

Оценка запаса стволовой древесины на основе данных Sentinel-2

Ворушилов И.И.¹, Барталев С.А.¹, Егоров В.А.¹

¹ *Институт космических исследований РАН*

**2025 г.
Москва**

Цель работы

- Построить оценку запаса стволовой древесины по данным высокого разрешения и адаптировать алгоритмы (предобработки и самой оценки запаса стволовой древесины) работающие для среднего разрешения (MODIS) под высокое
- Проанализировать полученный временной ряд запасов стволовой древесины

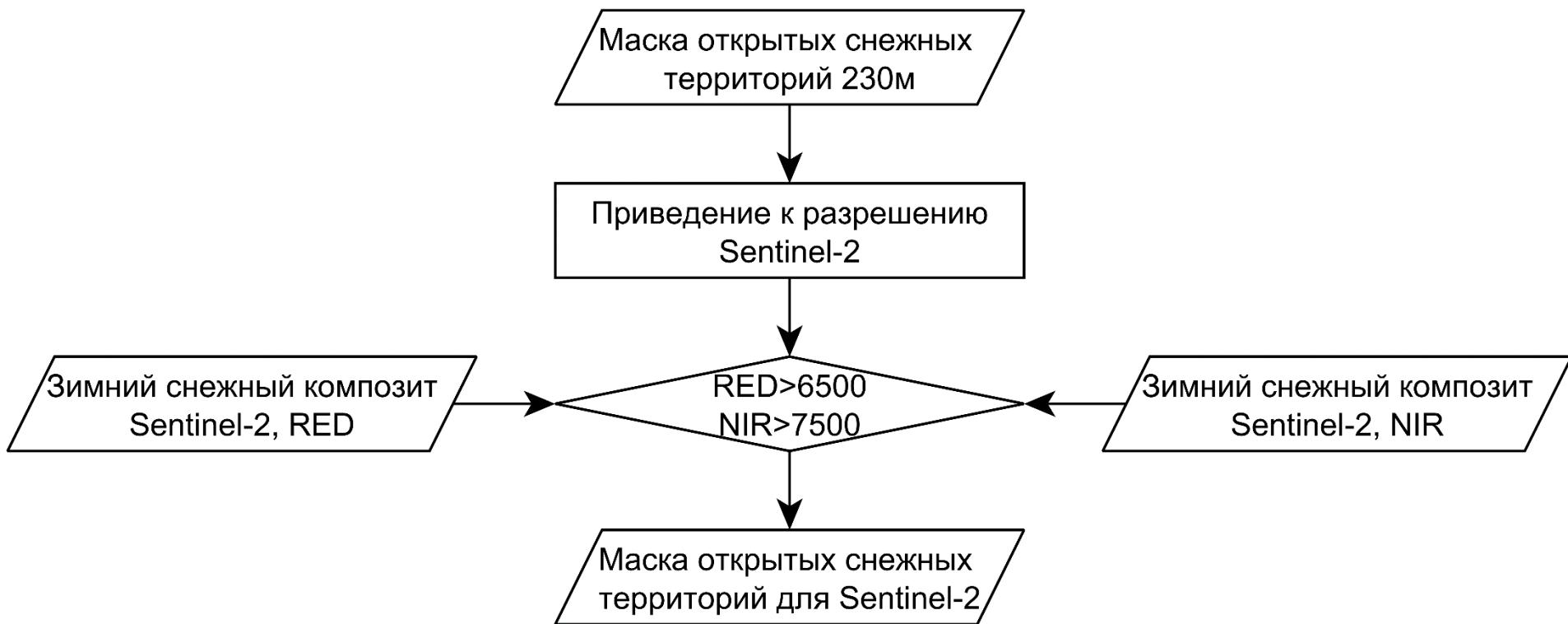
Входные данные

- Композитные снежные изображения Sentinel-2 за 2023 и 2024 год
- Временной ряд оценок запаса стволовой древесины в разрешении 230 м, полученный ранее из данных MODIS
- Временной ряд зимних снежных композитных изображений MODIS
- Летние, весенние, осенние данные MODIS за 2010 год

Адаптация модели предобработки спутниковых данных

Как и при работе с разрешением 230м в силу погодных причин яркость снега может иметь локальные однородные вариации. Поэтому аналогично с предобработкой данных на разрешении MODIS предлагаю провести радиометрическую нормализацию зимнего снежного композита по территории. Для этого необходима маска открытых снежных территорий. Так как у нас пока что нет всех необходимых продуктов в разрешении Sentinel-2 для её построения, возьмём за основу маску открытых снежных территорий построенную в разрешении MODIS. Приведём её к разрешению Sentinel-2 и отфильтруем то, что точно не может быть открытым снегом по яркости композитов.

