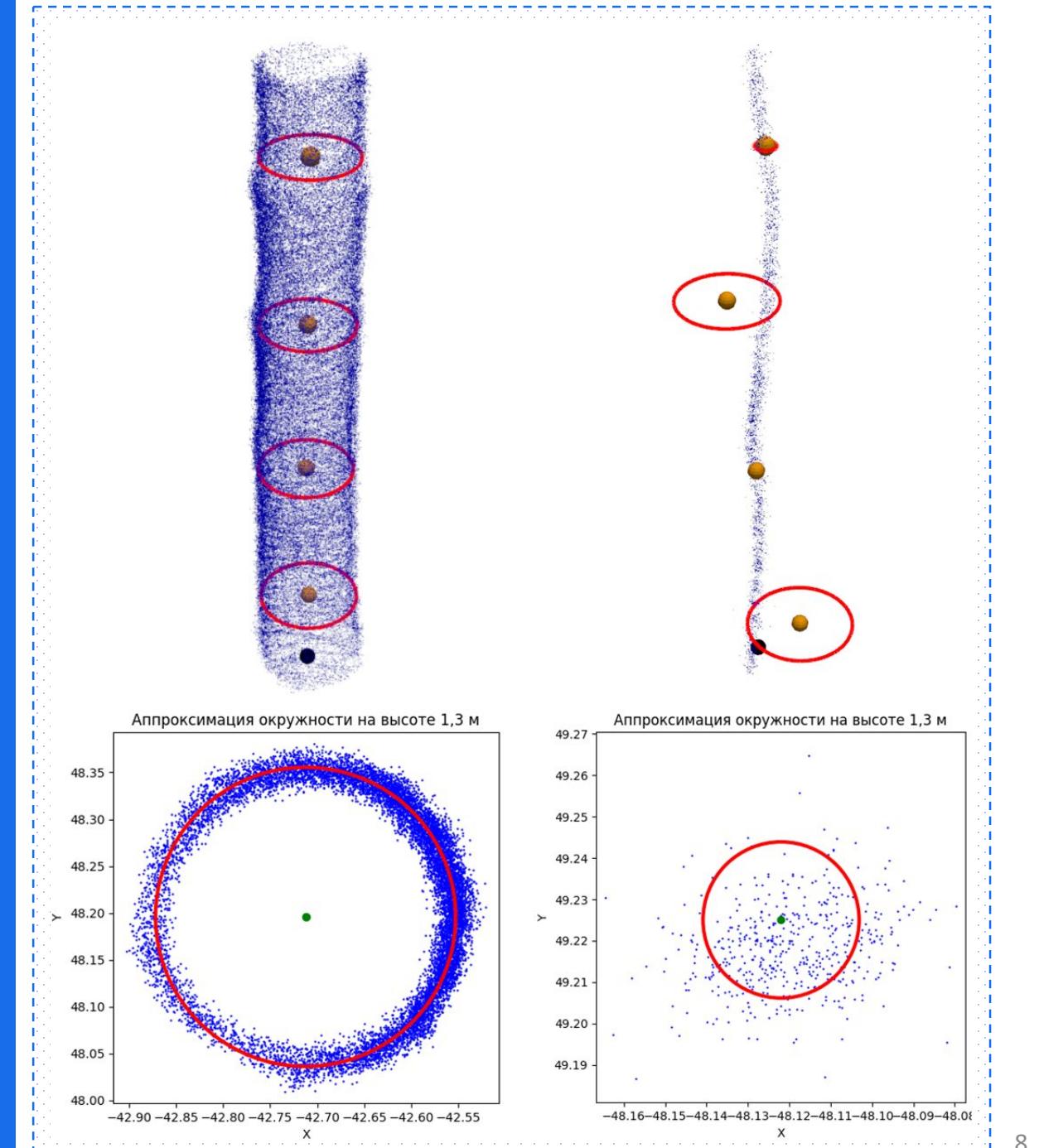
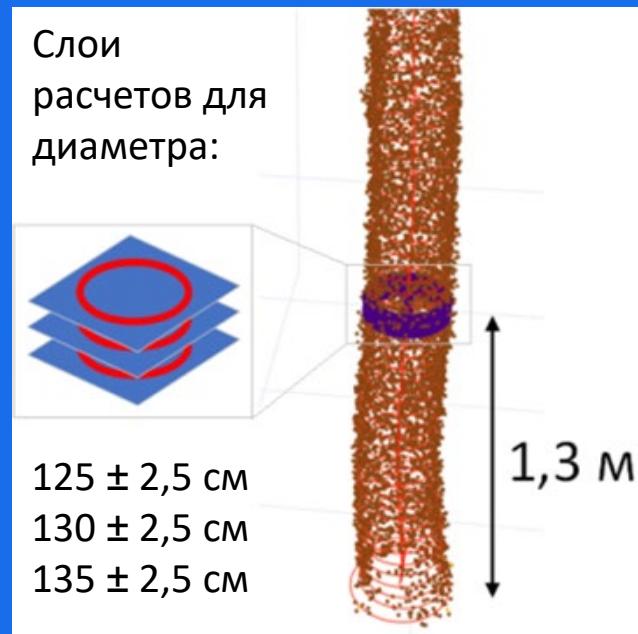
Вычисления диаметров и координат расположения деревьев

Hyper Least Squares fit – Метод гипернаименьших квадратов

Форма метода наименьших квадратов с нормализацией, устраняющей статистическую погрешность вплоть до членов шума второго порядка

Позволяет оценить диаметр и координату ствола

Включает ряд условий, ограничивающих распространение ошибок



Процедура определения координат расположения деревьев

0. Исходные данные

$p_i = \{x_i, y_i, z_i, \theta_i^1, \theta_i^2, \dots, \theta_i^m\}, i = 1 \dots n$. Каждый объект: $P = \{p_1, p_2, \dots, p_n\}^T$

1. Удаление точек поверхности земли – Remove Ground (RG)

$RG(data_{initial}) = data_{vegetation}$

2. Выделение слоя по высоте – Cut Height (CH)

$CH(data_{vegetation}, border_{down} = 0, border_{up} = 3) = data_{vegetation}|_0^3$

3. Фильтрация точек по порогу интенсивности – Filter Intensity (FI)

