

Влияние методов пространственной коррекции и фильтрации пробных площадей на точность оценки запаса стволовой древесины по данным Sentinel-2

¹Богодухов М.А.

¹Барталев С.А.

(1) ИКИ РАН

Двадцать третья международная конференция
«СОВРЕМЕННЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА»

10-14 ноября, 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:

- Оценить влияние методов пространственной коррекции и фильтрации ПП на точность оценки запаса запаса стволовой древесины при совместном использовании с данными Sentinel-2

Используемые данные:

Обучающая выборка:

- Материалы наземных обследований запаса ствольной древесины лесов на уровне регулярной сети ПП (реласкопические, шаг 200 метров, в пределах 2×2 км), заложенных на территории 240 Тестовых Полигонов (ТП), расположенных по всей территории России в 2023 и 2024гг.

Признаки для обучения модели (Random Forest):

- Нормализованный по углу солнца и территории зимний композит Sentinel-2:
 - RED, NIR
 - Линейная комбинация RED, NIR для учета различия между хвойными и листопадными;
- Летние композитные изображения Sentinel-2 с временным шагом 5 дней:
 - BLUE, GREEN, RED, NIR, SWIR1, SWIR2

Территориальное расположение ТП

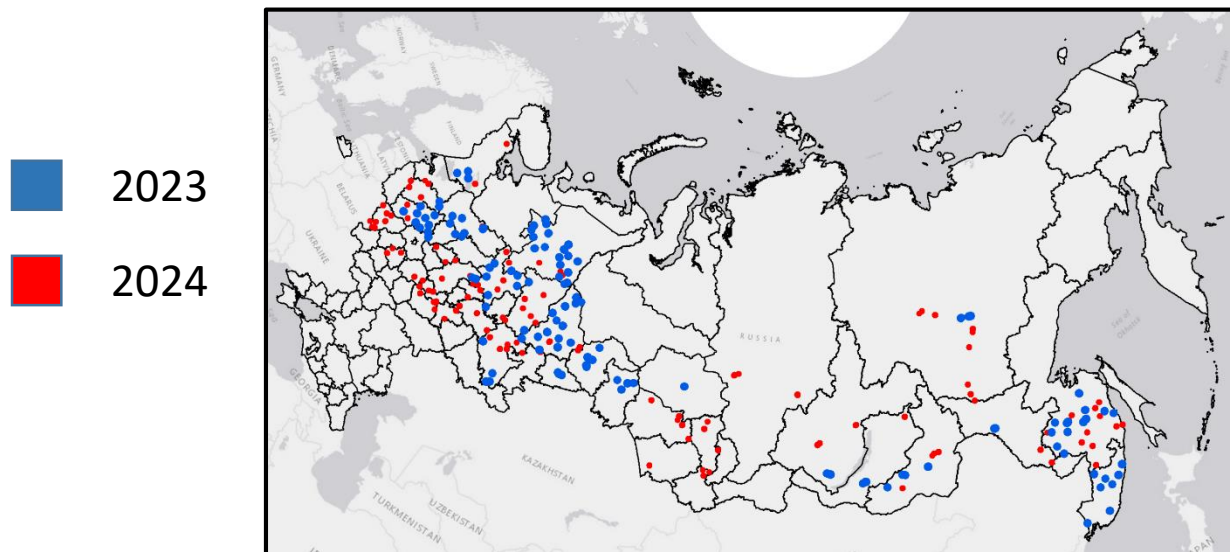
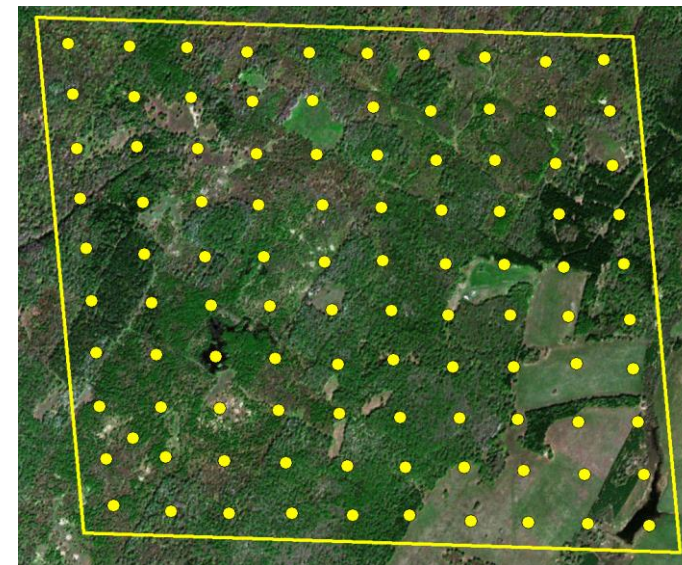
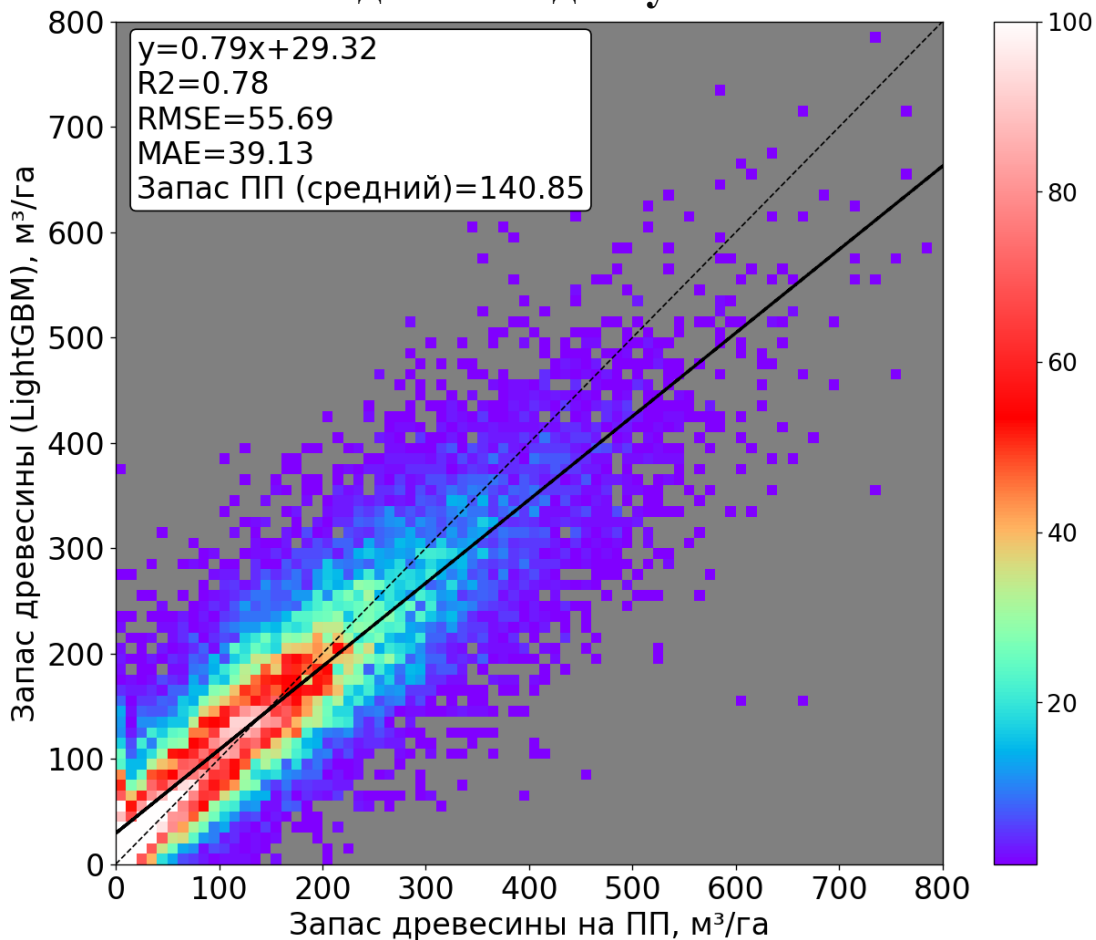


Иллюстрация расположения ПП в ТП



Предыдущие результаты (LightGBM)

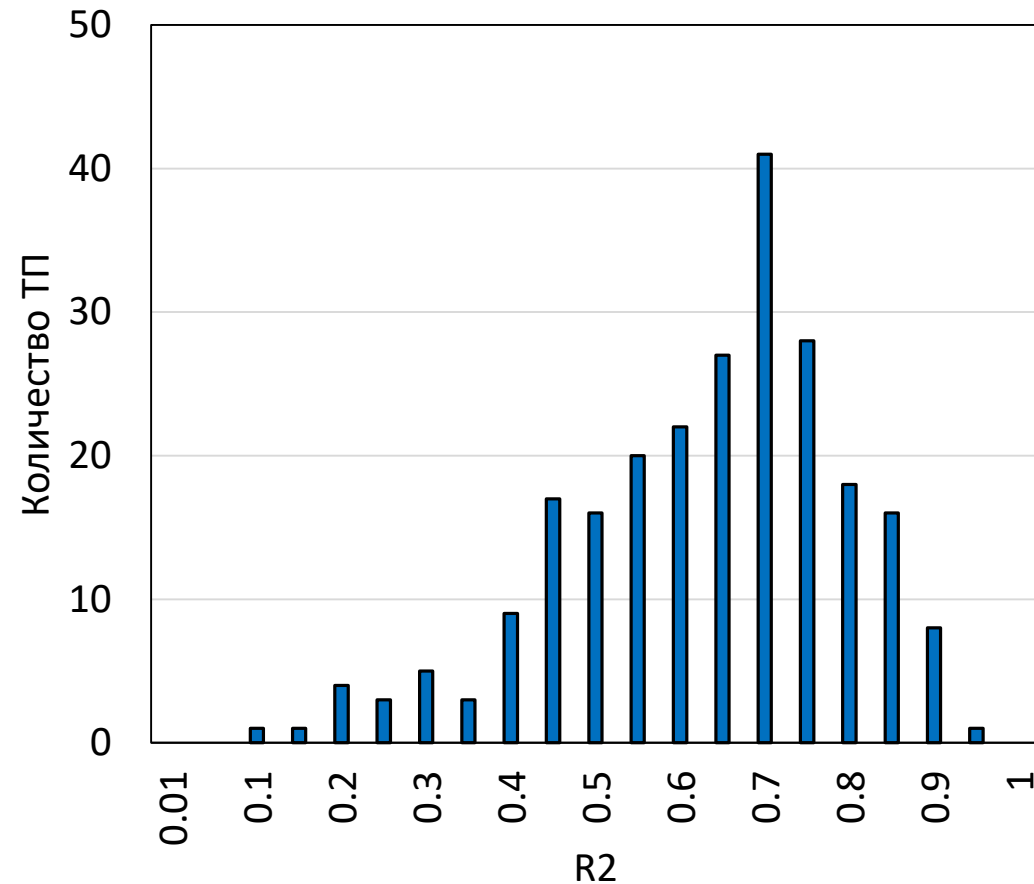
Интегральная оценка запаса стволовой древесины для всех доступных ТП