Схема построения маски открытых снежных территорий



Пространственная радиометрическая нормализация

Используя полученную маску открытых снежных площадей рассчитываю локальные средние значения яркости открытого снежного покрова. Места где в силу отсутствия открытых снежных территорий это невозможно среднюю яркость открытого снега пространственно интерполирую. Далее делю композитное изображение на яркость снега открытых снежных территорий и умножаю на медианное значение раstra средней яркости открытых снежных территорий.

$$\text{КСЯ}_{xy_p} = \frac{\text{КСЯ}_{xy} * med(\text{КСЯ}_{\text{снега}})}{\text{КСЯ}_{xy_снега}}$$

Различие яркости лиственных и хвойных лесов

Предполагаю что Red и NIR каналов Sentinel-2 по аналогии с Terra-MODIS, содержат неопределенность, относящуюся к вегетационным составляющим хвои. Данная неопределенность приводит к тому, что оценка запаса стволовой древесины становится заниженной/занышенной (в зависимости от выбранного канала и породного состава леса). Поскольку погрешности имеют разный знак - предлагаю сложить два признака, чтобы уменьшить влияние неопределённостей. Для того чтобы полностью избавится от погрешностей связанных с данными неопределенностями - предлагаю умножить один из них на коэффициент равный отношению этих неопределённостей.

Различие яркости лиственных и хвойных лесов

$$\begin{aligned}NIR &= NIR_{ist} \pm \Delta NIR \\RED &= RED_{ist} \mp \Delta RED \\k &= \frac{\Delta R}{\Delta N} \\I &= RED + k * NIR\end{aligned}$$

$$\begin{aligned}RED_{listv} &= RED_{ist_listv} + \Delta RED_{listv} \\NIR_{listv} &= NIR_{ist_listv} - \Delta NIR_{listv} \\RED_{hv} &= RED_{ist_hv} - \Delta RED_{hv} \\NIR_{hv} &= NIR_{ist_hv} + \Delta NIR_{hv}\end{aligned}$$

$$\begin{aligned}I &= RED_{ist} \mp \Delta RED + k(NIR_{ist} \pm \Delta NIR) \\I &= RED_{ist} + k * NIR_{ist}\end{aligned}$$

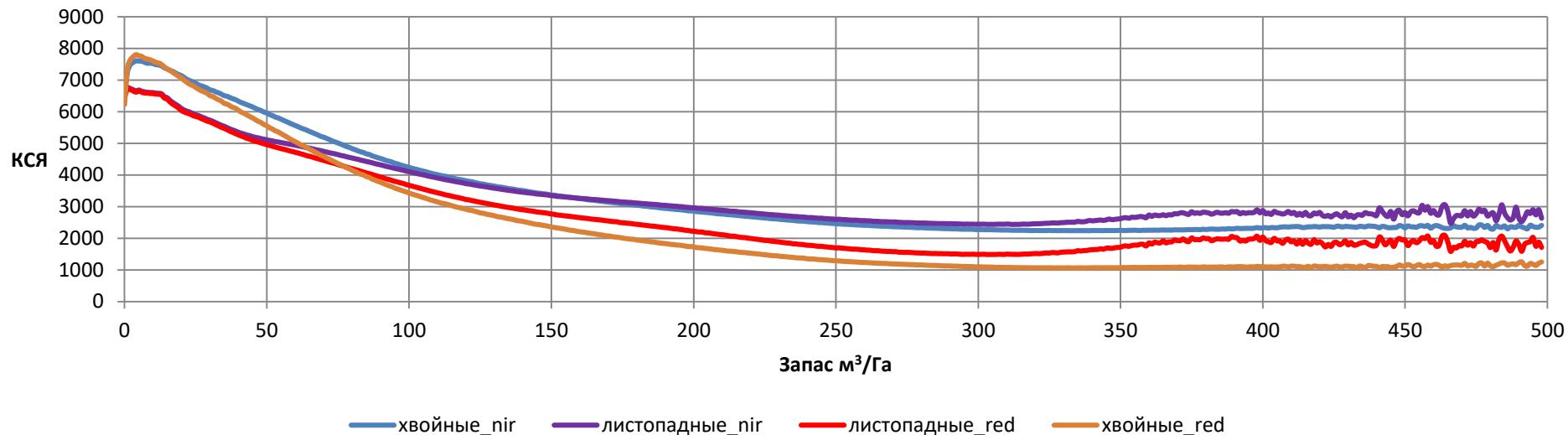
$$\begin{aligned}RED_{el} + k * NIR_{el} &= RED_{listv} + k * NIR_{listv} \\k &= \frac{RED_{listv} - RED_{el}}{NIR_{el} - NIR_{listv}}\end{aligned}$$

