

Использование нормализованных данных Sentinel-2 и материалов распределенных на больших территориях наземных обследований для оценки запасов стволовой древесины лесов

¹Богодухов М.А.

¹Барталев С.А.

¹Жарко В.О.

(1) ИКИ РАН

Двадцать вторая конференция молодых ученых
«ФУНДАМЕНТАЛЬНЫЕ И ПРИКЛАДНЫЕ КОСМИЧЕСКИЕ ИССЛЕДОВАНИЯ»

21-23 апреля, 2025

Москва, ИКИ РАН

Актуальность:

- Для формирования по спутниковым данным Дистанционного Зондирования Земли (ДЗЗ) карт характеристик леса, в том числе необходимых для оценки их запасов углерода, в национальном масштабе, используются данные ДЗЗ со средним пространственным разрешением (100-500 метров на пиксель).
- Однако, наиболее актуальные данные наземных обследований лесов, доступные для их валидации, характеризуют лесной покров на уровне Пробных Площадей (ПП) диаметром около 25-30 метров, что затрудняет их прямое сопоставление.
- Следовательно, актуальным является создание карт характеристик лесов высокого разрешения с размером пикселя 10-30 метров, основанных на материалах ПП и ДЗЗ, как для возможности валидации продуктов обработки данных ДЗЗ среднего пространственного разрешения, так и в виде отдельного продукта.

Цель работы:

- Исследование возможностей определения запаса стволовой древесины на основе данных Sentinel-2 и материалов ПП

Используемые данные:

Обучающая выборка:

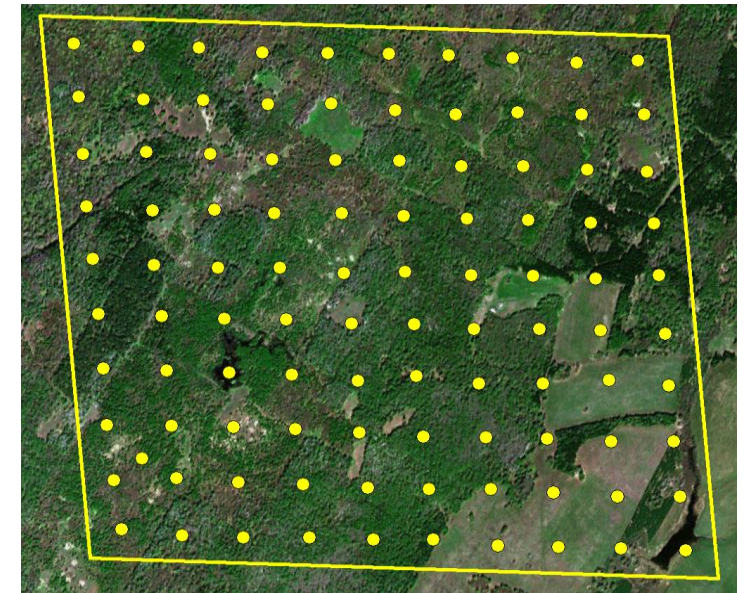
- Объединенные материалы наземных обследований запаса стволовой древесины лесов на уровне регулярной сети ПП, заложенных на территории более 100 Тестовых Полигонов (ТП), расположенных по всей территории России;

Признаки для обучения модели (Random Forest):

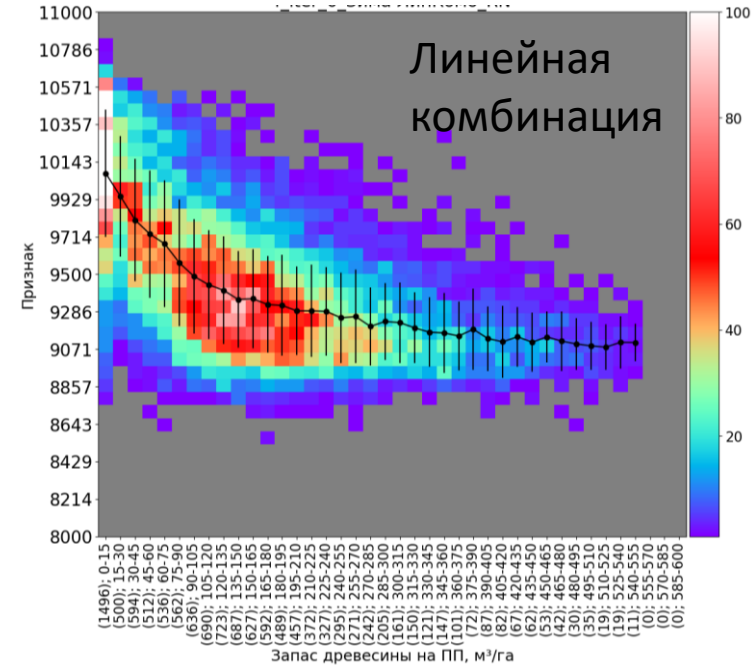
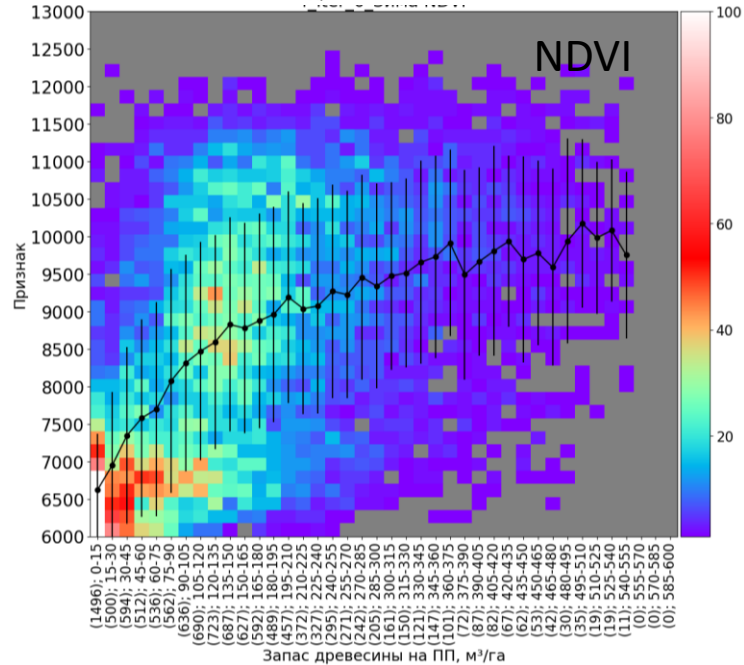
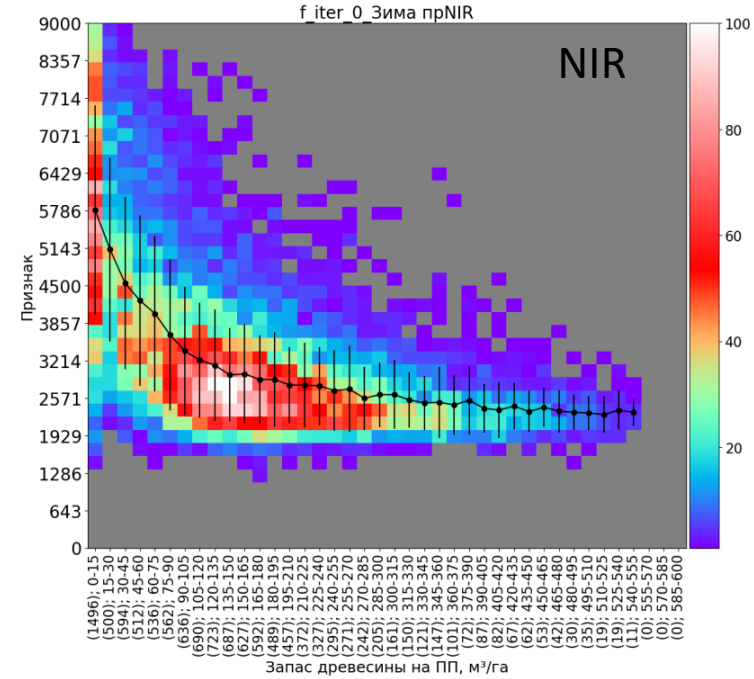
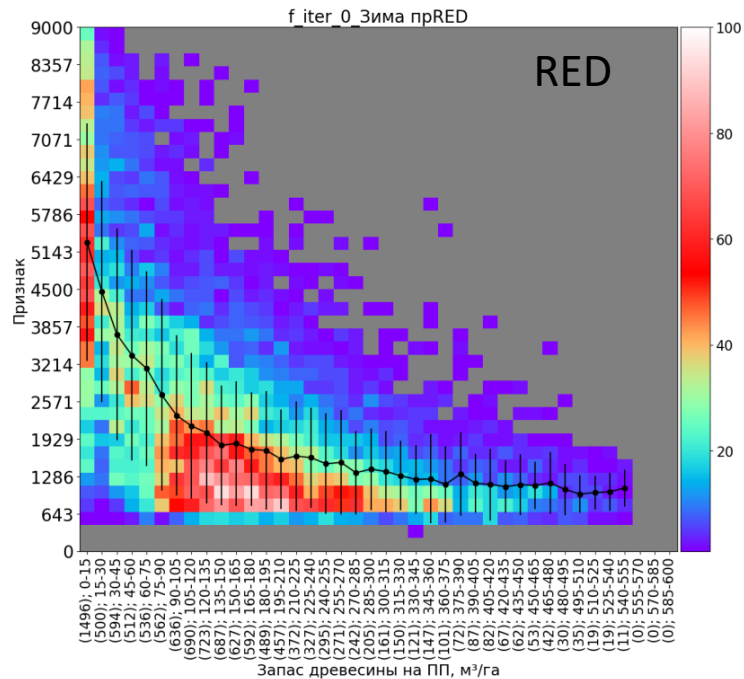
- Нормализованный по углу солнца и территории зимний композит Sentinel-2:
 - RED, NIR
 - NDVI
 - Линейная комбинация RED, NIR для учета различия между хвойными и листопадными;
- Летние композитные изображения Sentinel-2 в пике сезона по NDVI:
 - BLUE, GREEN, RED, NIR, SWIR1, SWIR2, NDVI