$FI(data_{vegetation}|_0^3, intensity_{threshold} = 7000) = data_{vegetation}|_0^3_{filtered(7000)}$

4. Сегментация участка леса на малые участки – Segment Area (SA) – Flood Fill или Voronoi Tessellation

$SA(area, N) = \{area_{small_i}\}, i = 1 \dots N$, где N – число кластеров малых участков

5. Сегментация полученного сегмента на отдельные деревья – Segment Tree (ST) – DBSCAN, HDBSCAN, SOR

$ST(area_{small}) = \{tree_j\}, j = 1 \dots M$, где M – число найденных деревьев

6. Вычисление координат расположения стволов деревьев – Coordinate Tree (CT)

$CT(tree) = [x, y]$

7. Классификация на основе глубокого обучения

Исходные данные сканера LiDAR по участку леса при наземном сканировании

1. Удаление точек поверхности земли

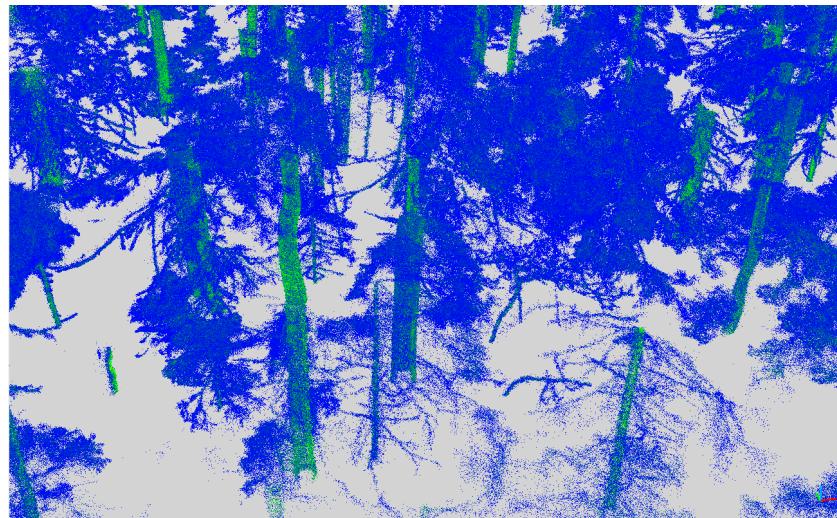
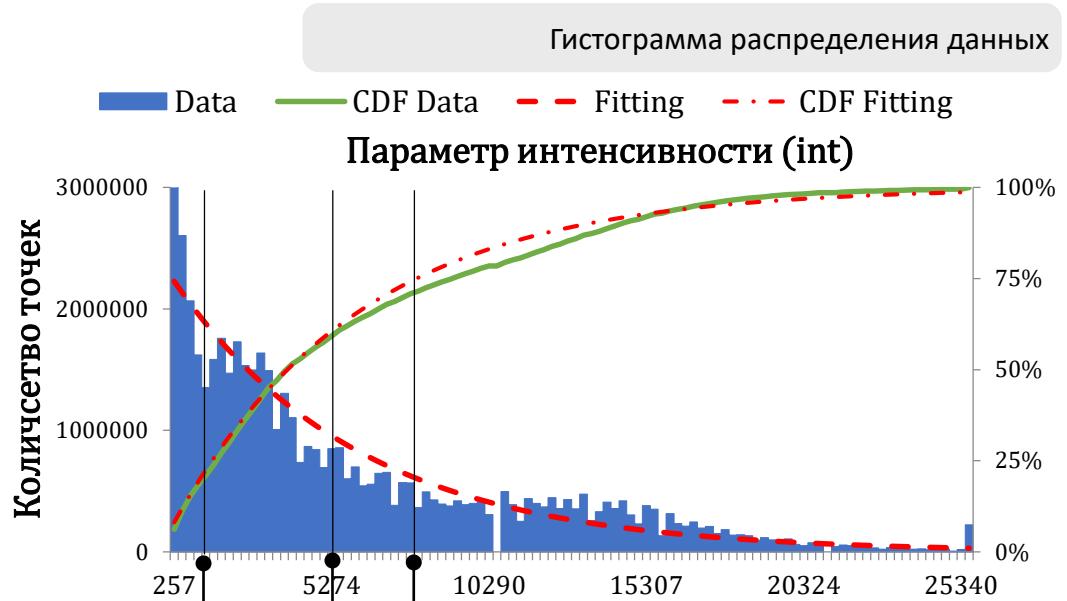
2. Удаление точек выше 3 метров над поверхностью земли

Сегментация локального участка леса на небольшие секции

Сегментация полученного сегмента на отдельные деревья

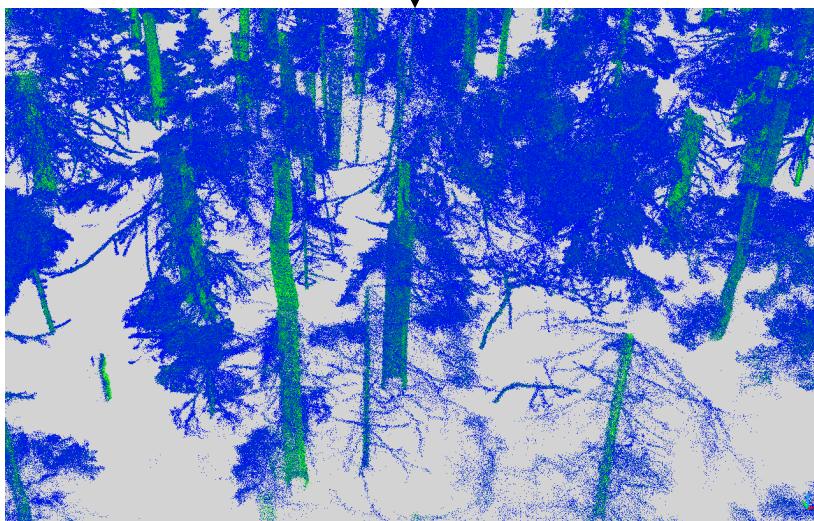
Вычисление координат расположения стволов деревьев