Подготовка данных для оценки на более высоком разрешении

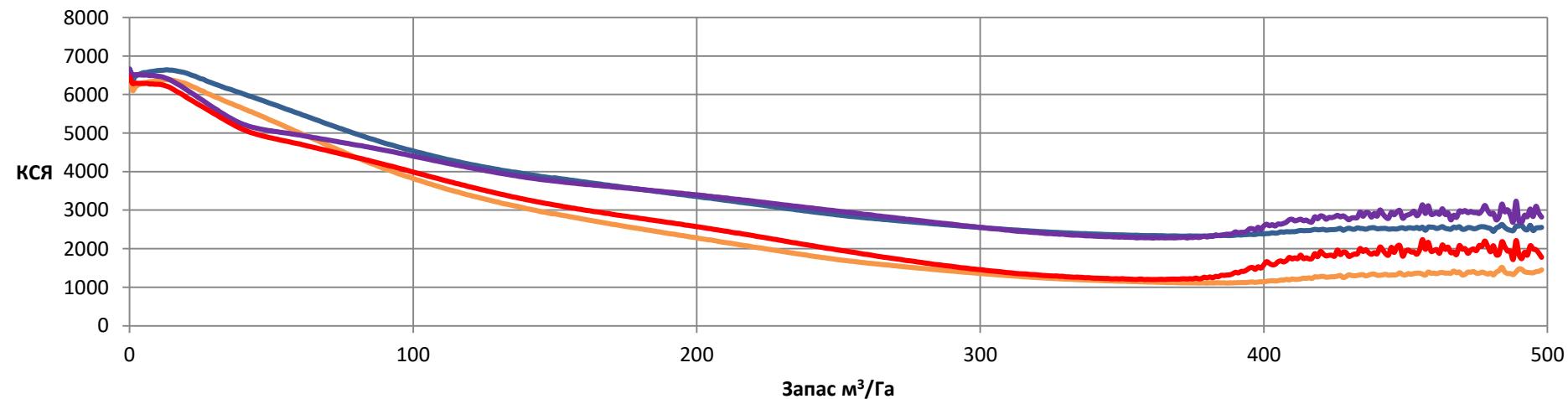
Планирую использовать в качестве опорной выборки один из результатов оценки запаса стволовой древесины по данным MODIS. Чтобы рассчитать коэффициент “ k ” и подготовить модель для регрессора предлагаю привести данные Sentinel-2 до разрешения MODIS. После найти нужный коэффициент и построить признаки. Обучить модель на приведённых данных и применить её к данным высокого разрешения.