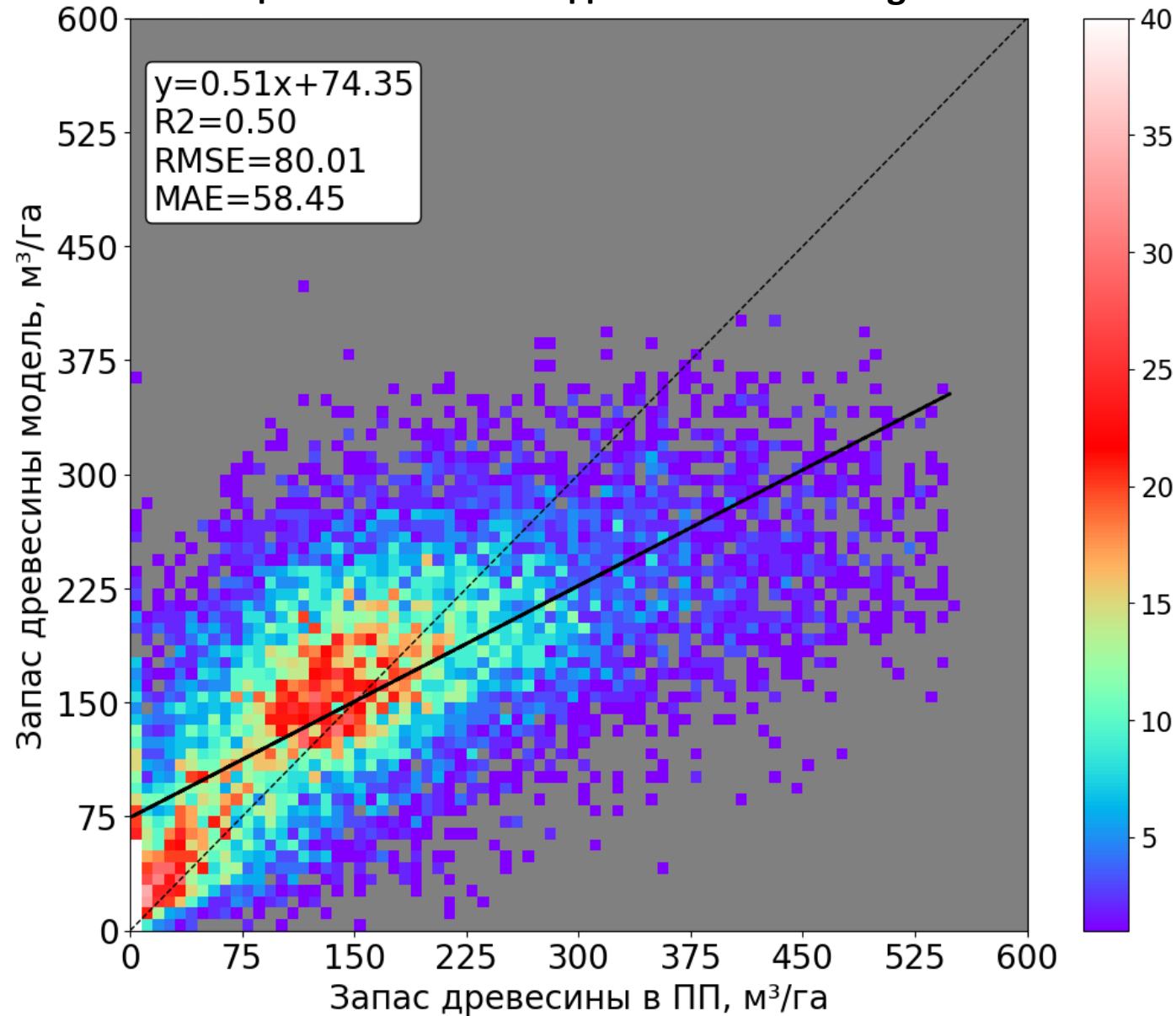
Иллюстрация распределения ПП в ТП



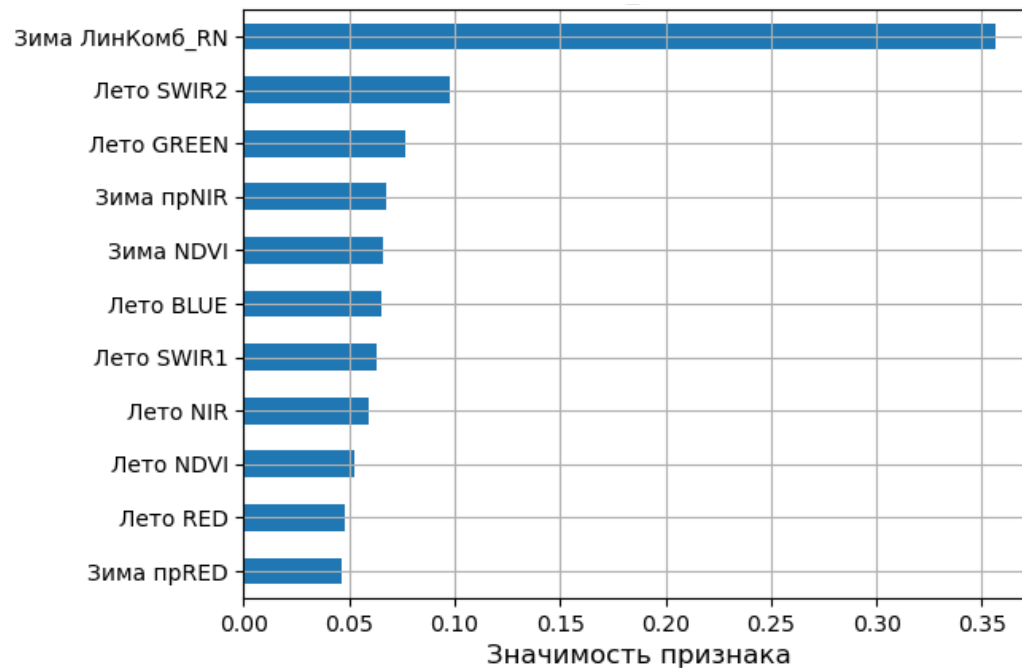
Зависимость между запасом и зимними признаками



Оценка точности модели по Ouf-Of-Bag



Значимость признаков