Фильтрация данных

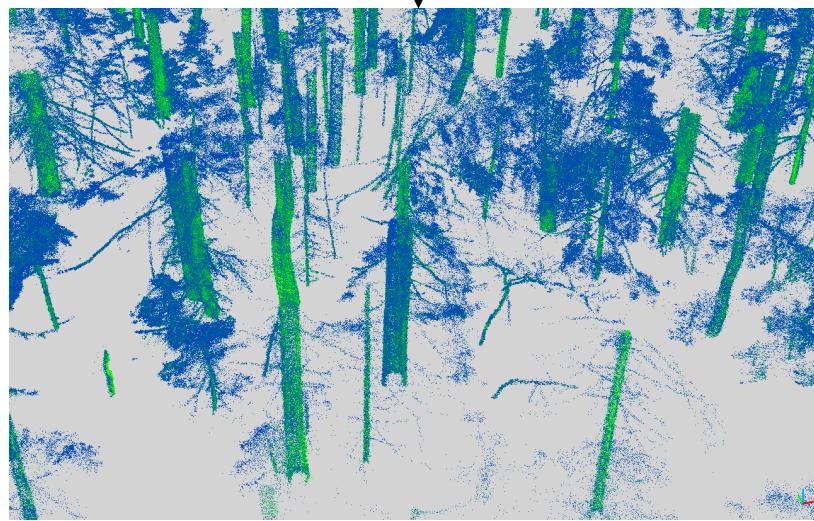


$data = data_{vegetation}|_0^3$

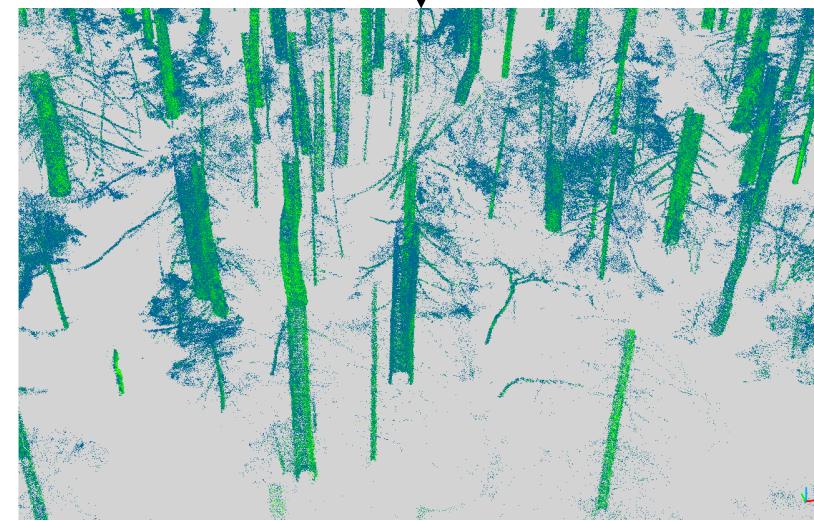
Данные слоя до 3 метров



$data_{filtered(int=1000)}$



$data_{filtered(int=5000)}$

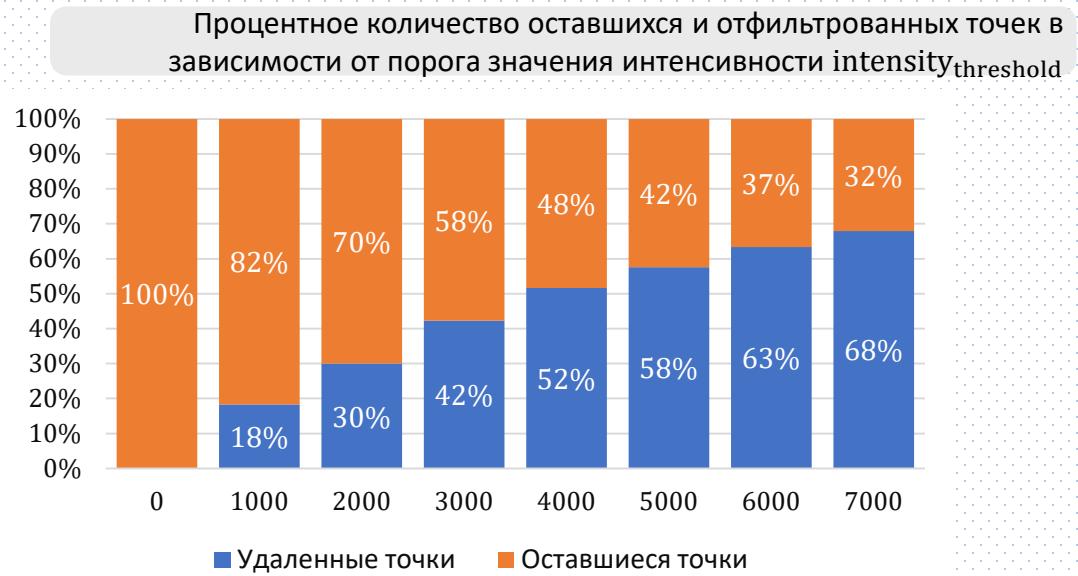
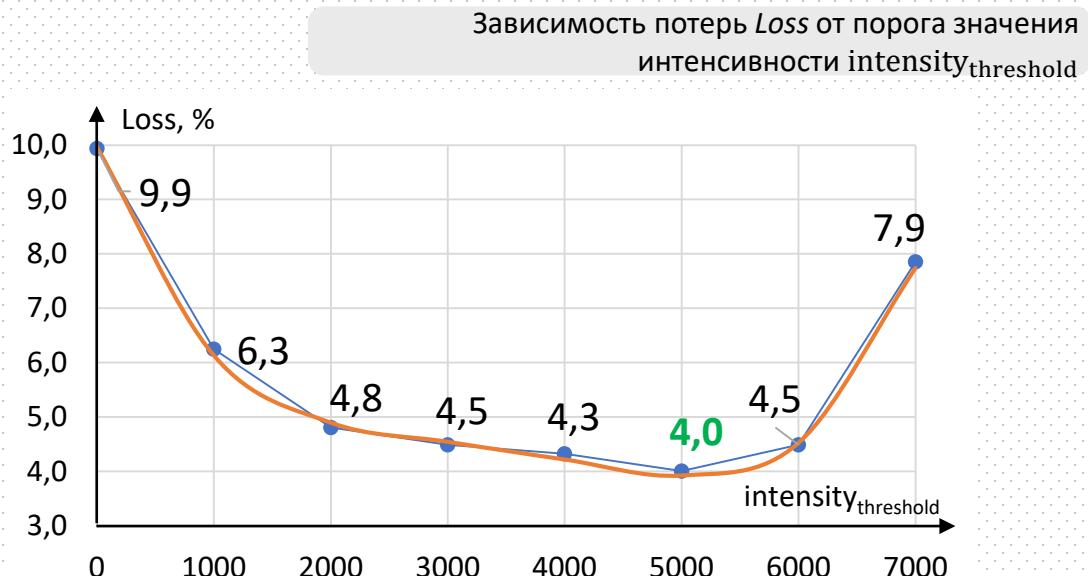