Различие яркости лиственных и хвойных лесов

Зависимость медианной яркости Sentinel-2 от запаса

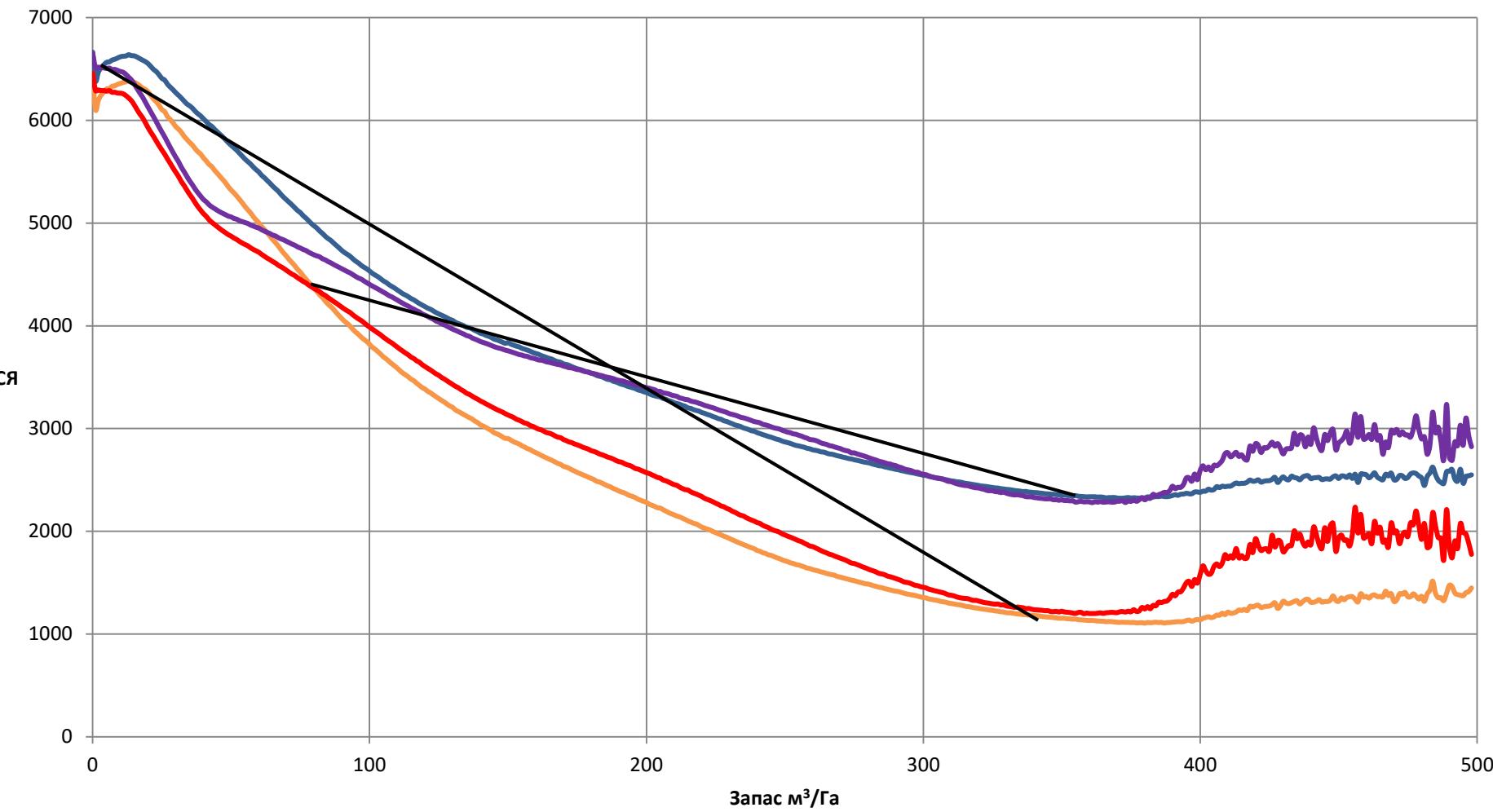


Зависимость медианной яркости MODIS от запаса



Различие яркости лиственных и хвойных лесов

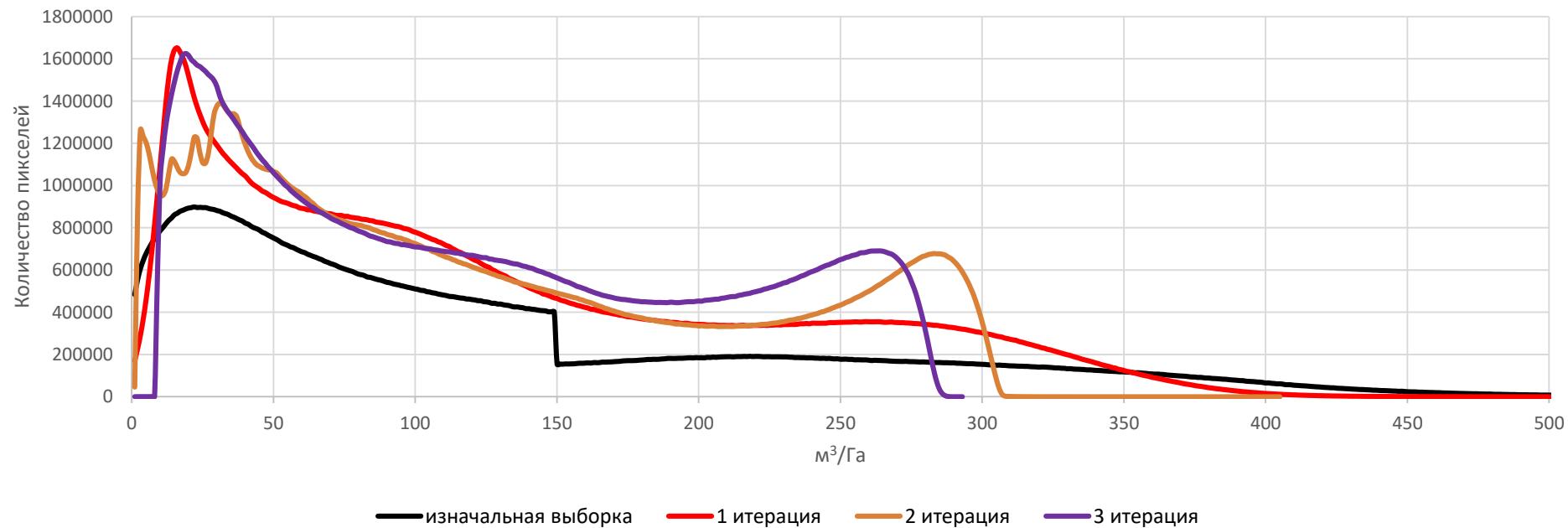
Зависимость медианной яркости MODIS от запаса



Погрешности связанные с множественным использованием методов машинного обучения

При итеративном применении методов машинного обучения (использовании результата в качестве новой опорной выборки) были выявлены изменения в гистограмме. Она сжимается и кластеризуется.

Гистограммы запасов за 2010 год



Нарастает смещение связанное с множественным использованием методов машинного обучения

Для преодоления этих эффектов, предлагается частично переносить опорную выборку (это можно делать так, как мы получаем результат за тот же год на котором и обучаемся) и при переносе на данные Sentinel-2 не использовать в качестве обучающей выборки запас полученный с помощью RF за 2023 год, а наоборот привести признаки на основе MODIS за 2010 год к признакам Sentinel-2 так, чтобы модель обучившись на них повторила результат (суммарный для России запас по MODIS) применив её к признакам Sentinel. В итоге у нас вместо 3 итераций с запасом будет 2 и 1 итерация с признаками.