Интегральная оценка:

- Высокий уровень согласованности;
- Большой разброс, выбросы

Точность локальных оценок запаса стволовой древесины



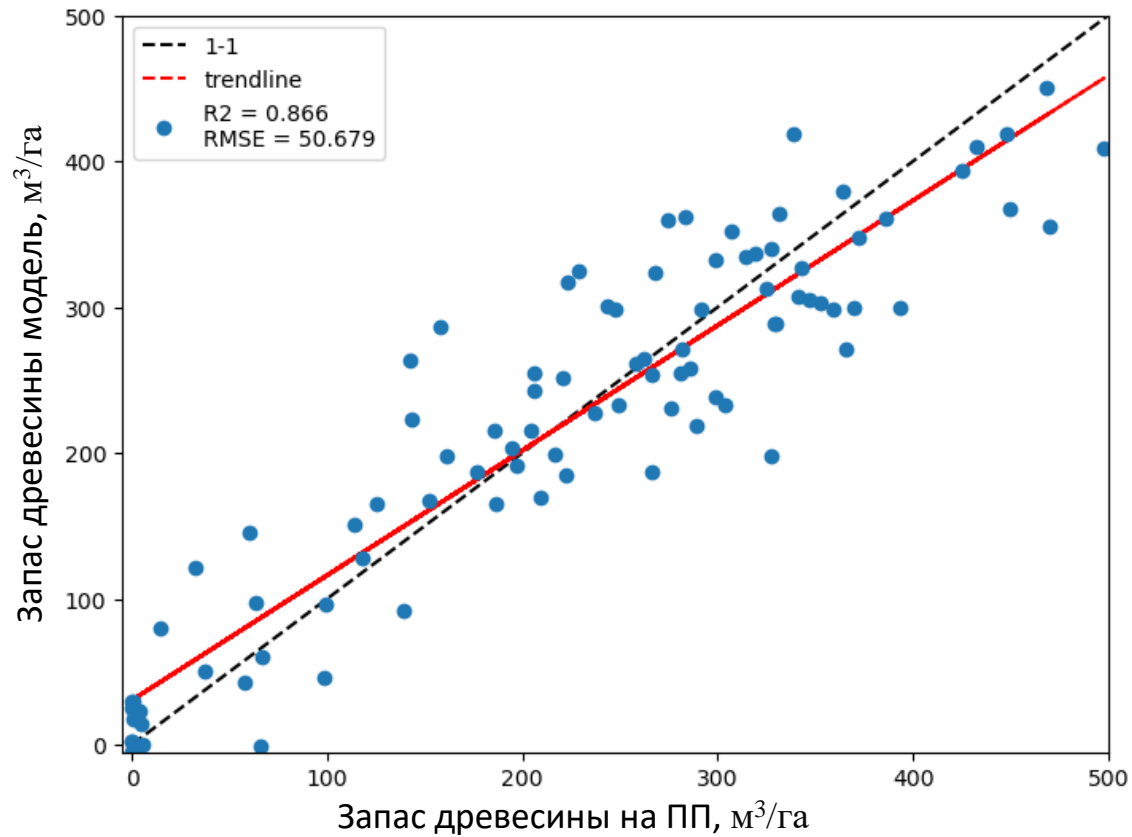
Локальная оценка:

- Основная масса моделей имеет $R^2 \approx 0.7$, демонстрируя умеренное качество предсказания
- Некоторые оценки демонстрируют низкую точность $R^2 < 0.5$,

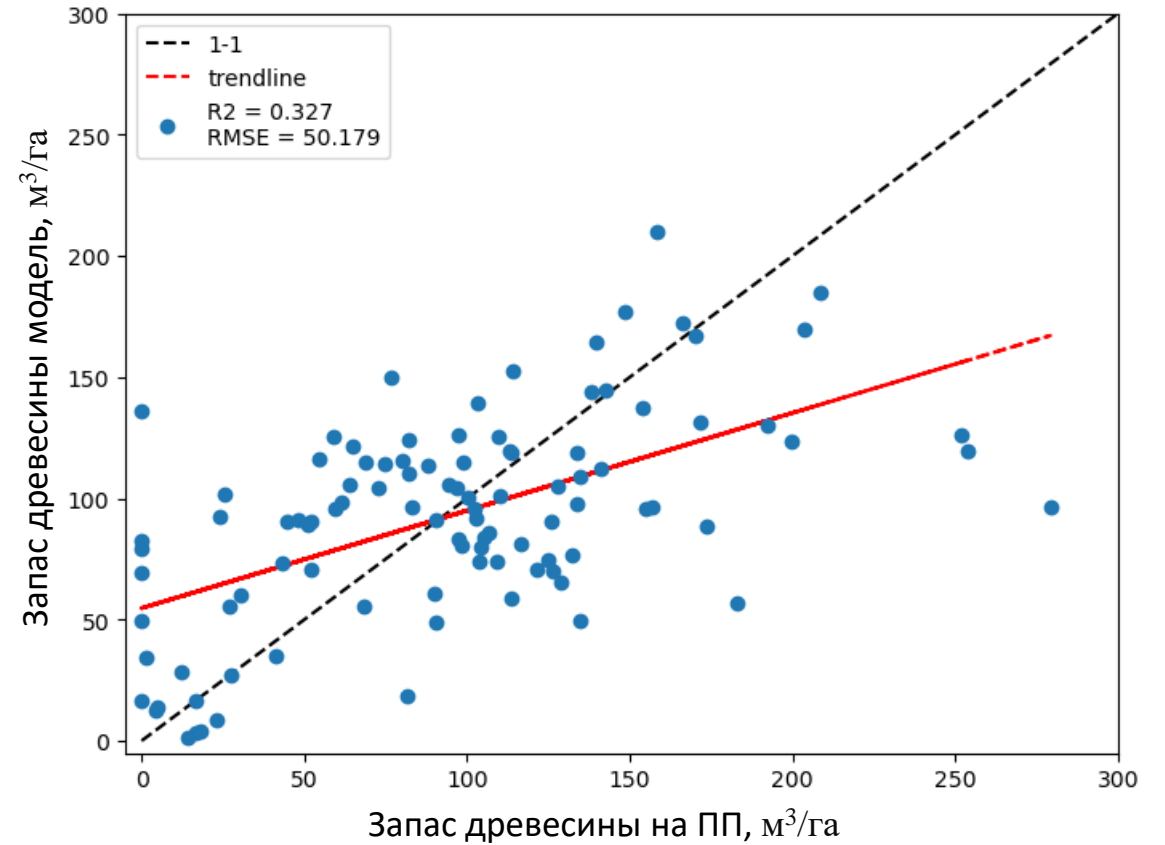
Предыдущие результаты

Локальная оценка запаса ствольной древесины (LightGBM)

Высокий уровень согласованности



Низкий уровень согласованности



Возможные трудности при совместном использовании реласкопических измерений и спутниковых данных в ТП

1. Несбалансированность выборки в области низких значений запаса стволовой древесины (~до 10м³/га, включая безлесные участки)

- Из-за особенностей методики сбора данных:
 - Могут находиться деревья с небольшим диаметром стволов (молодой, восстанавливающийся лес, кустарники), которые не подсчитываются и определяются в подрост
 - Возможное ограничение видимости эксперта неподходящими для измерения объектами

2. Граничный эффект + потенциальная неточность привязки в труднодоступных зонах

- Использование регулярной координатной сети может приводить к расположению ПП в неоднородных и граничных участках (лес – не лес – инфраструктура)
- Потенциальная неточность привязки может сместить измерения в соседнюю зону

3. Другие особенности

- Измерения зависят от эксперта;
- Некорректный ввод информации в базу данных.

1. Особенности в области низких значений запаса стволовой древесины

- Из-за особенностей методики сбора данных:
 - Могут находиться деревья с небольшим диаметром стволов (молодой, восстанавливающийся лес, кустарники)
 - Возможное ограничение видимости эксперта неподходящими для измерения объектами

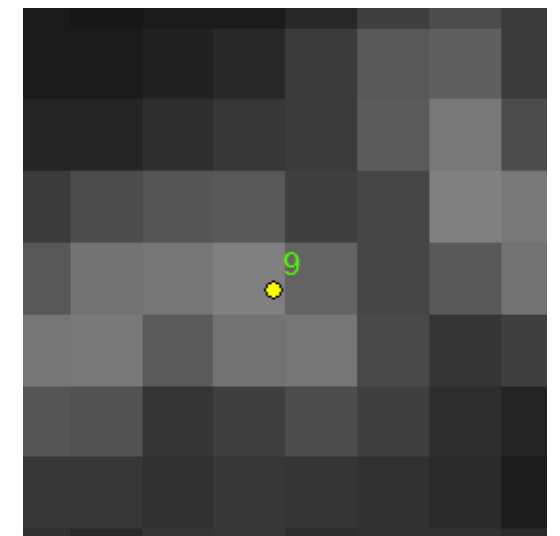
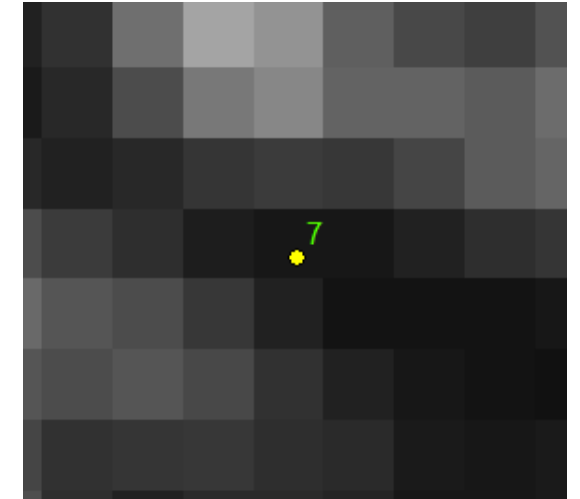


2. Граничный эффект

Иллюстрация особенностей расположения ПП

ПП №7 и ПП №9:

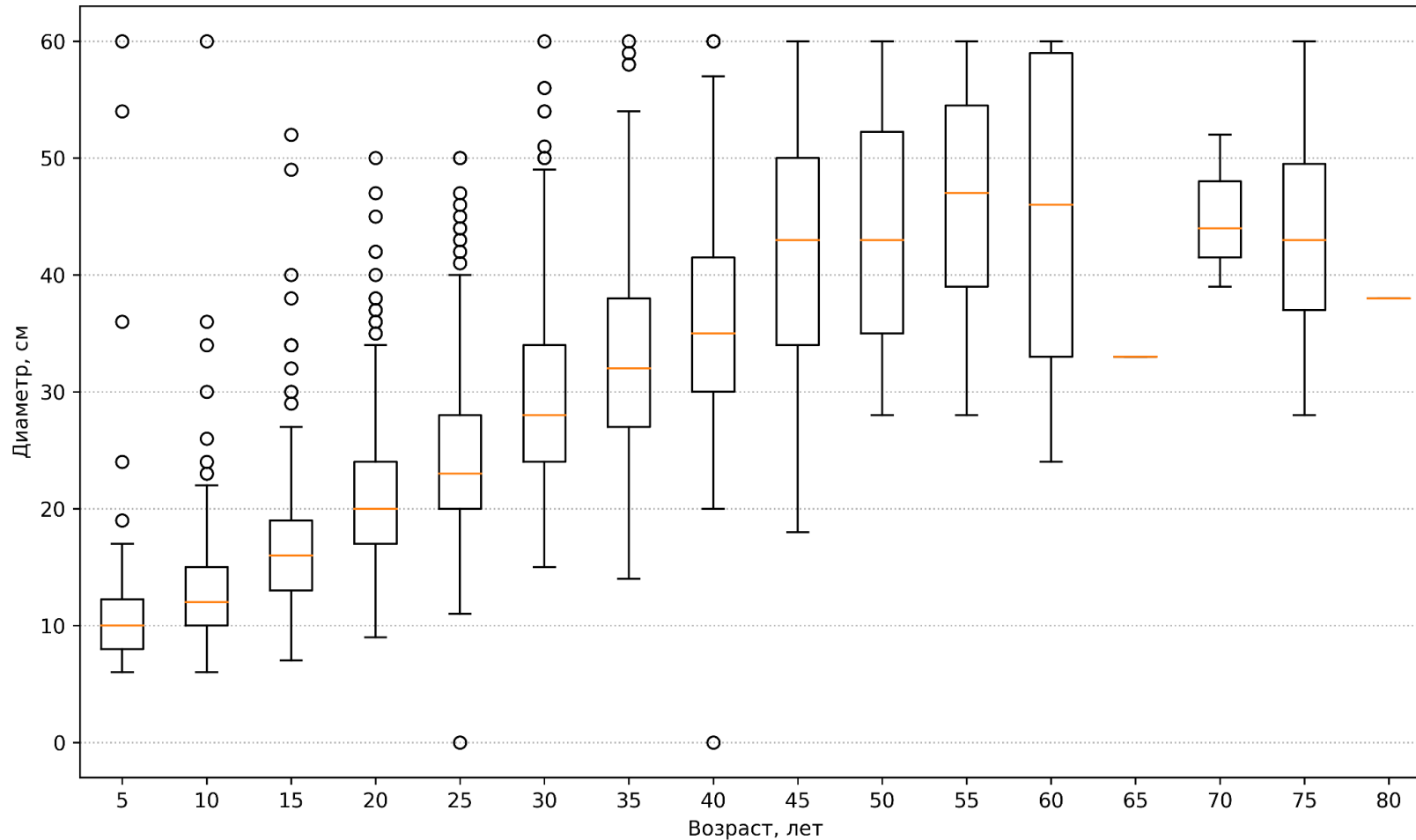
- Имеют достаточно похожие параметры в базе данных: запас ($\text{м}^3/\text{га}$), количество деревьев, модельных деревьев, реласкопический коэффициент и т.д.
- По зимнему композиту ПП имеют кардинально разные яркости, но при обучении ставятся в соответствии практически одному запасу
- Возможно необходимо корректировать расположение или исключать неоднозначные группы элементов выборки



3. Другие особенности

- В базе данных периодически встречаются аномальные, несогласованные записи (отрицательные значения, сверхвысокие/сверхнизкие измерения для конкретных элементов леса и т.д.)

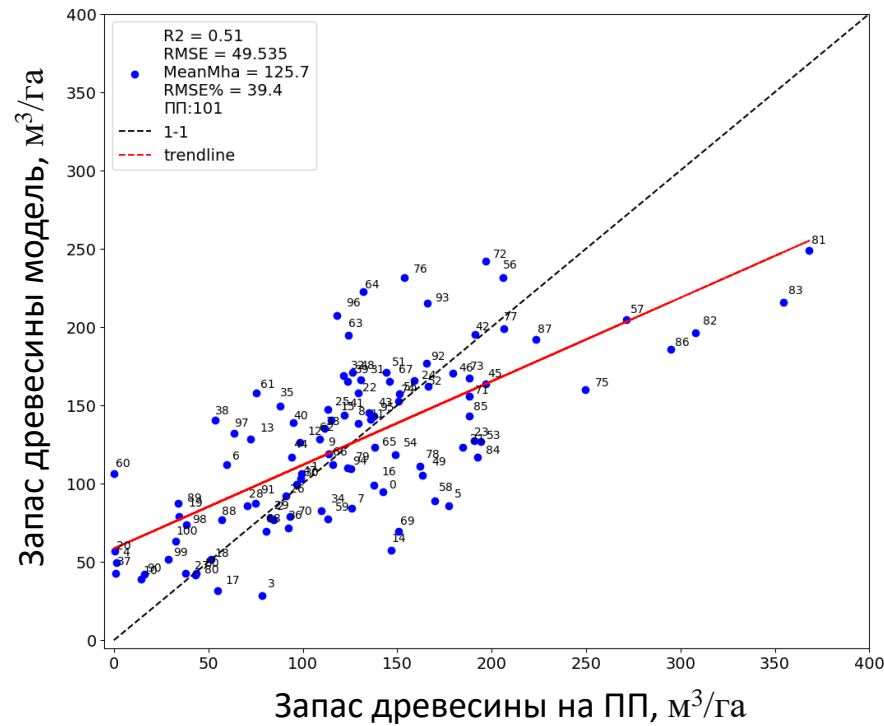
Иллюстрация связи возраста и диаметра модельных деревьев при вводе данных для березы II бонитета



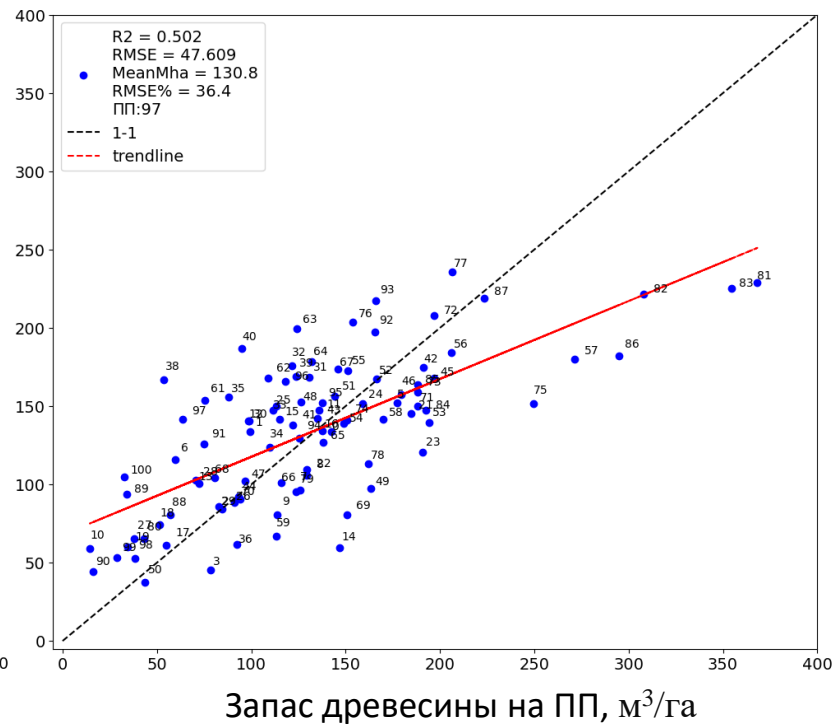
Локальная оценка запаса ствольной древесины.

Добавление безлесных эталонов

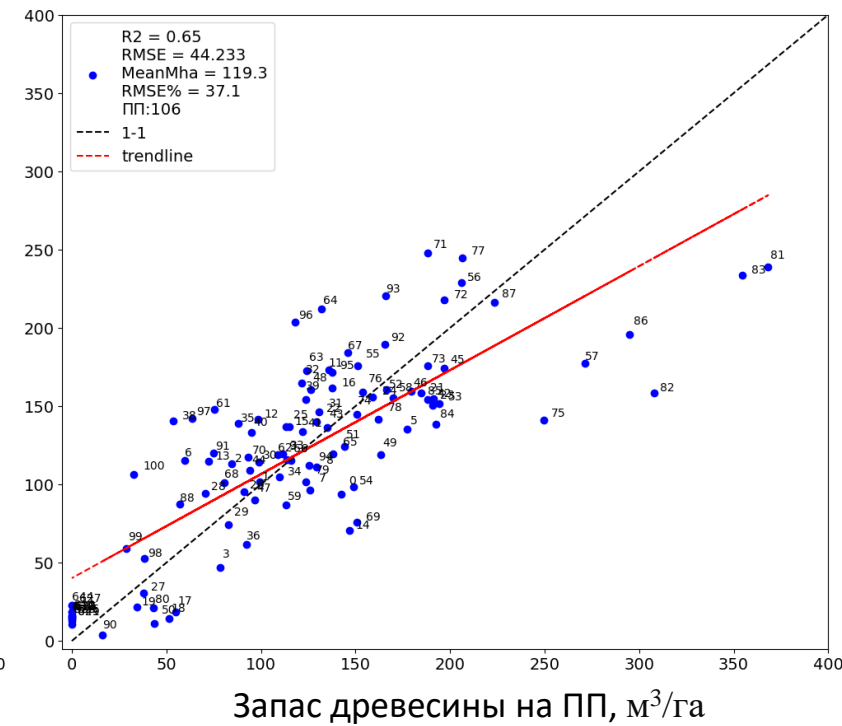
(1) Исходные данные (101 ПП)