$data_{filtered(int=7000)}$

Анализ фильтрации данных

$Loss = \frac{NF}{NP} * 100\%$, – потери при обнаружении

NF – количество найденных деревьев
NP – количество деревьев, указанных в паспорте участка



Вероятность обнаружения дерева в предложенной системе можно представить в виде зависимости:

$$P_{detection} = Tree(pts_{stump}, pts_{noise}, intensity_{avg}, RadDist_{ClosestTree})$$

pts_{stump} – количество точек ствола

$intensity_{avg}$ – среднее значение интенсивности точек ствола

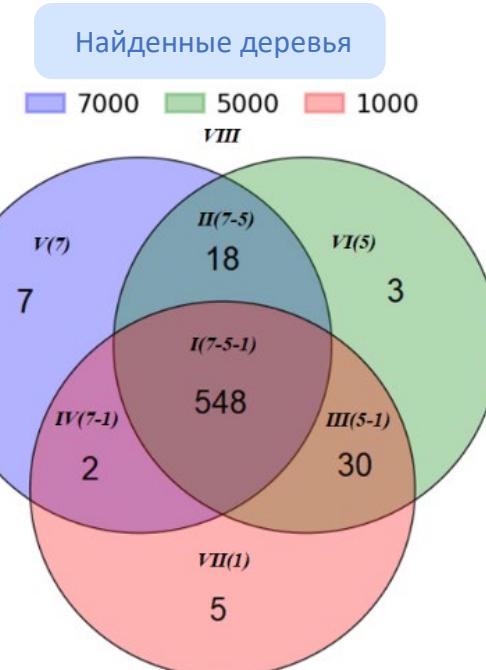
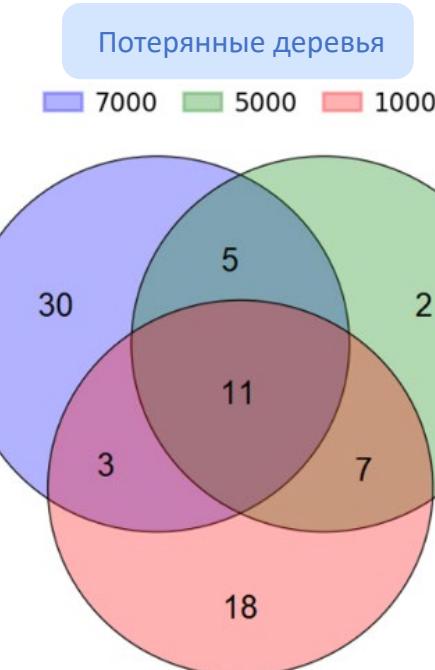
pts_{noise} – количество точек шума

$RadDist_{ClosestTree}$ – факт, когда сумма оценки радиусов рассматриваемого дерева и ближайшего дерева к нему меньше расстояния между ними

Анализ классов

Таблица 1 – Характеристика классов

Класс	Кол-во деревьев	median(pts _{stump})	median(pts _{noise})
I (7-5-1)	548	18995	48695
II (7-5)	18	23522	87968
III (5-1)	2	5859	33104
IV (7-1)	30	3457	51904
V (7)	7	4010	32736
VI (5)	3	1384	62023
VII (1)	5	3882	35597
VIII	11	564	41865