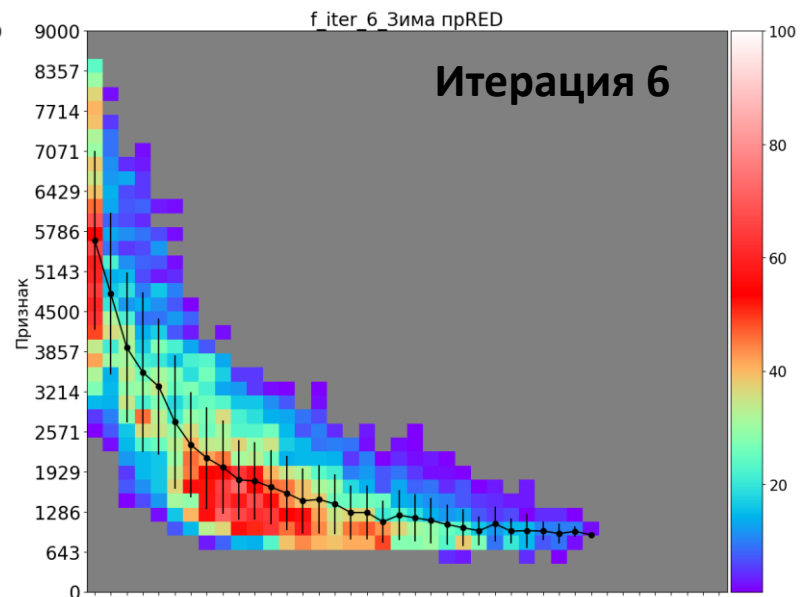
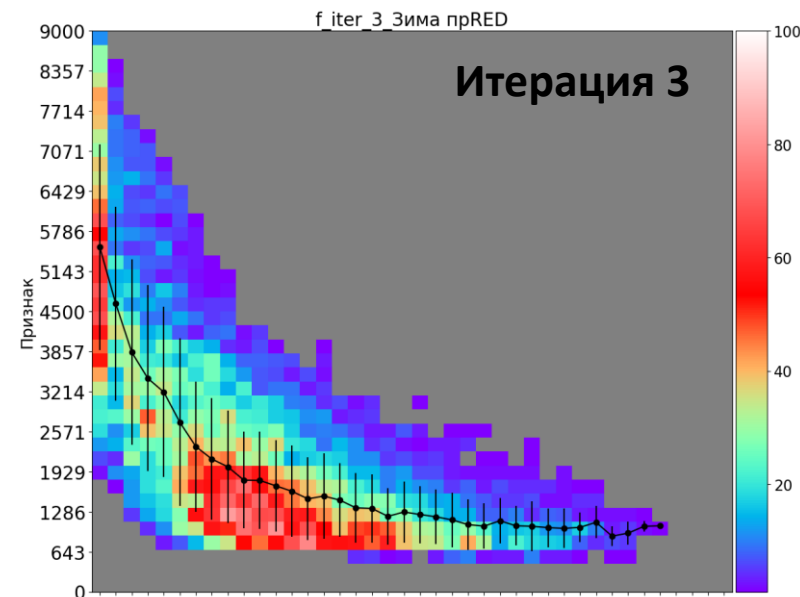
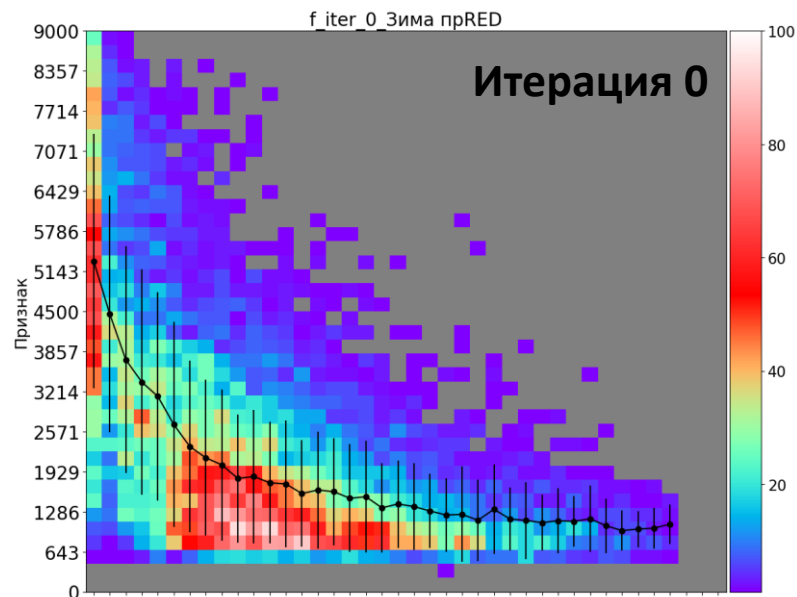
Итеративный фильтр

Для фильтрации выборки был предложен следующий итеративный подход фильтрации:

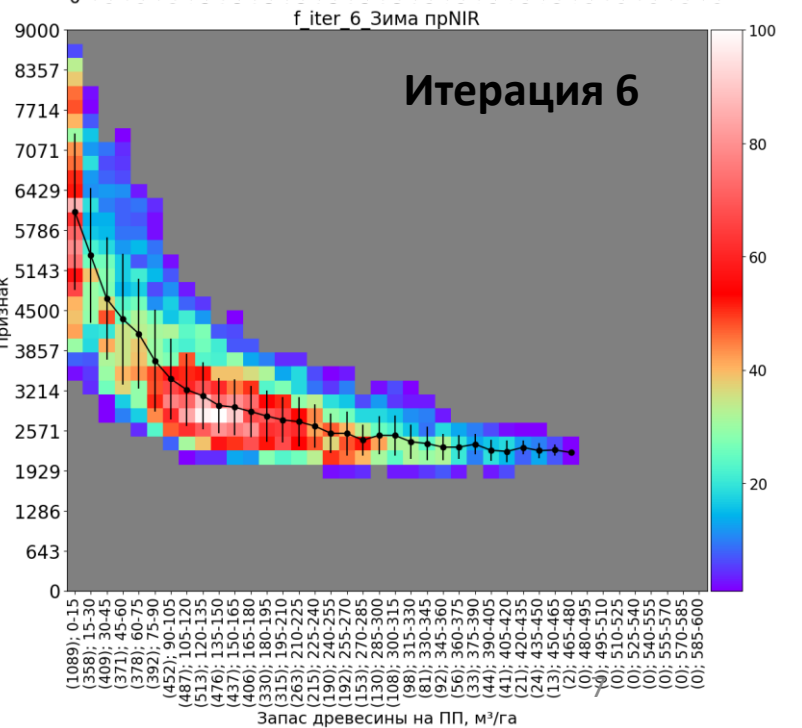
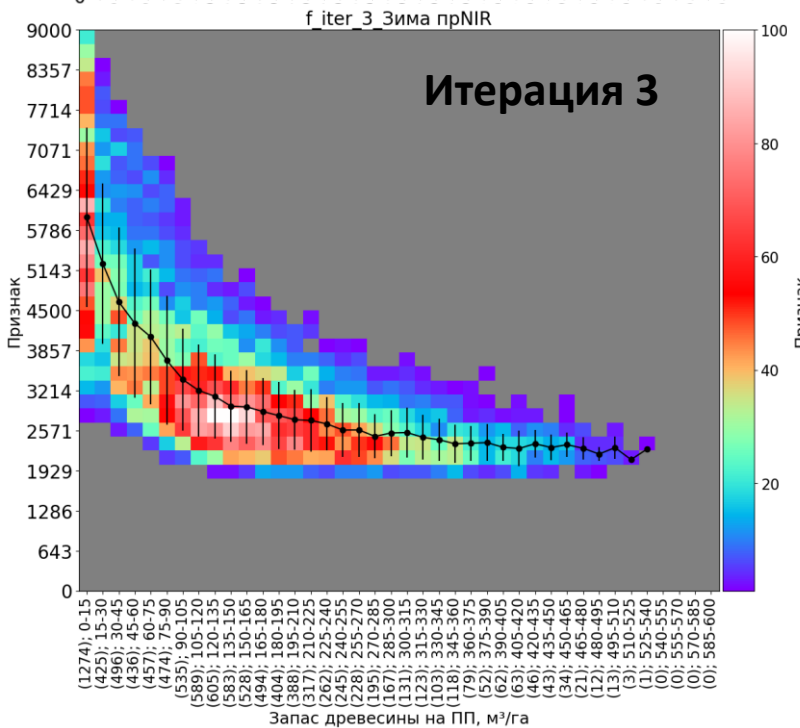
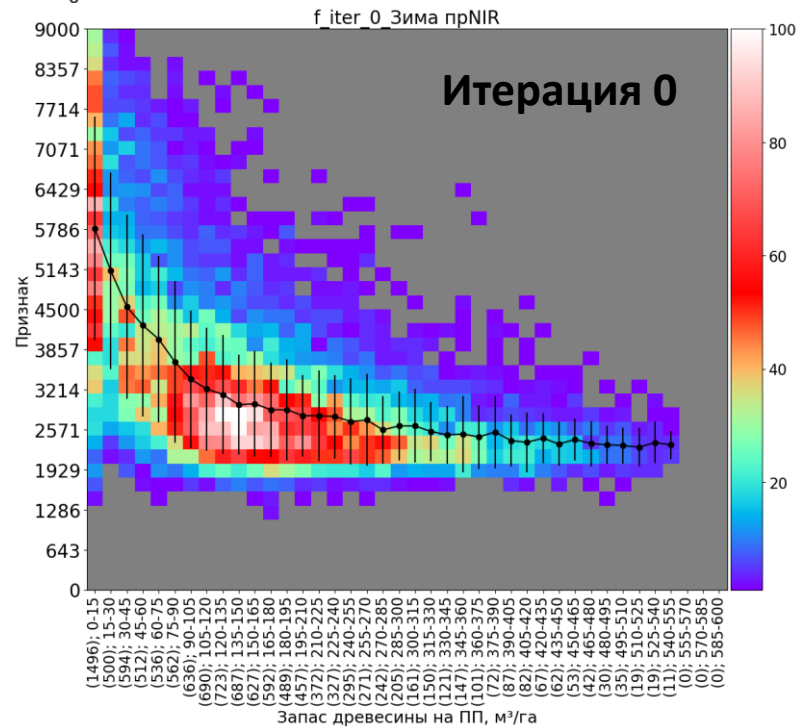
1. Диапазон значений запаса стволной древесины в ПП разбивается на диапазоны по $10\text{м}^3/\text{га}$;
2. Фильтрация по зимним признакам RED, NIR для каждого диапазона запаса:
 - 2.1. Для каждого диапазона запаса определяется 1-й и 99-й перцентили в зимних каналах RED, NIR;
 - 2.2. Отбрасываются значения, выходящие за пределы с 1-го по 99-й перцентиль в обоих каналах в соответствующих диапазонах запаса.
3. Фильтрация по прогнозу модели:
 - 3.1. Обучение модели по отфильтрованной выборке и прогноз значений запаса (методом out-of-bag);
 - 3.2. Для каждого диапазона запаса:
 - 3.2.1. Считается относительное отклонение прогноза модели от истинных значений запаса в ПП по модулю;
 - 3.2.2. Определяется 98-й перцентиль относительного отклонения;
 - 3.2.3. Отбрасываются ПП, в которых относительное отклонение превышает 98-й перцентиль из п.3.2.2.
4. Повторяются пункты 1-3 до тех пор, пока не будет достигнута оптимальная точность модели