(2) Исключены безлесные ПП



(3) Добавлены безлесные эталоны



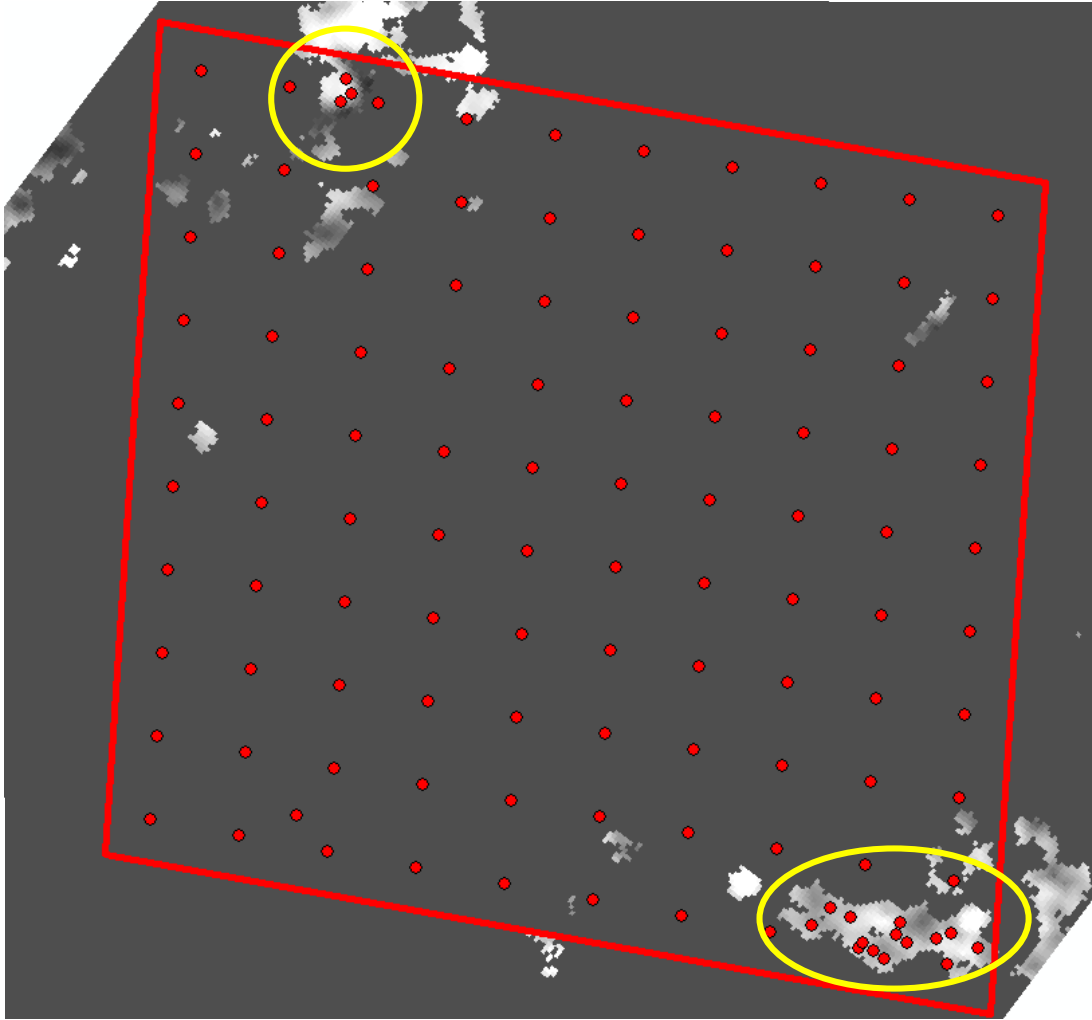
(1) Безлесные ПП прогнозируются в диапазоне 40 ~ 100 м³/га (по Out-Of-Bag). Из-за крайне небольшой обучающей выборки эти элементы значительно влияют на обучение модели.

(2) При их исключении из выборки (4 элемента) формально ничего не меняется: облако точек по прежнему сконцентрировано в пике гистограммы запаса ствольной древесины с высоким разбросом. недооценка высоких запасов

(3) При добавлении безлесных эталонов (9 элементов) разброс уменьшился, элементы более сгруппированы, произошло смещение прогнозов в область высоких значений, появились прогнозы < 30 м³/га

Формирование безлесных эталонов

Иллюстрация добавления безлесных эталонов в ТП



- Для поиска безлесных эталонов используется карта леса ESA WorldCover 10м
- В окрестностях ТП заполняем нелесные области точечными измерениями таким образом, чтобы до лесных участков и элементов между собой было достаточное расстояние
- Извлечем значения яркостей зимнего композита и найдем медианное значение для каждого точечного измерения
- Для исключения предельных случаев и шума найдём наиболее характерные нелесные участки в окрестностях ТП. Выберем в интервале вокруг медианы необходимое количество ПП для добавления в набор в качестве безлесных эталонов

- Красные точки, граница – ПП в пределах ТП
- Желтая область – сформированные безлесные эталоны

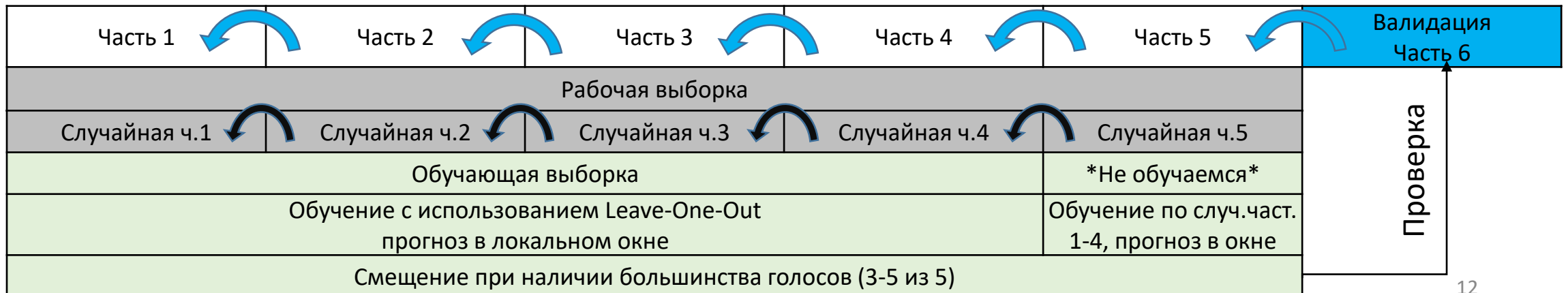
Пространственная коррекция расположения ПП

Алгоритм:

- 1) Выборка разделяется на 6 равномерных частей по распределению запаса стволовой древесины
- 2) Одна часть отводится под валидацию, остальные части – рабочая выборка
- 3) Рабочая выборка случайным образом разбивается на 5 частей
 - 3.1) 4 части – обучающая выборка, 1 часть – не используется для обучения (прогнозная)
 - 3.2) Для каждой ПП в рабочей выборке:
 - 3.2.1) Прогноз в локальном окне (обучение: для обучающей выборки – Leave-One-Out; для оставшейся – по обучающей выборке)
 - 3.2.2) Расчет вариации и дисперсии среди деревьев ансамбля Random Forest
 - 3.2.3) На основе вариации фиксируется голос, в какой пиксель локального окна сместить ПП
- 4) Повторяем пп. 3.1-3.2, пока каждая часть рабочей выборки не будет прогнозной 1 раз
- 5) Проверяем на валидации обученную модель по исходной и смещенной (смещаем при большинстве голосов: 3-5 из 5) рабочей выборке
- 6) Повторяем пп. 2-3, пока каждая часть не будет валидационной 1 раз

В результате работы алгоритма:

- 25 голосов (5 смещений, 5 разбиений) о направлении смещения в локальном окне → при большинстве голосов (13-25 из 25) пространственная коррекция расположения ПП
- Для каждой ПП вариация и дисперсия ансамбля Random Forest → данные для статистической фильтрации выборки.

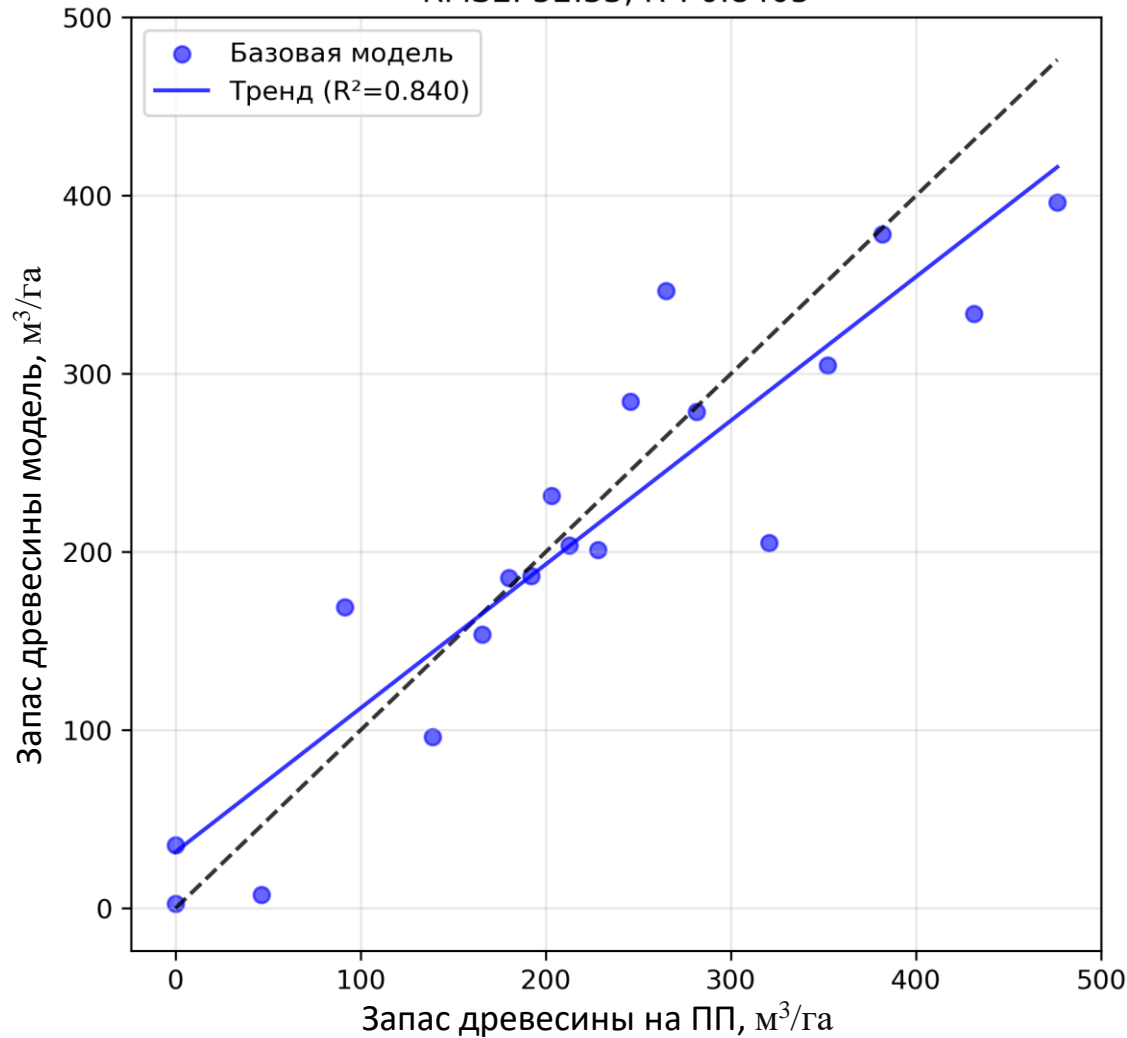


Пространственная коррекция расположения ПП

Валидация (1)