Диаграммы Венна при разных настройках intensity_{threshold}

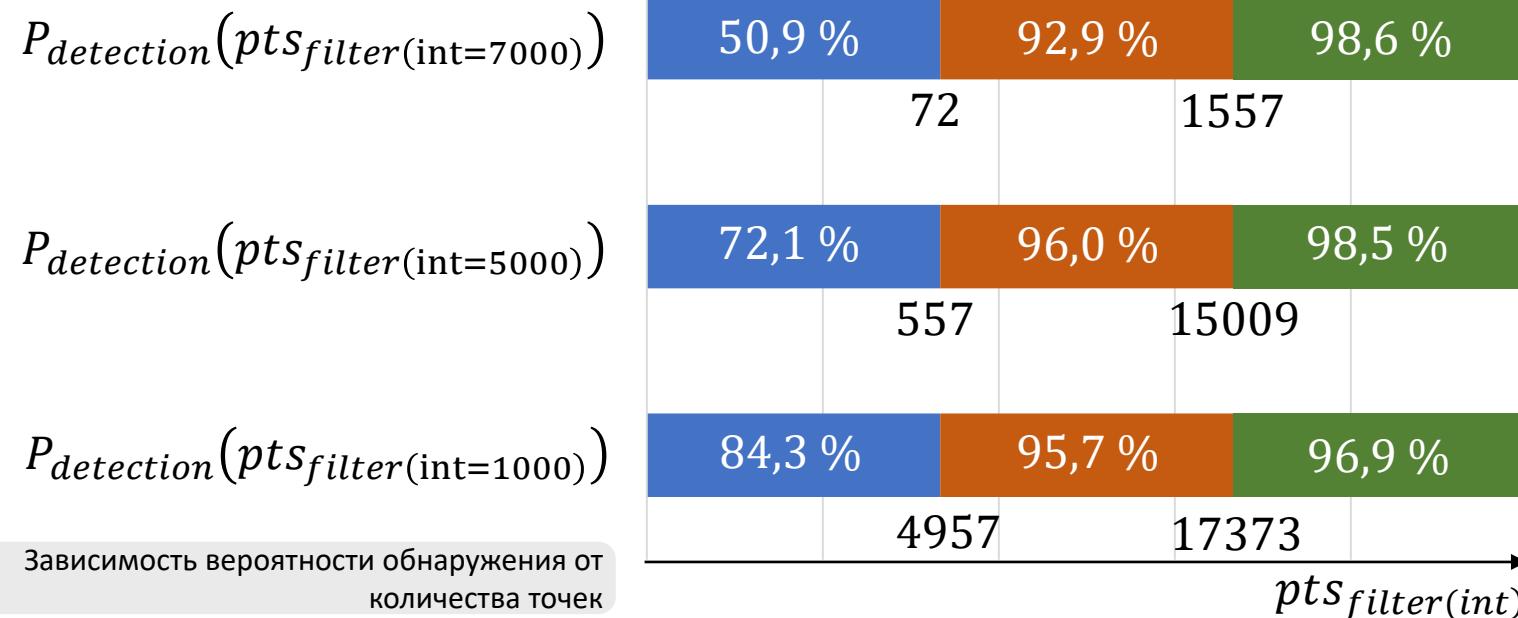


Таблица 2 – Проценты потерь точек после фильтрации

intensity _{threshold}	Потеря точек шума	Потеря точек ствола
7000	95,34 %	92,88 %
5000	62,94 %	18,46 %
1000	15,84 %	2,72 %

Глубокое обучение на облаках точек

Проблема: требуется упорядоченный формат данных

Следующие сети работают непосредственно с плотными облаками точек

PointNet

Первооткрыватель в этом направлении

3 ключевых модуля:

- слой Max Pooling
- две сети трансформации с MLP
- две сети для предсказания с обученной матрицей T-Net

PointNet++

Базируется на использовании классической сети

Добавлены слои, которые умеют агрегировать информацию, учитывать неоднородную плотность точек в данных

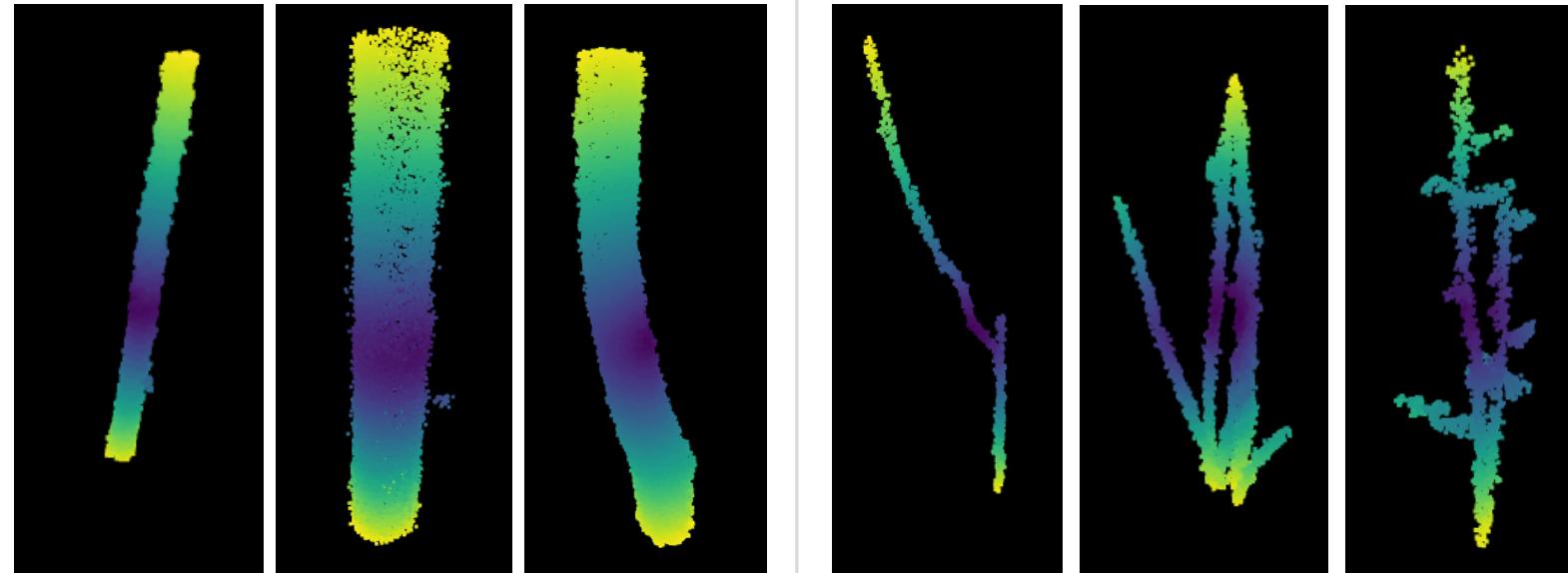
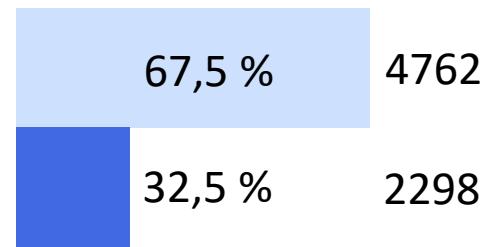
CurveNet

Получение форм трехмерной геометрии

Последовательности точек представляются в виде кривых, этап обучения сети предполагает группировку полученных кривых

Обучение моделей нейросетей

Распределение
классов



Примеры **паспортных** деревьев

Примеры **лиших** объектов

Алгоритм сжатия до N точек:
farthest point sampling

Таблица 3 – Результаты обучения универсальной модели в зависимости от количества точек на облако

Итеративный
выбор самой
дальней точки от
текущей.