Алгоритм действий подготовки данных для корректной обучения модели

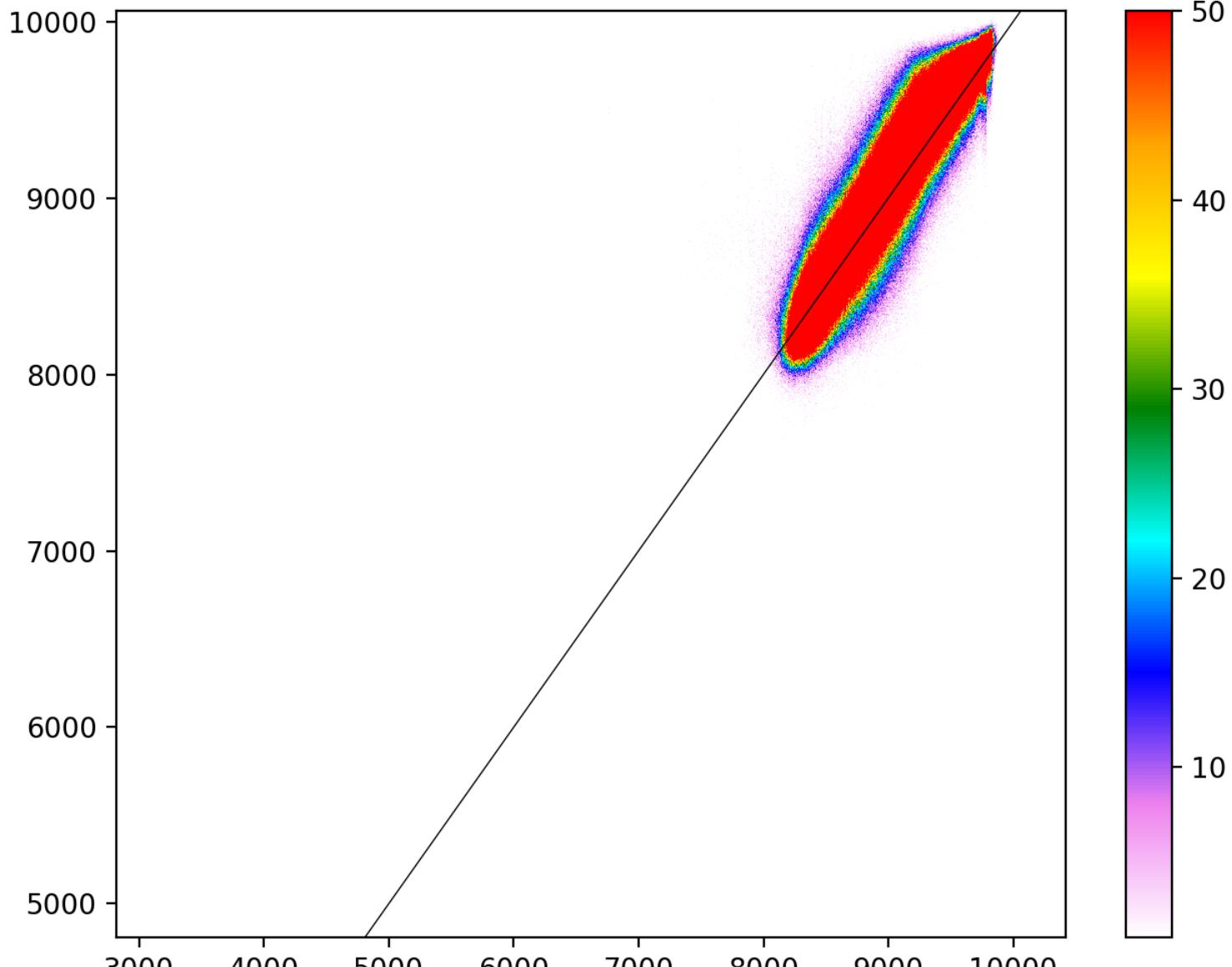
- 1) Приводим данные Sentinel-2 до пространственного разрешения MODIS
- 2) Используем приведённые данные Sentinel-2 в качестве опорной выборки, а данные MODIS в качестве признаков и получаем временной ряд приведённых данных Sentinel-2, полученных на основе MODIS
- 3) Для 2023 года производим сравнение приведённых данных Sentinel-2 с его аналогом восстановленным по MODIS
- 4) Строим уравнение связи вида $y = a * x + b$, находим a и b
- 5) Применяем a и b к ряду приведённых данных Sentinel-2, полученных на основе MODIS