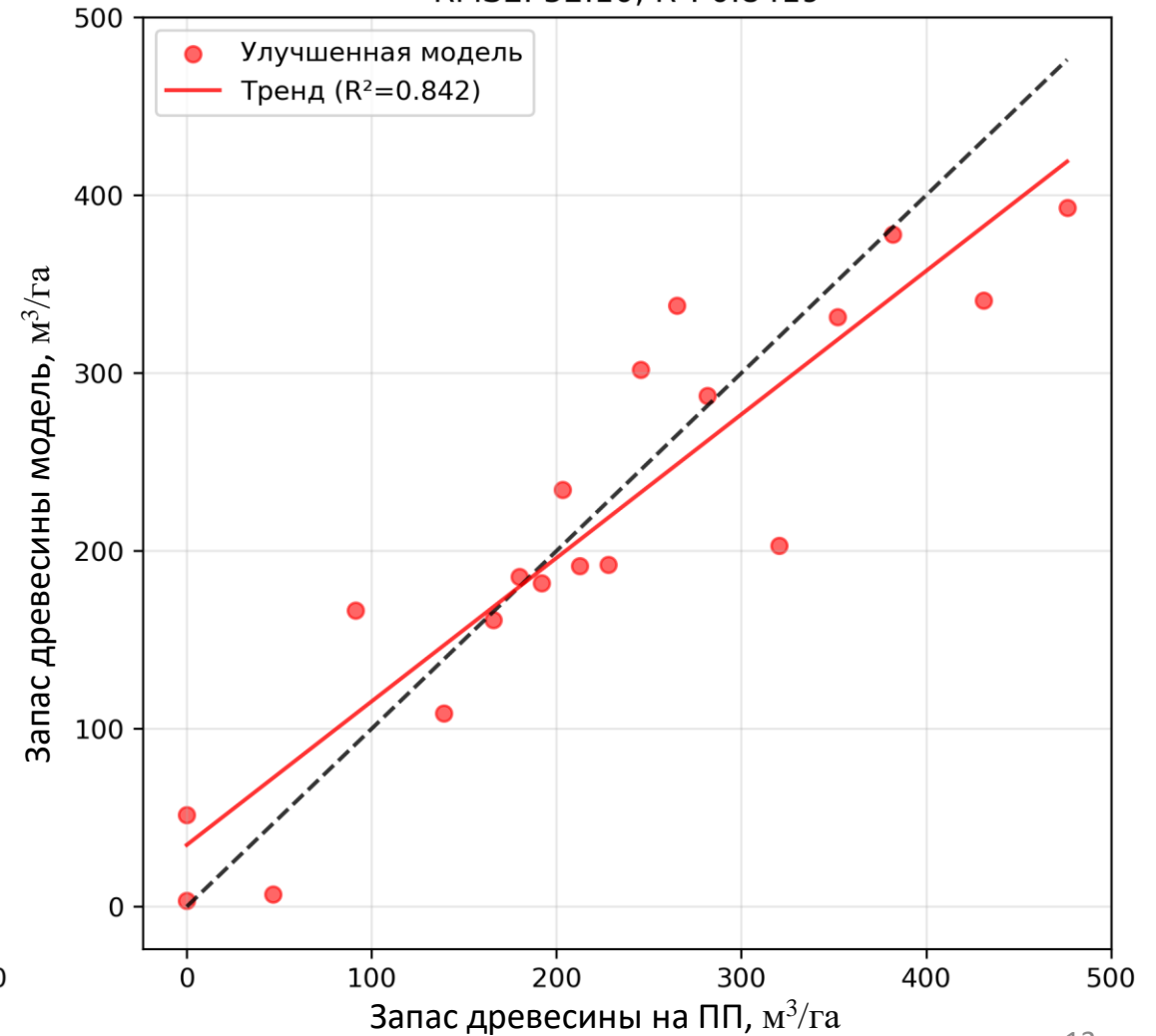
Исходная выборка

RMSE: 52.33, R^2 : 0.8405



Смещенная выборка

RMSE: 52.10, R^2 : 0.8419

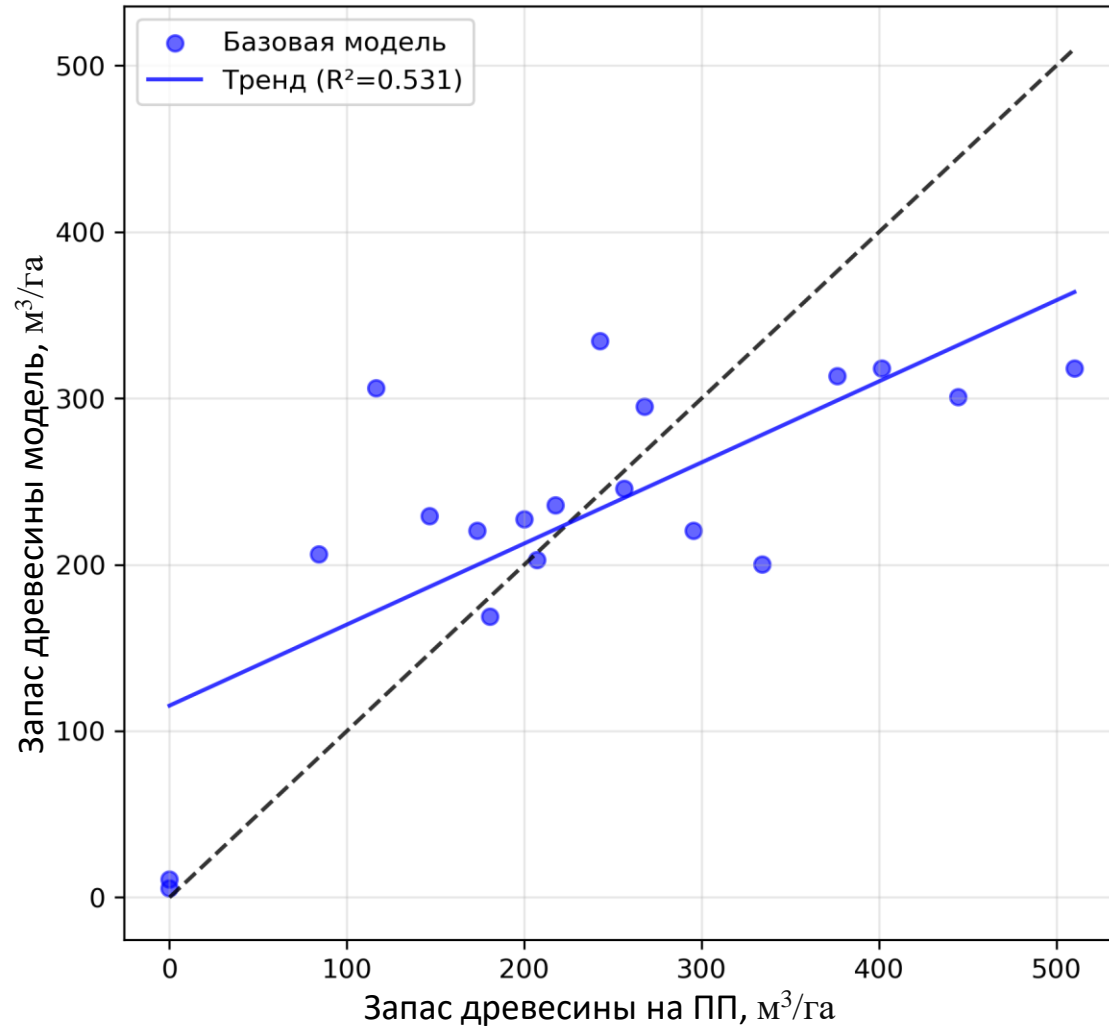


Пространственная коррекция расположения ПП

Валидация (2)