	Количество точек на облако точек (одно дерево)				
	128	256	512	1024	2048
PointNet++	0,938	0,960	0,969	0,959	0,952
CurveNet	0,940	0,962	0,964	0,959	0,946

Обучение моделей и результаты

Таблица 4 – Результаты обучения моделей для разных порогов фильтрации

	Модели для разных порогов фильтрации			
	<i>int1000</i>	<i>int5000</i>	<i>int7000</i>	<i>Общая модель</i>
Обучающая выборка	952	866	736	7060
accuracy	0,920	0,958	0,956	0,969
recall	0,920	0,947	0,935	0,963
precision	0,912	0,951	0,935	0,966

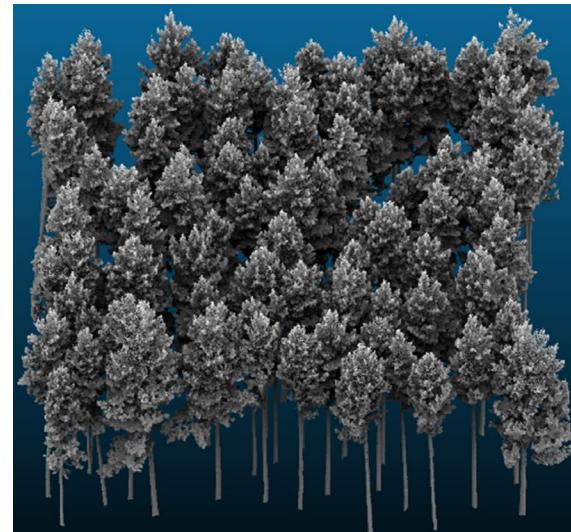
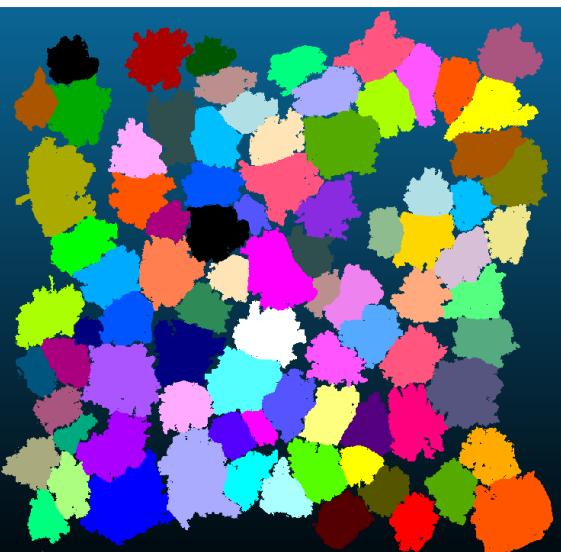
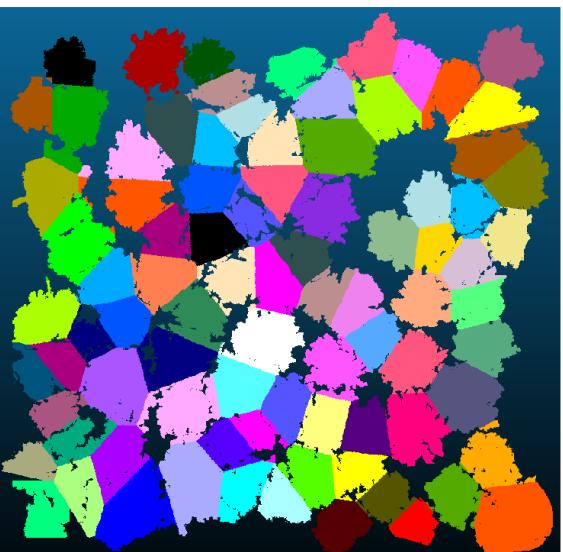
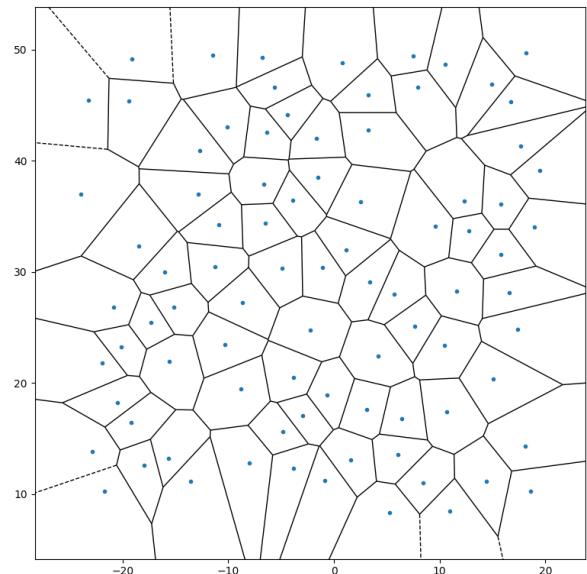
Таблица 5 – Результаты удаления лишних деревьев

intensity threshold	Кол-во найденных деревьев	Кол-во лишних деревьев	Кол-во подтвержденных найденных деревьев	Кол-во подтвержденных лишних деревьев
7000	575	160	526	154
5000	599	267	554	261
1000	585	367	530	359

Удаление до 97,3% лишних объектов, увеличение потерь Loss до 15,7%.

Анализ сразу трех результатов предсказания – Loss = 4,0%, при удалении 76,2% лишних объектов.

Использование координат



Пример использования диаграммы Вороного для разделения облака точек участка в сравнении с ручной разметкой