Схема подготовки данных для корректной обучения модели



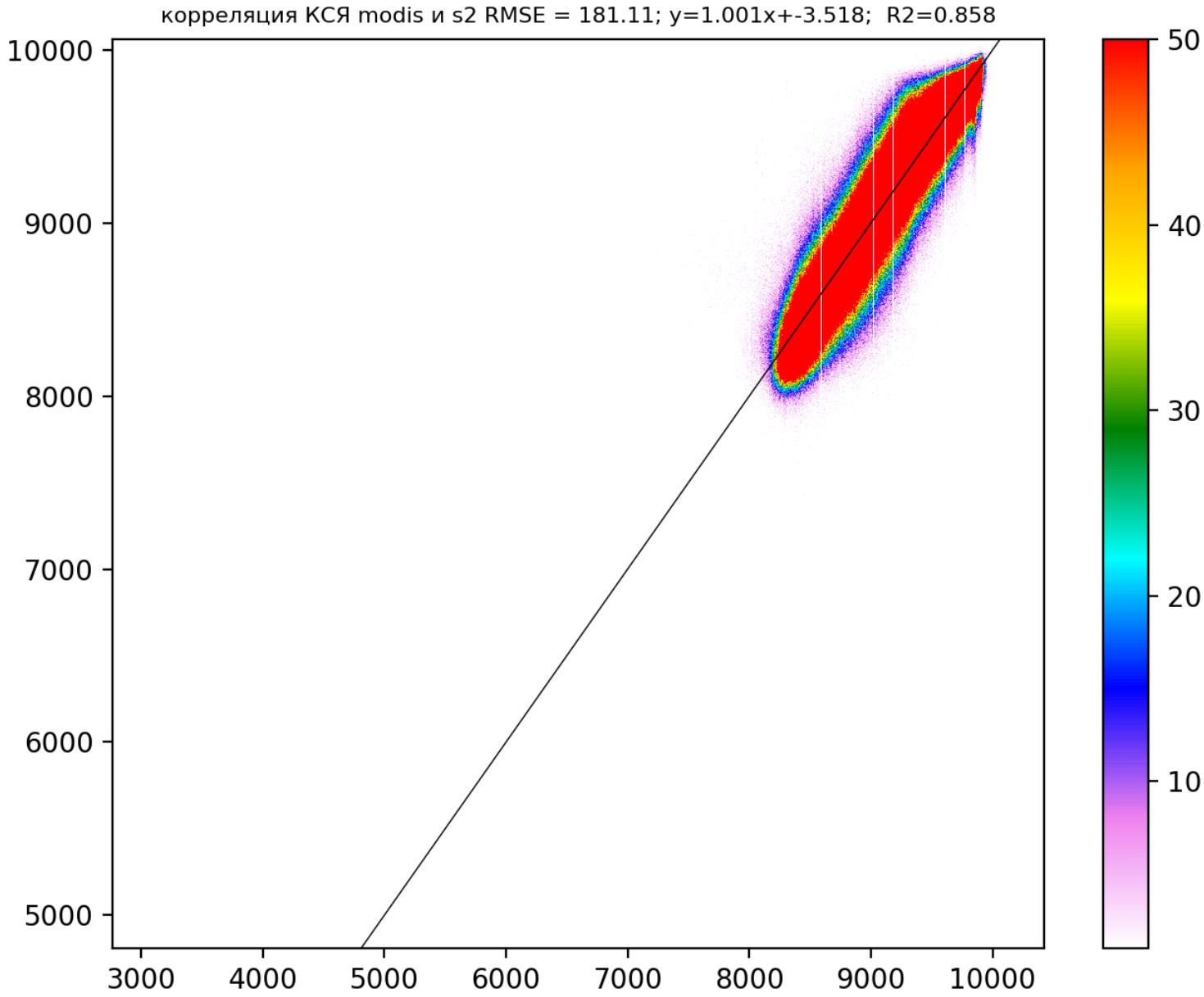
Построенный признак Sentinel за 2023 год на основе данных