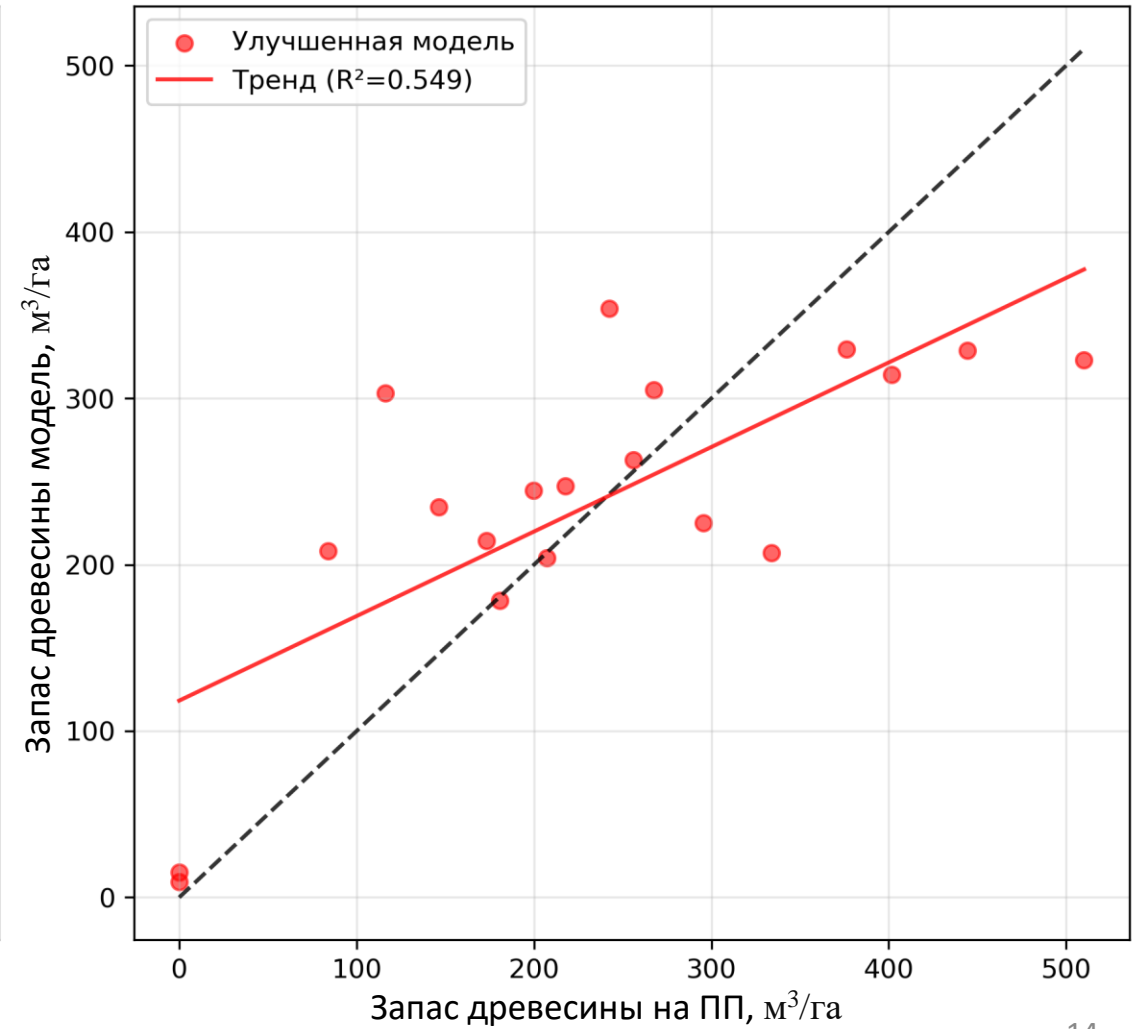
Исходная выборка

RMSE: 92.49, R^2 : 0.5309



Смещенная выборка

RMSE: 90.70, R^2 : 0.5489

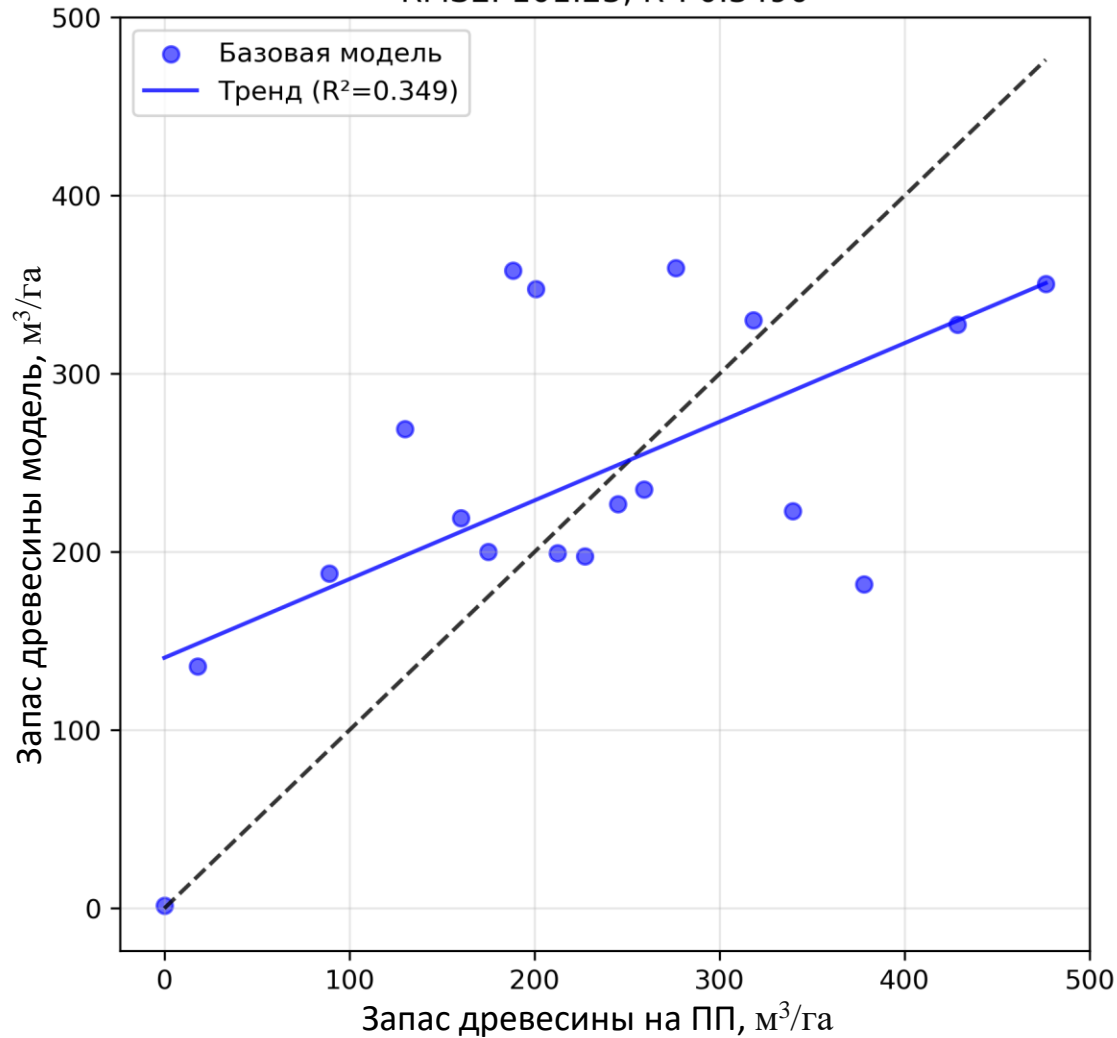


Пространственная коррекция расположения ПП

Валидация (3)