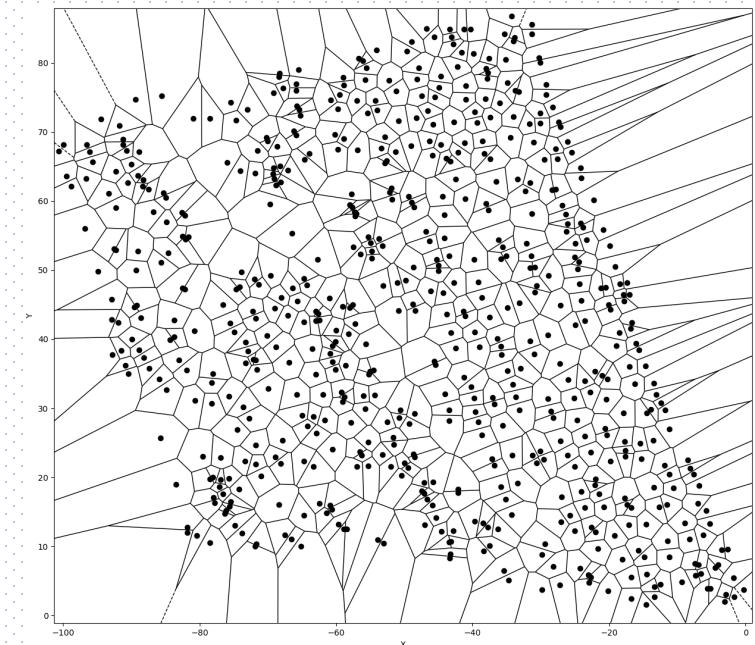
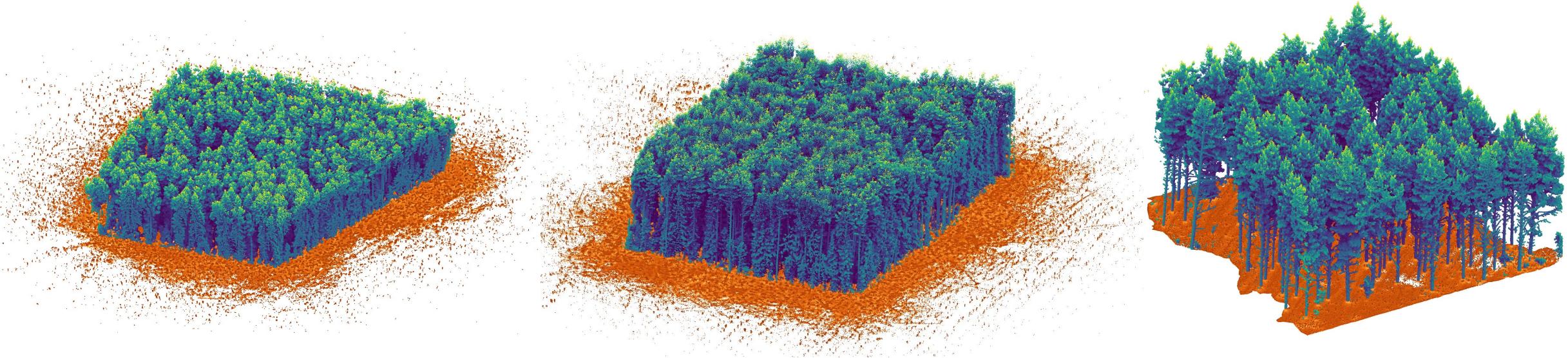


Диаграмма Вороного на найденных координатах

Координаты использованы для автоматического разделения участка на подобласти, каждая из которых содержит только одно дерево.

Диаграмма Вороного – основа для сегментации деревьев.

Верификация методики



Первый участок лесной местности

- 624 дерева
- 6 пород деревьев (сосна, ель, рябина, осина, дуб, береза)
- 0,63 гектар

Полевые данные:

- координаты
- диаметры

Второй участок лесной местности

- 1001 дерево
- 6 пород деревьев (сосна, ель, рябина, осина, дуб, береза)
- 0,92 гектар

Полевые данные:

- координаты
- диаметры

Третий участок лесной местности

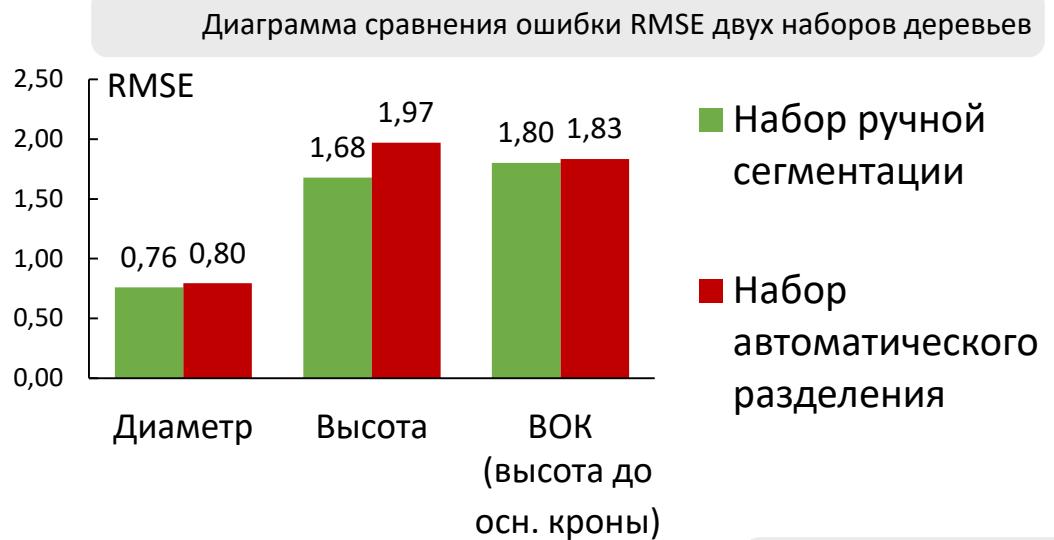
- 86 дерева
- 1 порода дерева (сосна)
- 0,26 гектар

Полевые данные:

- | | |
|--------------|----------------------------|
| — координаты | — высоты |
| — диаметры | — высоты до оснований крон |

Оценка параметров деревьев

Таблица 6 – Результаты ошибок измерений двух наборов деревьев для третьего участка лесной местности



	Длина дерева	Объем выпуклой оболочки кроны	Площадь выпуклой оболочки кроны
MSE	0,62	1758,22	1210,36
RMSE	0,79	41,93	34,79
Ошибка, %	1,12	26,20	19,07

Таблица 7 – Результаты ошибок измерений двух наборов деревьев относительно полевых измерений для третьего участка лесной местности