Итеративная фильтрация на основе RED и NIR; ошибок out-of-bag

RED

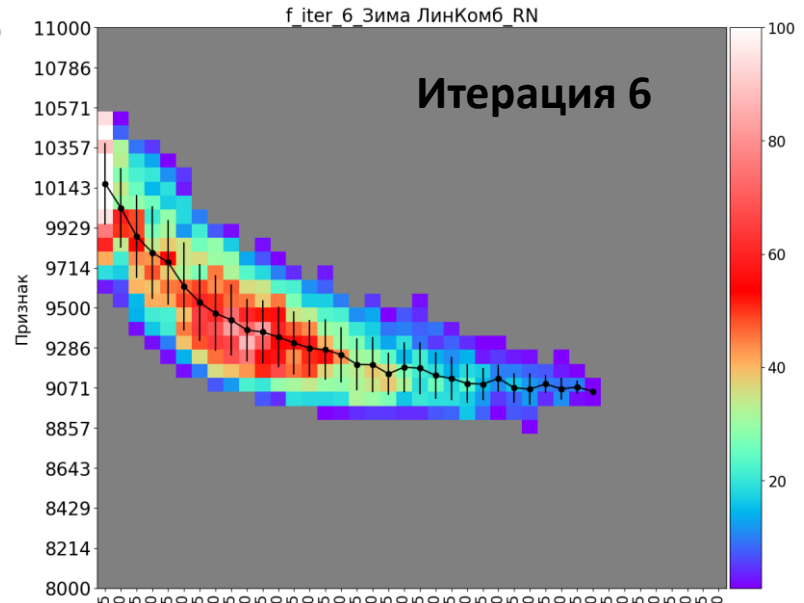
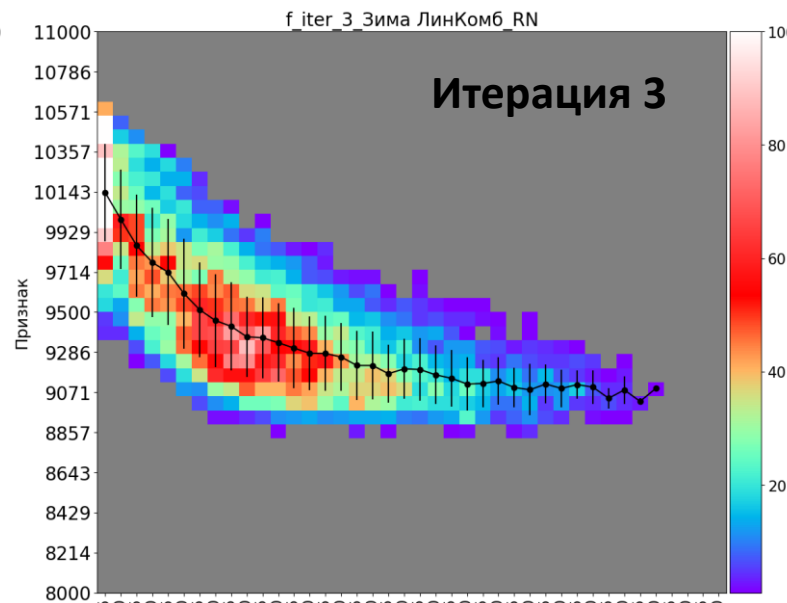
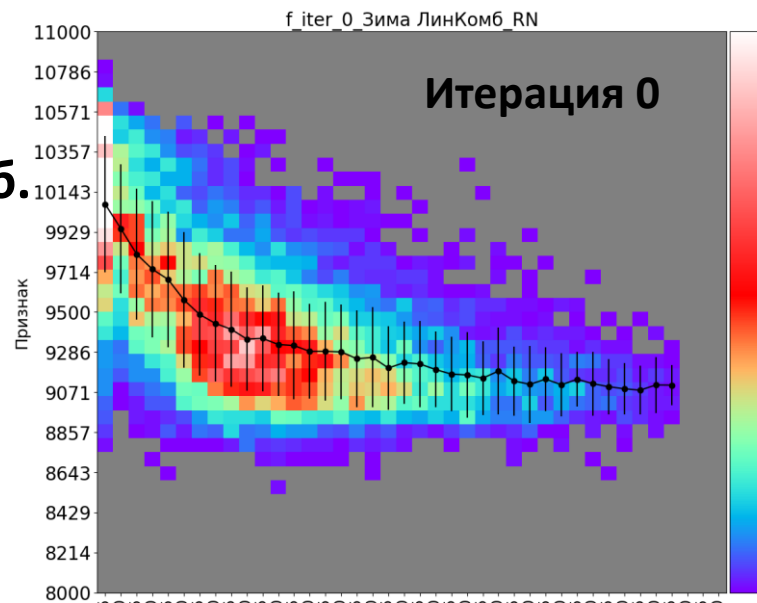


NIR

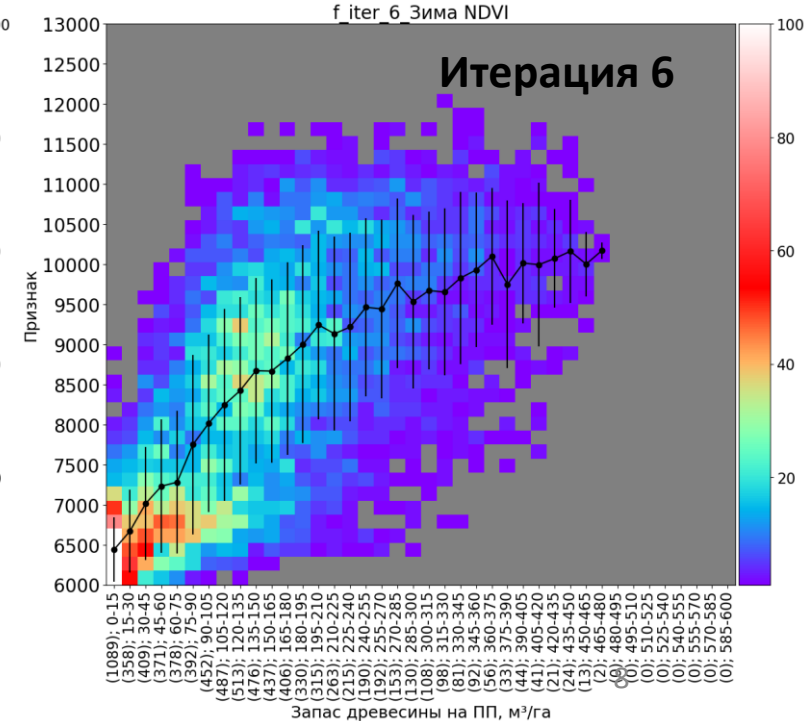
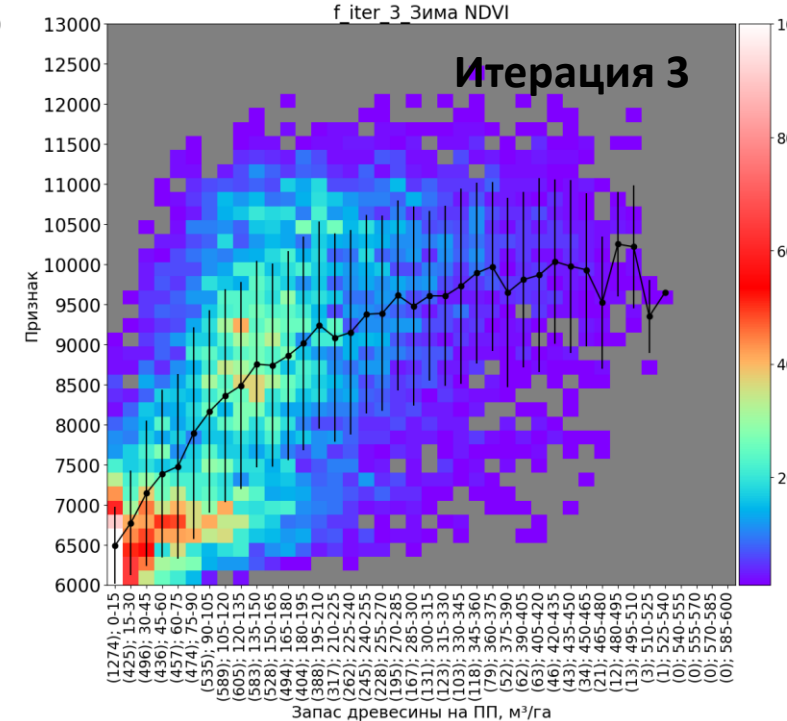
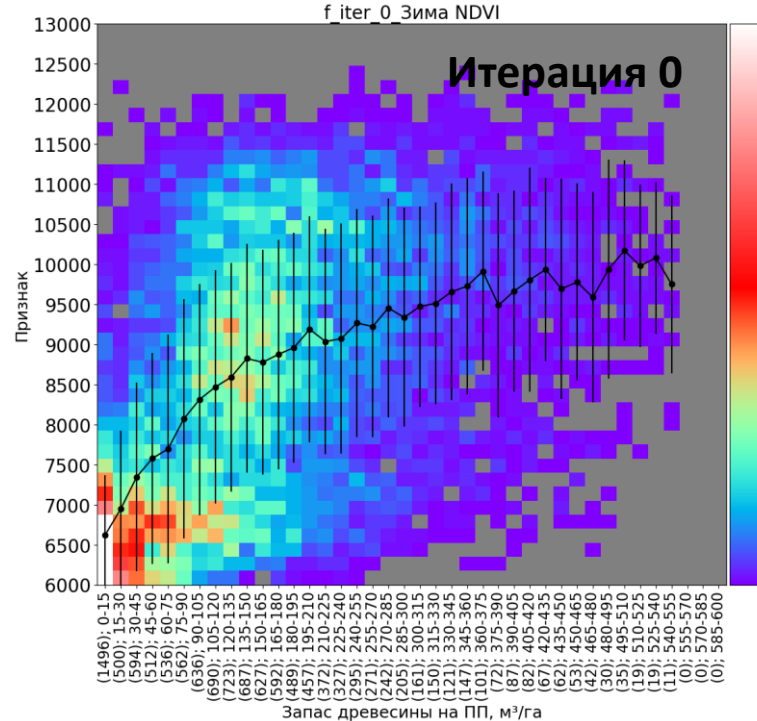


Итеративная фильтрация на основе RED и NIR; ошибок out-of-bag (все ПП)

Лин.
Комб.