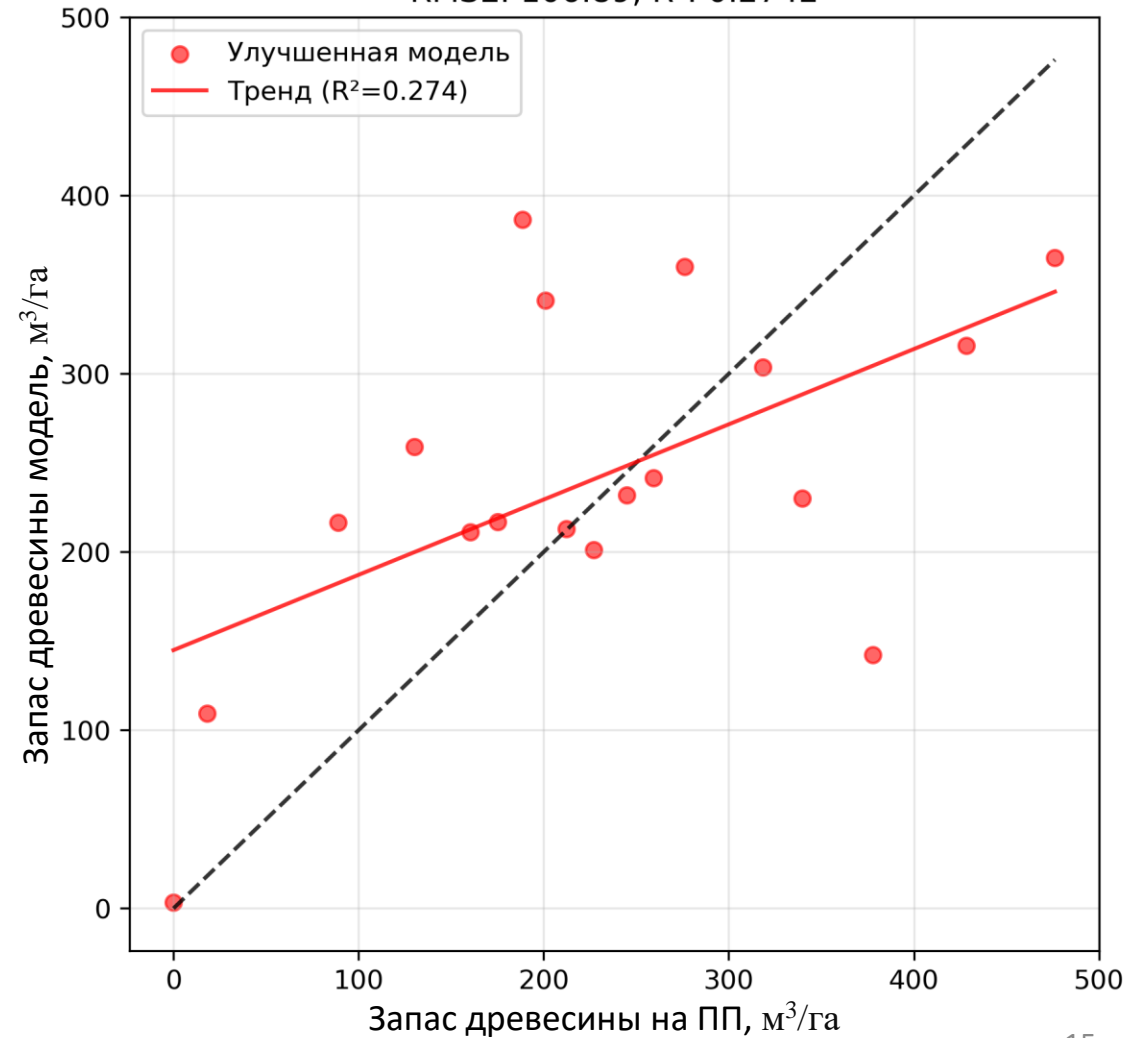
Исходная выборка

RMSE: 101.23, R^2 : 0.3490



Смещенная выборка

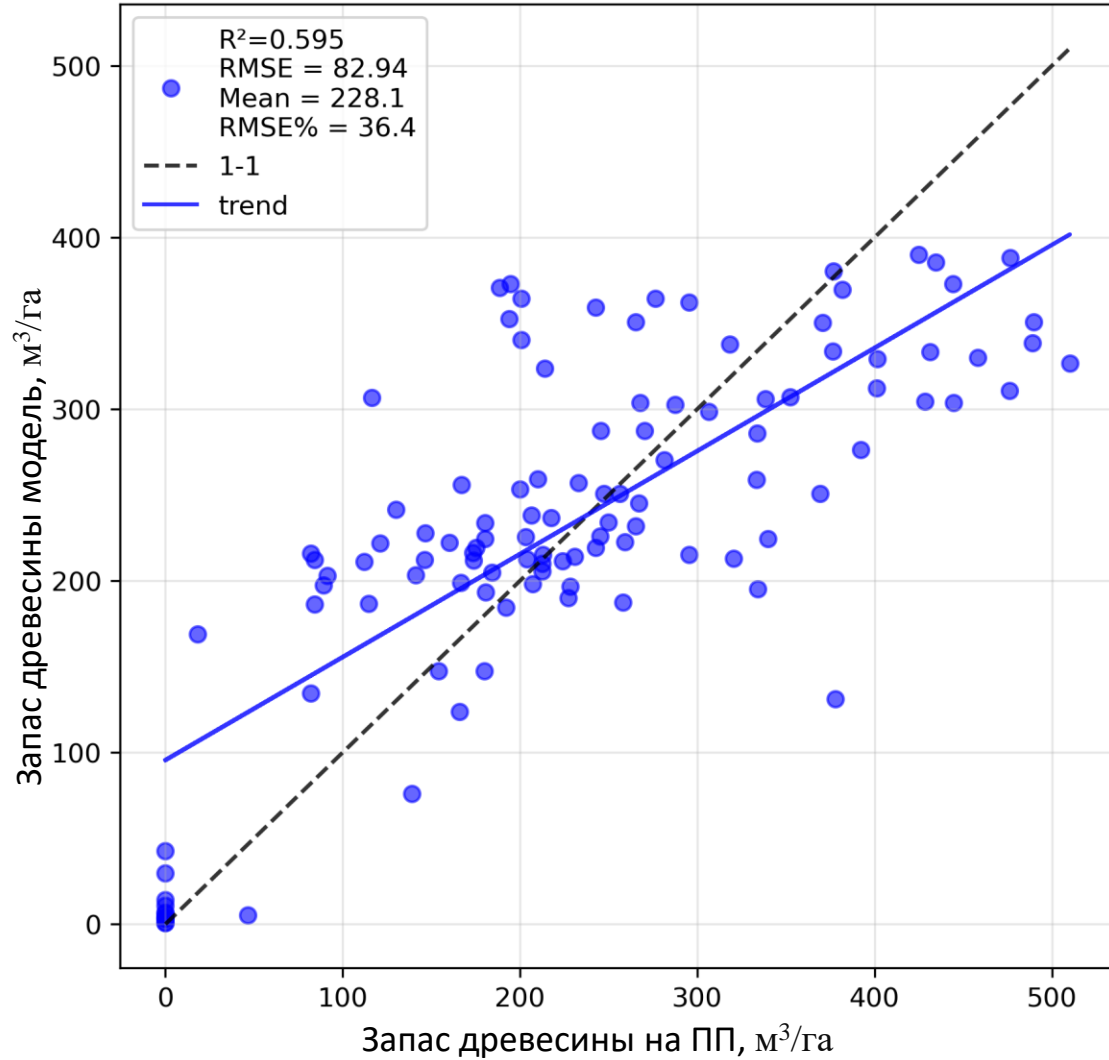
RMSE: 106.89, R^2 : 0.2742



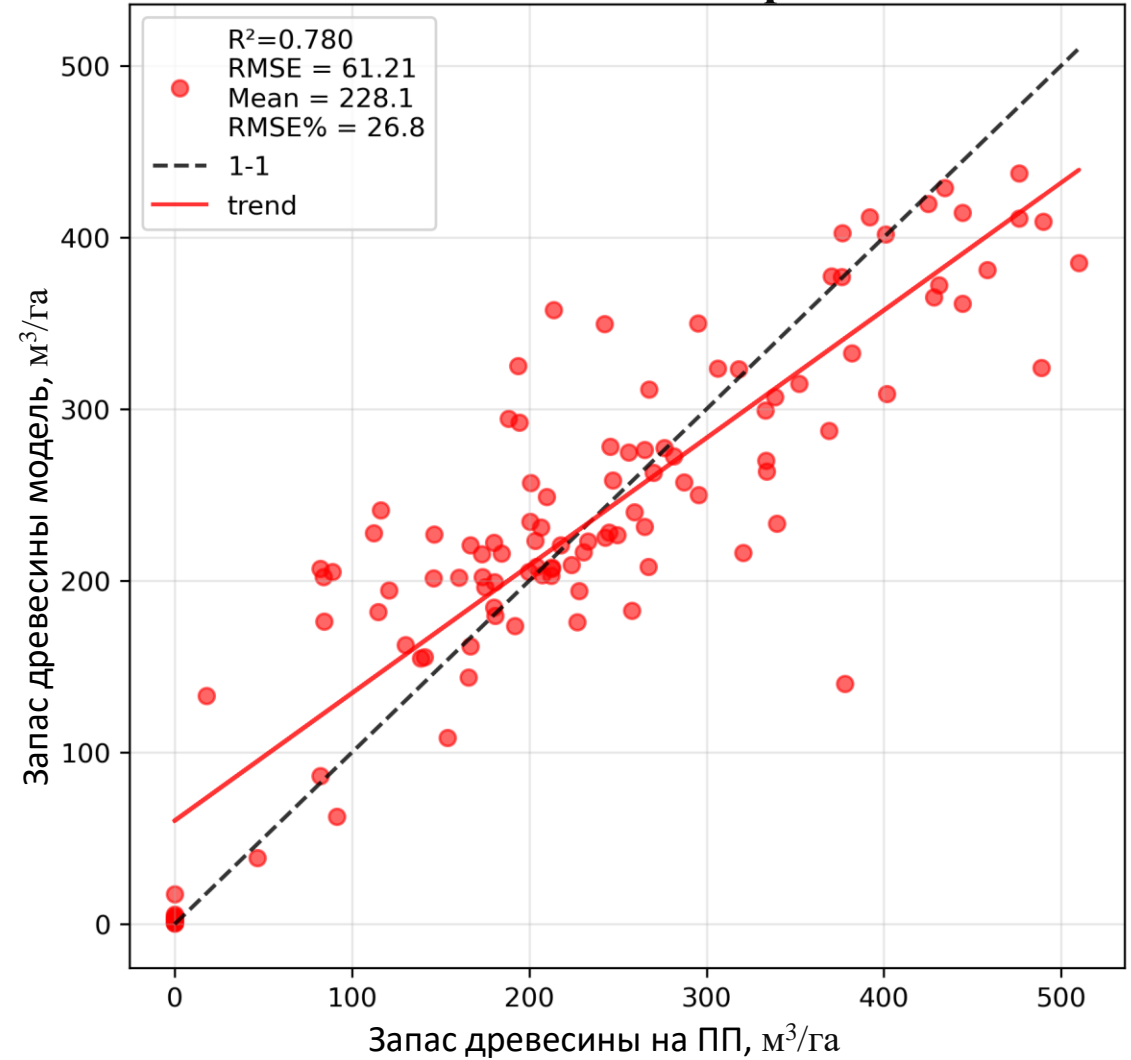
Пространственная коррекция расположения ПП