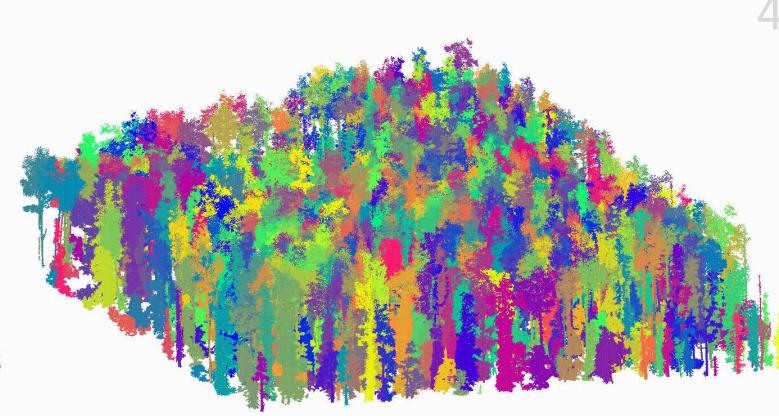
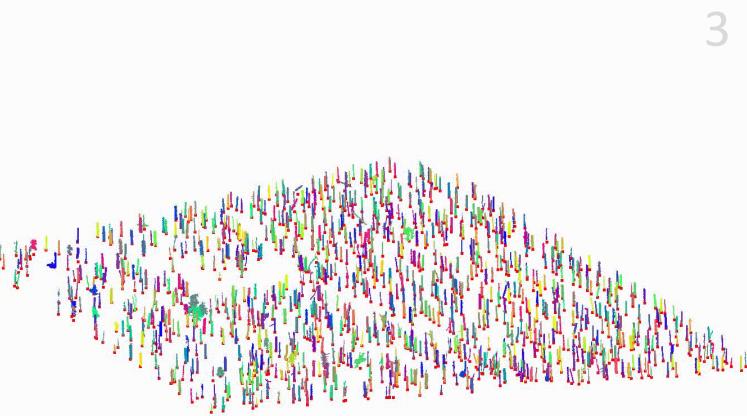
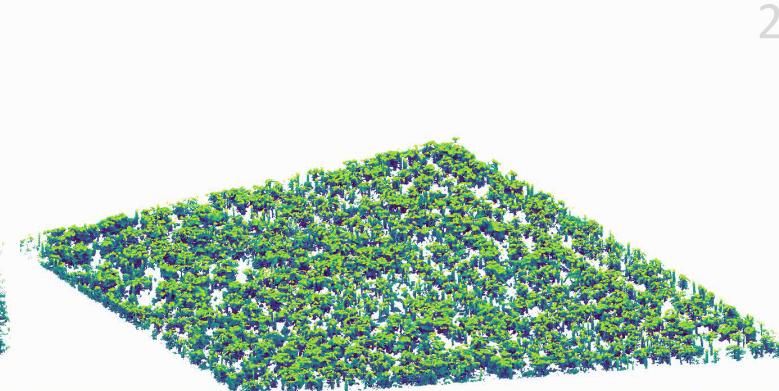
Для первого и второго участка:

RMSE оценки диаметров 3,40 см и 4,78 см соответственно.

	Набор деревьев ручной сегментации			Набор деревьев автоматического разделения		
	Диаметр	Высота	Высота до осн. кроны	Диаметр	Высота	Высота до осн. кроны
MSE	0,58	2,82	3,24	0,63	3,87	3,36
RMSE	0,76	1,68	1,80	0,80	1,97	1,83
Ошибка, %	-1,26	0,89	-8,77	-0,92	2,06	-8,86

Результаты

Визуализация облаков точек на разных этапах обработки



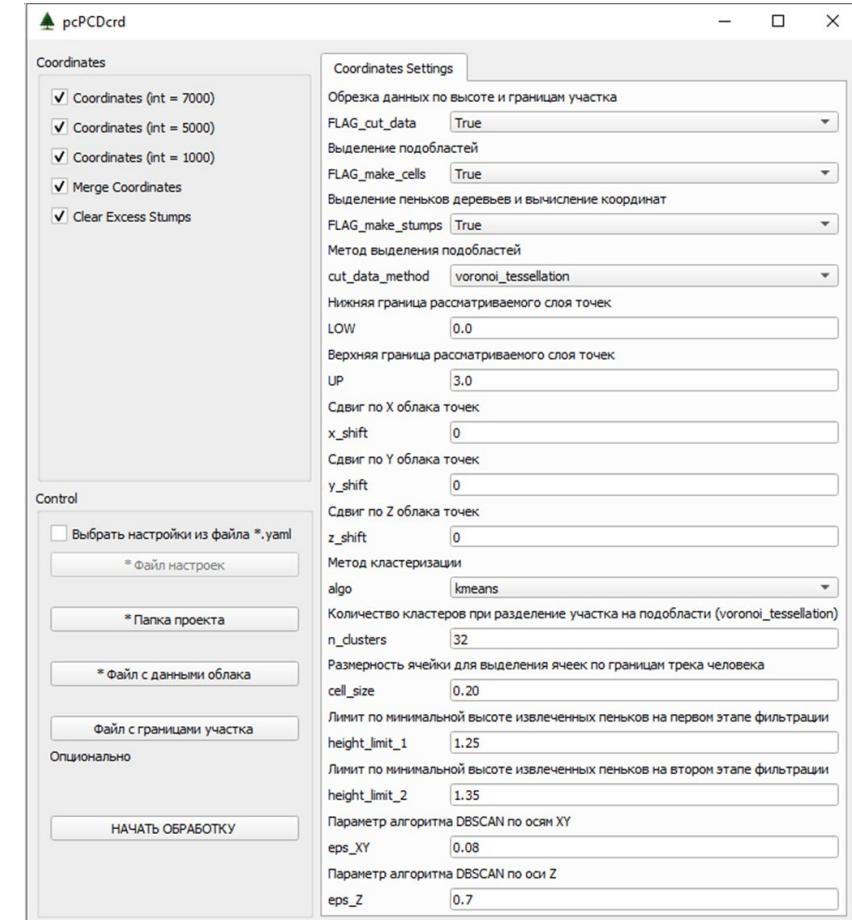
В результате проведенной работы:

1
Разработана методика определения местоположений и характеристик деревьев на участке леса

2
Рассмотрены методы и алгоритмы, рекомендуемые для обработки данных LiDAR лесных местностей

3
Методы были адаптированы для обработки больших массивов точечных данных

4
Найдено эффективное решение задачи поиска и идентификации стволов деревьев при помощи методов сегментации, удаления шумов и методов глубокого обучения



Пользовательский интерфейс демонстрационного стенда

Заключение

В результате проведенной работы:

Результаты исследования дают возможность подготовить основу для расчета всей необходимой информации структуры лесного участка и его количественного наполнения для построения цифровой модели местности

Описаны и формализованы основные этапы обнаружения стволов деревьев и определения их координат

Произведена апробация методики

1. Первый участок

4,0% потерь, при удалении 76,2% лишних объектов

2. Второй участок

5,5% потерь, при удалении 84,4% лишних объектов

Отобраны подходящие архитектуры нейронных сетей, обучены модели PointNet++ и CurveNet

Точность классификации: 97%

Точности оценки параметров деревьев схожи в сравнении с ручной сегментацией

Методика предлагает надежное обнаружение стволов деревьев

Информация о местоположениях деревьев поможет в разметке и сборе данных для создания наборов данных

20

9

10

20