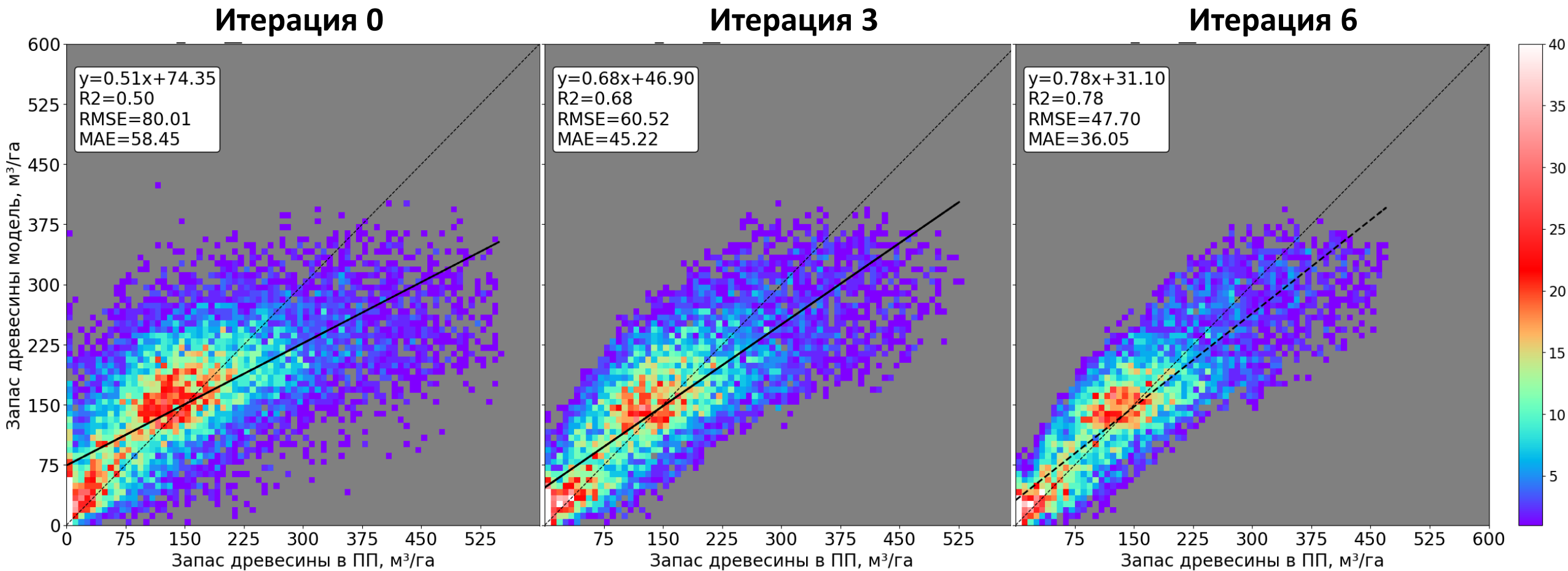


NDVI



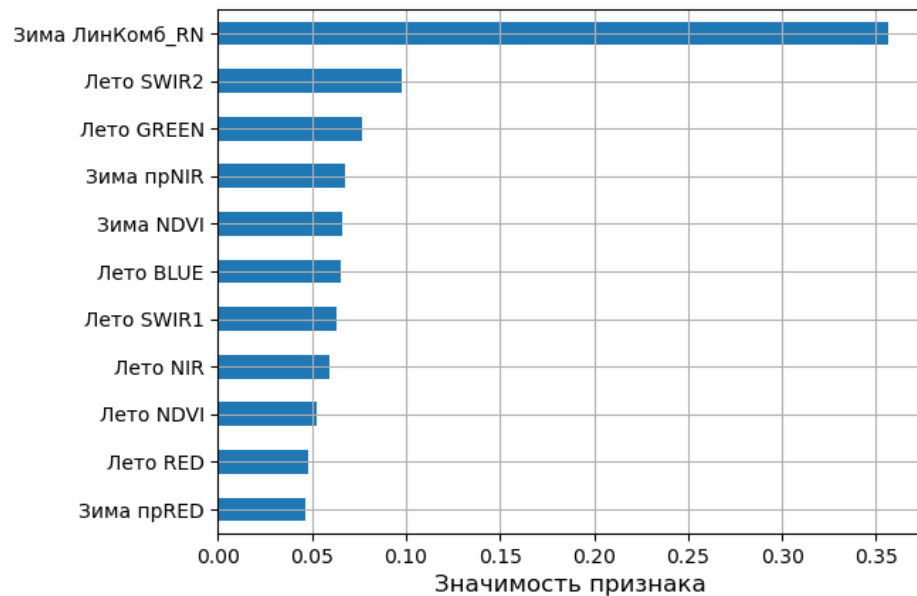
Оценка точности модели по Out-Of-Bag

Итеративная фильтрация на основе RED и NIR; ошибок out-of-bag (все ПП)