Смещение по всем ПП при наличии 13-25 из 25 голосов

Исходная выборка



Смещенная выборка

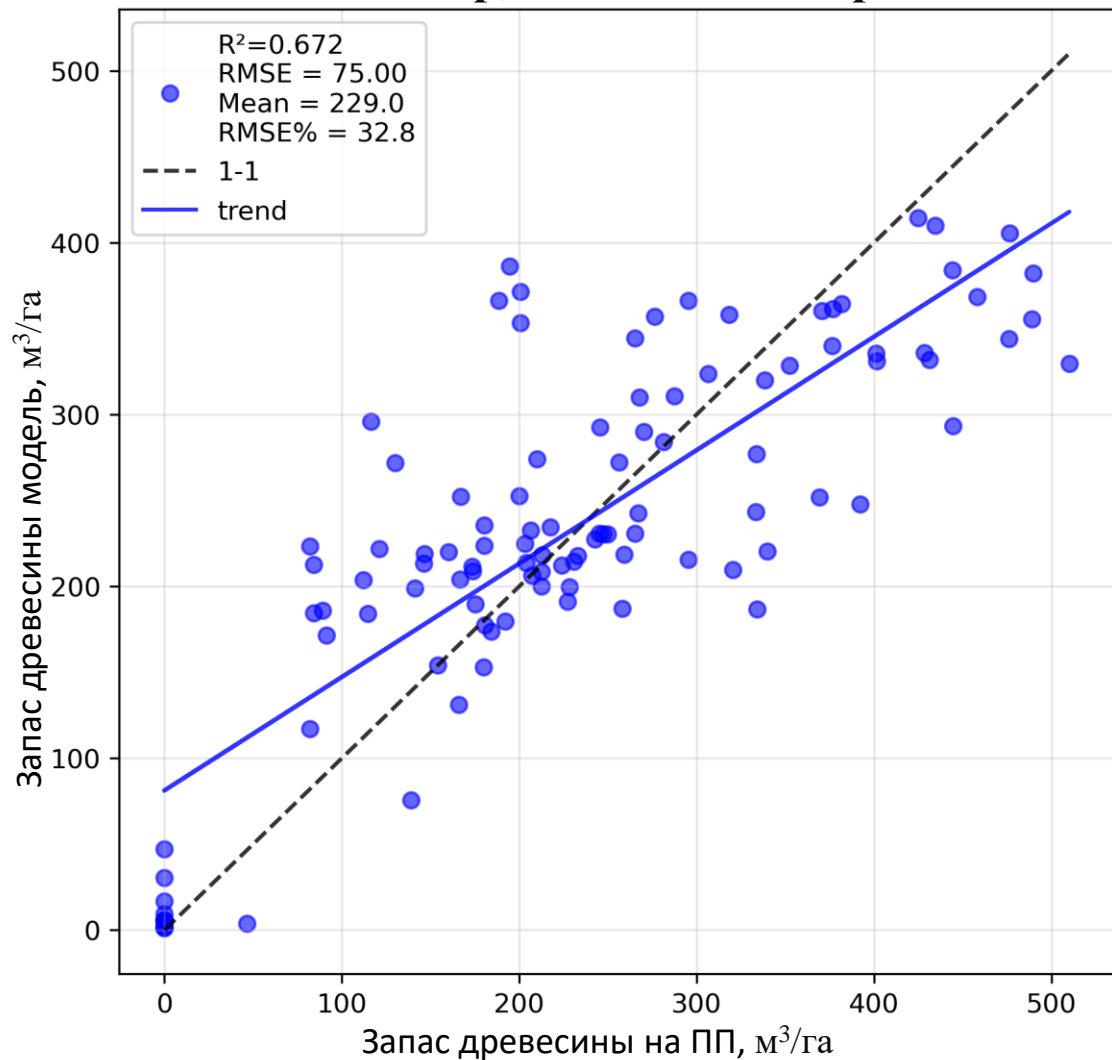


Пространственная коррекция расположения ПП

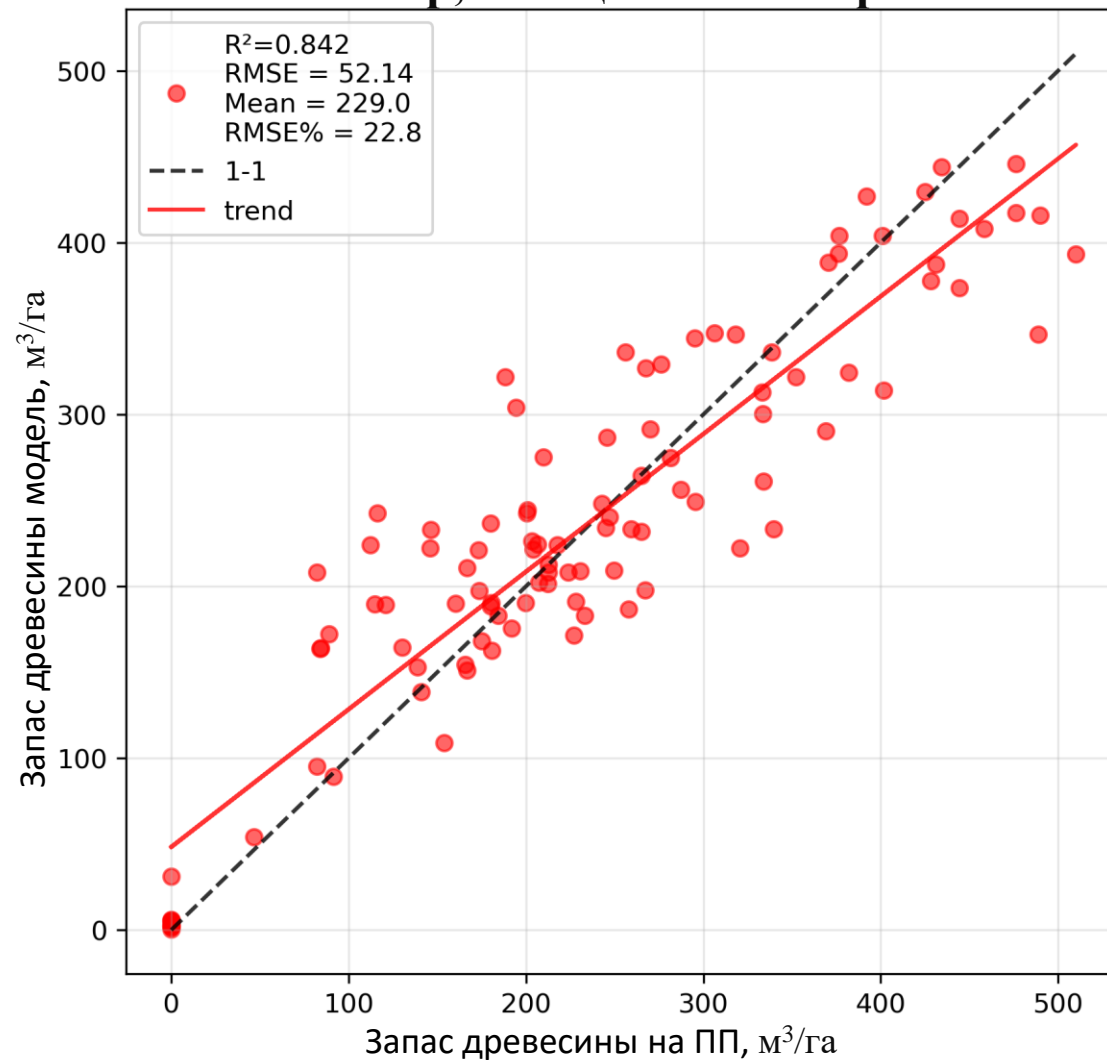
Статистическая фильтрация

Смещение по всем ПП при наличии 13-25 из 25 голосов

Фильтр, Исходная выборка



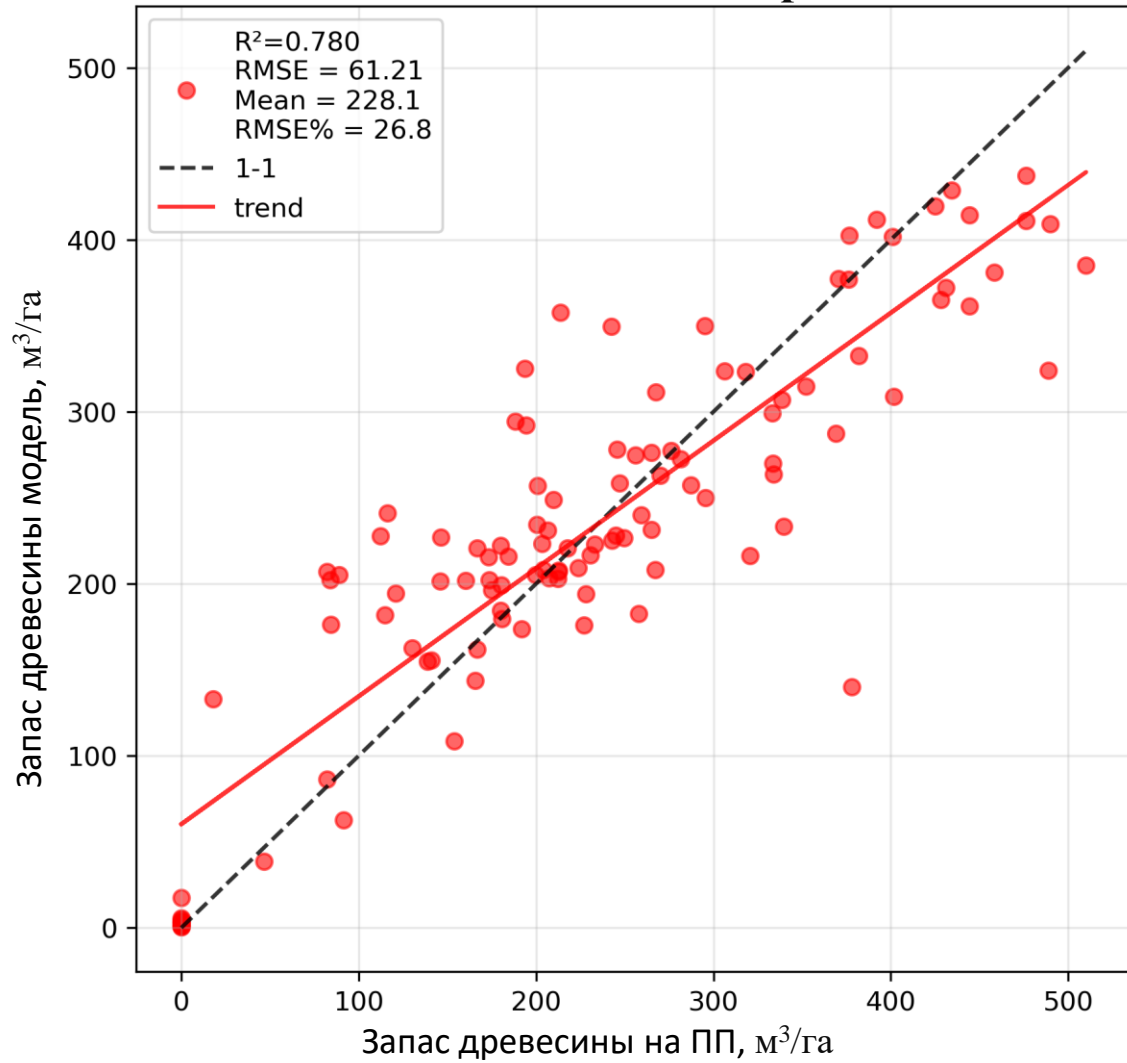
Фильтр, Смещенная выборка



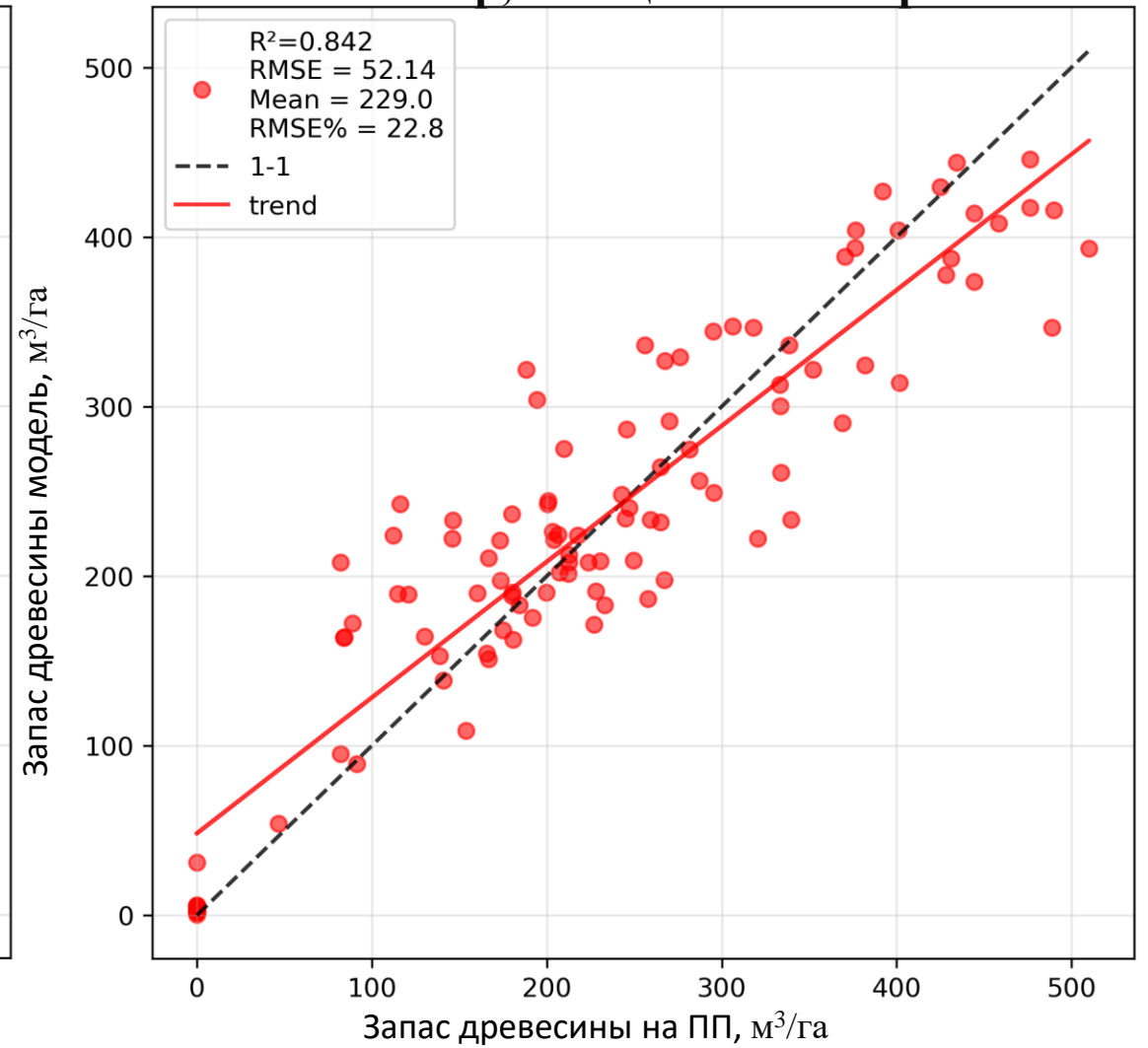
Пространственная коррекция расположения ПП

Статистическая фильтрация, смещение

Смещенная выборка



Фильтр, Смещенная выборка



Выводы

- Для выбранного ТП рост показателей:

Набор	R ²	RMSE, м ³ /га	RMSE%
Исходный	0.595	82.94	36.4
Смещение	0.780	61.21	26.8
Фильтрация + смещение	0.842	52.14	22.8

- Предложенный подход продемонстрировал повышение точности и устойчивости модели:
 - **Коррекция выборки:** фильтрации и добавления безлесных эталонов в области низких запасов стволовой древесины;
 - **Пространственная коррекция:** оптимизация расположения ПП в локальном окне для максимальной корреляции спектральных признаков с запасом стволовой древесины;
 - **Идентификация выбросов:** определение и исключение нерепрезентативных ПП.

Дальнейшие исследования

- Усовершенствование процедуры выбора оптимального направления смещения ПП в локальном окне: применение взвешенных показателей дисперсии и вариации
- Применение предложенного метода коррекции на других ТП, формирование локальных карт запаса стволовой древесины