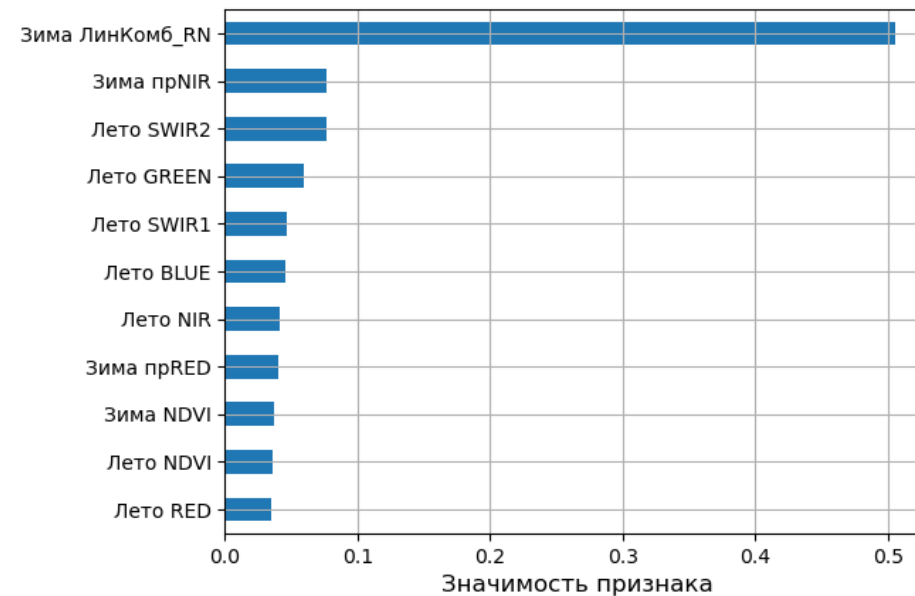


Значимость признаков при обучении модели

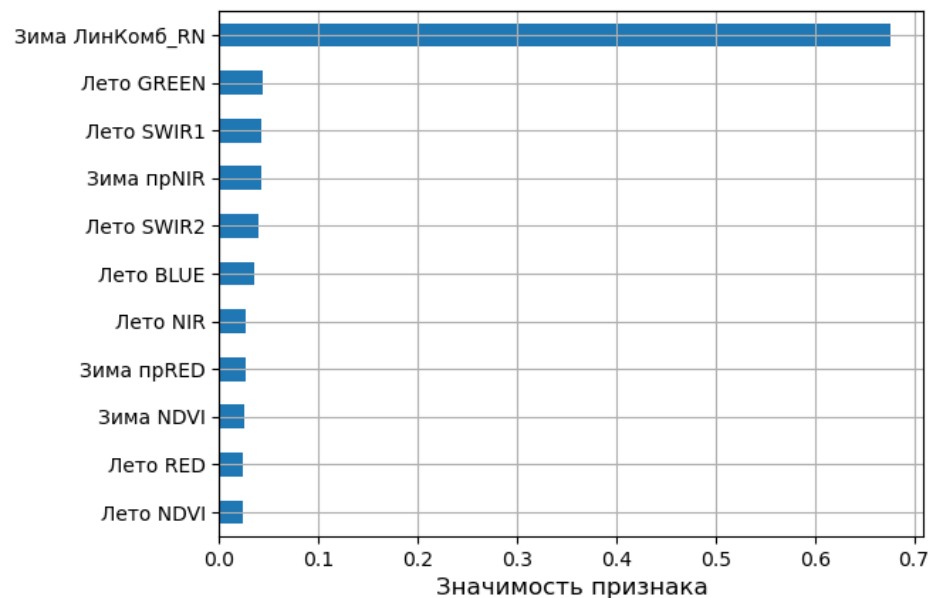
Итерация 0



Итерация 3

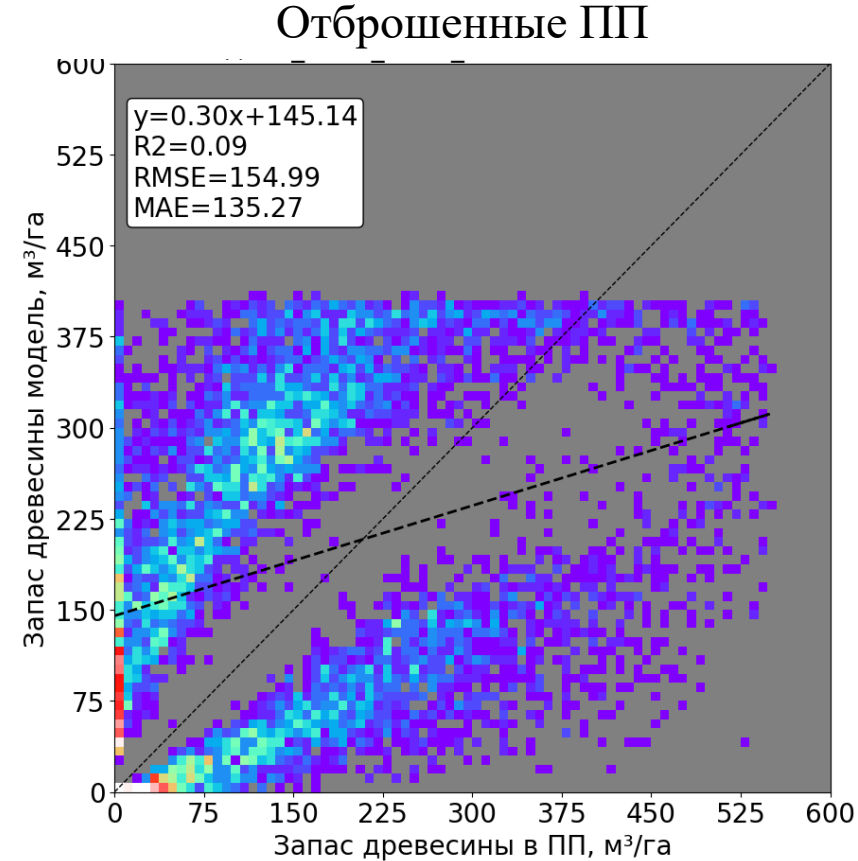
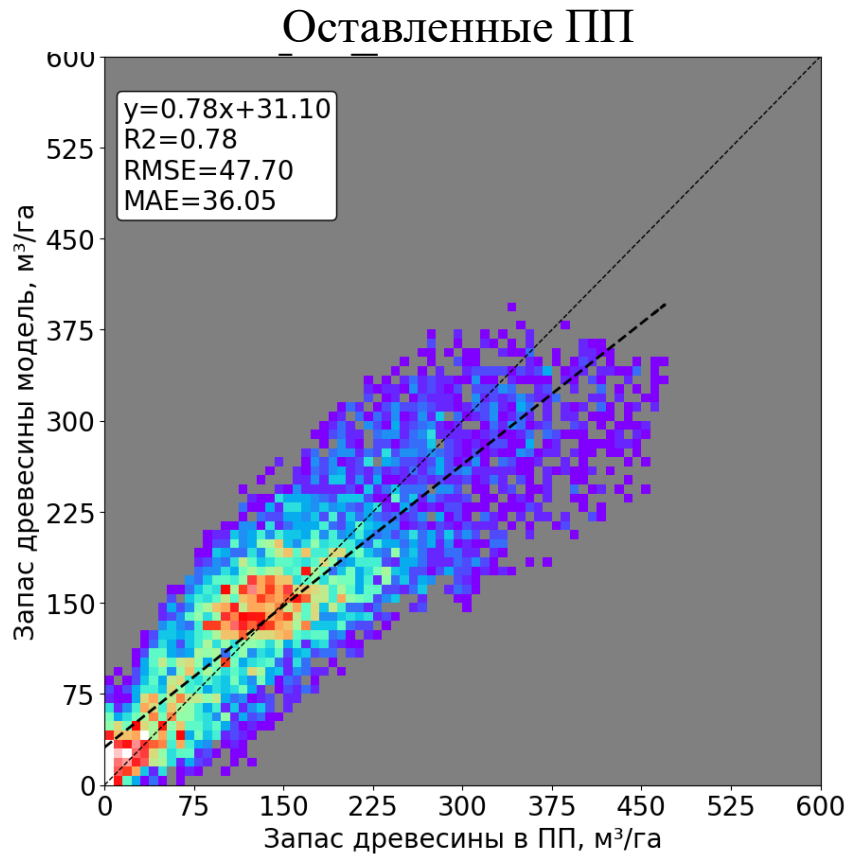


Итерация 6



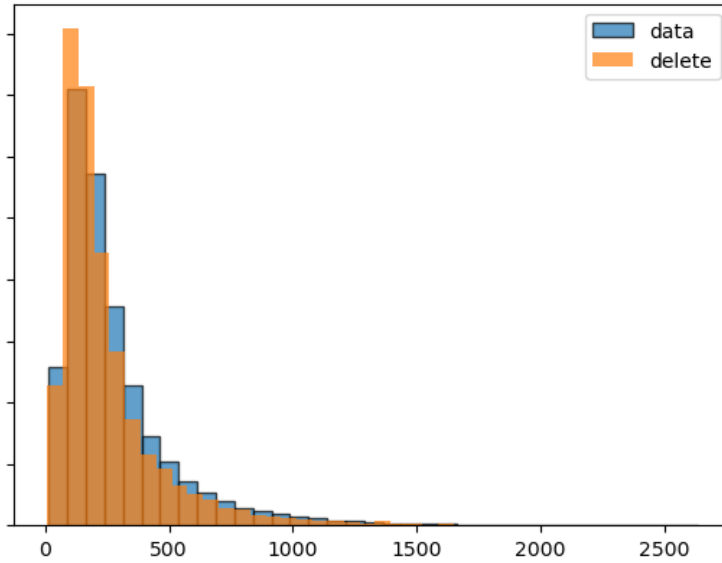
Отброшенные из модели ПП

- При использовании предложенного итеративного подхода фильтрации отбрасывается ~30% общего количества ПП.
- Некоторые измерения потенциально возможно вернуть в модель. Например:
 - Предсказанные финальной моделью (до ~10% от общего количества ПП)
 - Нелесные категории: водные объекты, дороги, пастбища и т.д. (до ~2.5%)
 - Высокий запас стволовой древесины в ПП (не моделируется из-за насыщения. ~2% ПП с значением >380 м³/га)

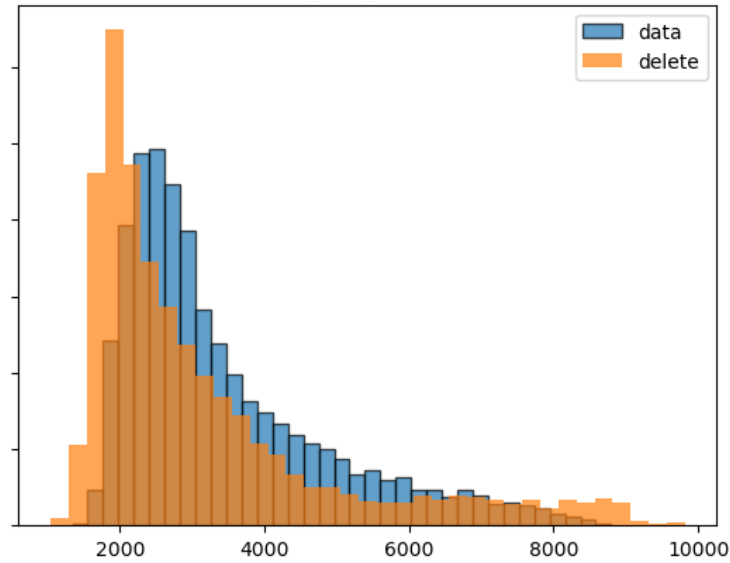


Статистика в окне 3x3 в окрестностях ПП по NIR зимы для оставленных (синий) и отброшенных ПП (оранжевый)

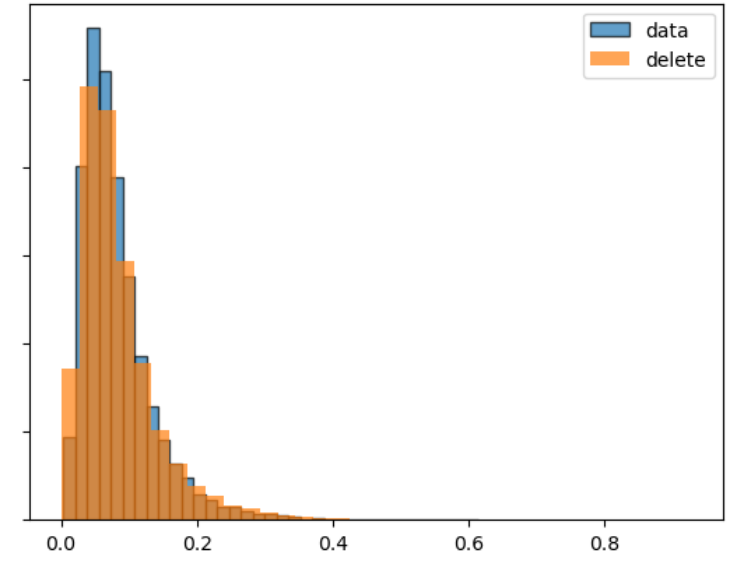
Стандартное отклонение



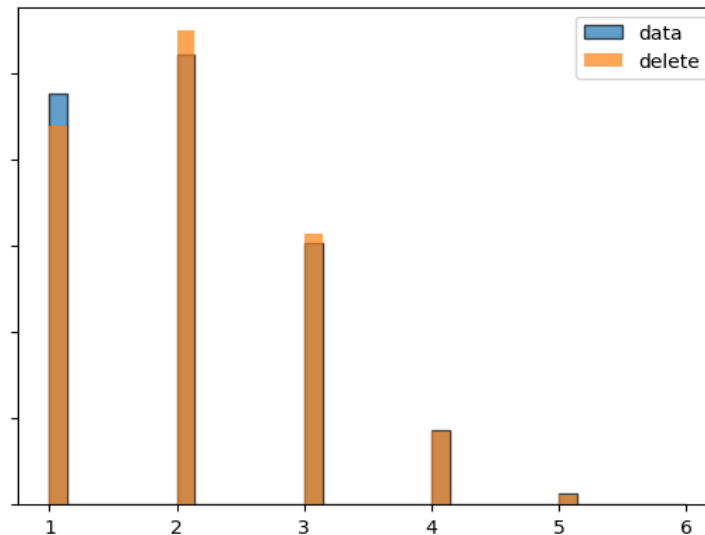
Среднее



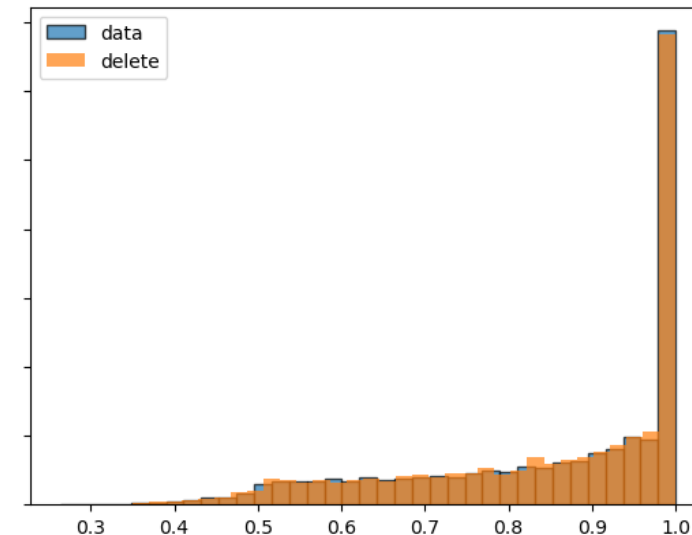
Стандартное отклонение / Среднее



Количество ярусов



Доля запаса в преобладающем ярусе



Выводы

- При использовании предложенного итеративного подхода фильтрации отбрасывается ~30% наземных обследований.
- Некоторые измерения могут быть ошибочно отброшены. Возможные причины:
 - сложная структура леса (многоярусные насаждения);
 - специфика проведения наземных обследований;
 - ошибка привязки данных;
 - насыщение в больших значениях запаса стволовой древесины;
- Часть отброшенных ПП возможно вернуть в обучающую выборку и использовать для дальнейших работ.

Дальнейшие исследования

- На данный момент не получается найти метрику, по которой можно было бы явно разделить оставленные и отброшенные предложенным итеративным фильтром данные наземных обследований при использовании глобальной модели;
- Возможные варианты:
 - расширение обучающей выборки данными наземных обследований за другие годы;
 - совместное использование локальных моделей по отдельно взятым ТП (поиск некорректных измерений);
 - использование других методов (*LightGBM).