корреляция КСЯ modis и s2 RMSE = 229.294; $y=1.019x+ -106.256$; $R^2=0.858$



Построенный признак Sentinel за 2023 год (результат RF) на основе данных MODIS

Построенный признак Sentinel за 2023 год на основе данных

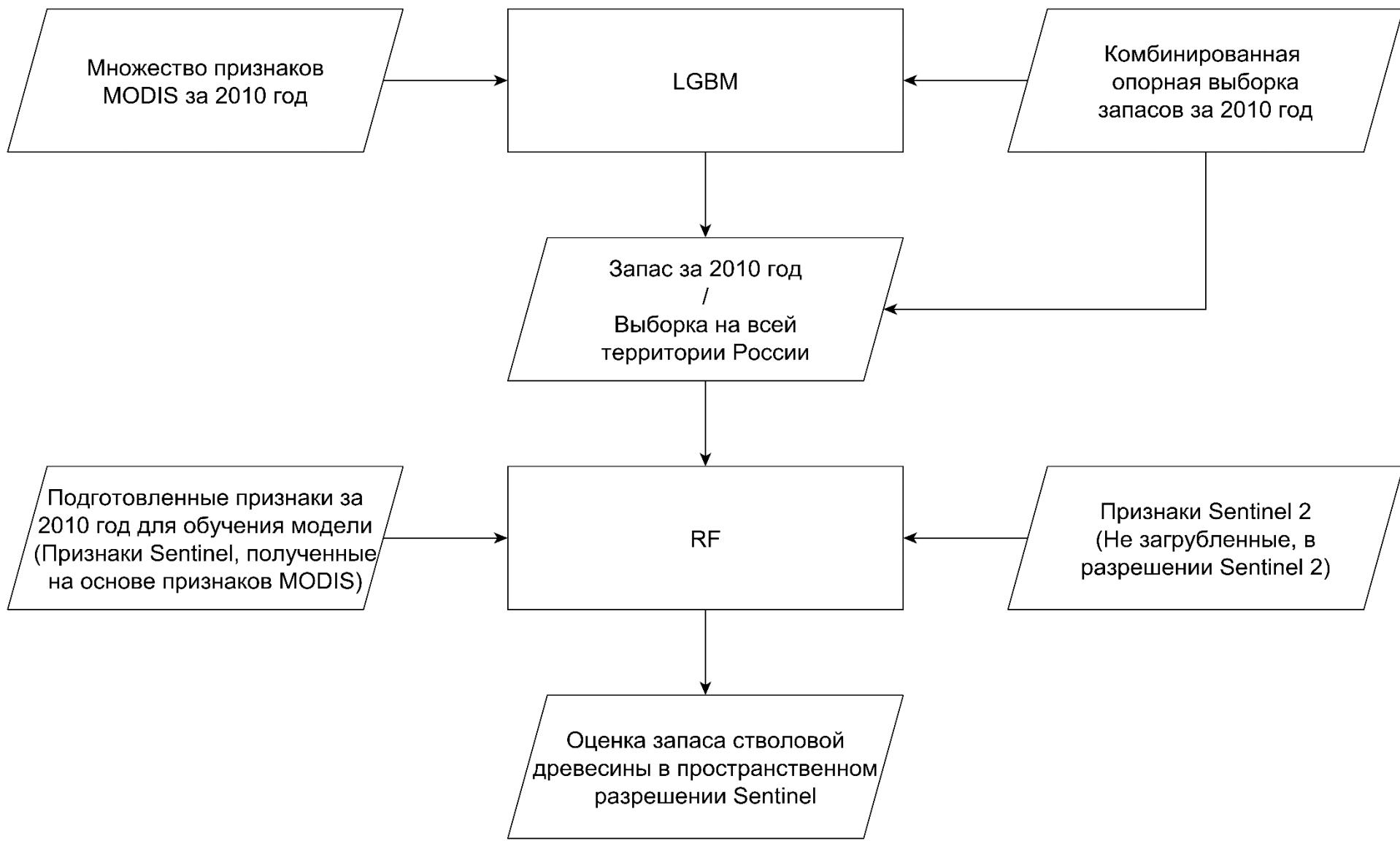


Скорректированный Построенный признак Sentinel за 2023 год (результат RF) на основе данных MODIS

Алгоритм действий для оценки запасов в разрешении Sentinel

- 1) Так как выборка у нас из двух источников и представлена не на всей территории России, используем методы машинного обучения, чтобы расширить её, для избежания ошибок воспользуемся огромным количеством признаков MODIS (41 признак). Получив тем самым оценку запасов стволовой древесины за 2010 год.
- 2) Зная что применённый метод может искажать гистограмму, скорректируем результат с помощью опорной выборки (перенесём большие запасы от туда).
- 3) Используя подготовленные признаки за 2010 год (Признаки Sentinel, полученные на основе признаков MODIS и в разрешении MODIS) и скорректированную опорную выборку (в разрешении MODIS) для обучения модели.
- 4) Обученную модель применим на признаки Sentinel 2 изначального разрешения, получив тем самым оценку запаса стволовой древесины в разрешении Sentinel-2

Схема оценки запасов в разрешении Sentinel



Оценка результатов

Для оценки результатов будем сравнивать интегральные характеристики, такие как суммарный запас стволовой древесины на лесах России за определённый год. А также будем строить двумерные гистограммы между опорной выборкой (разрешение MODIS) и результатом оценки по данным Sentinel-2. Для этого пересчитаем запас стволовой древесины на пиксель MODIS, чтобы результаты были одного пространственного разрешения

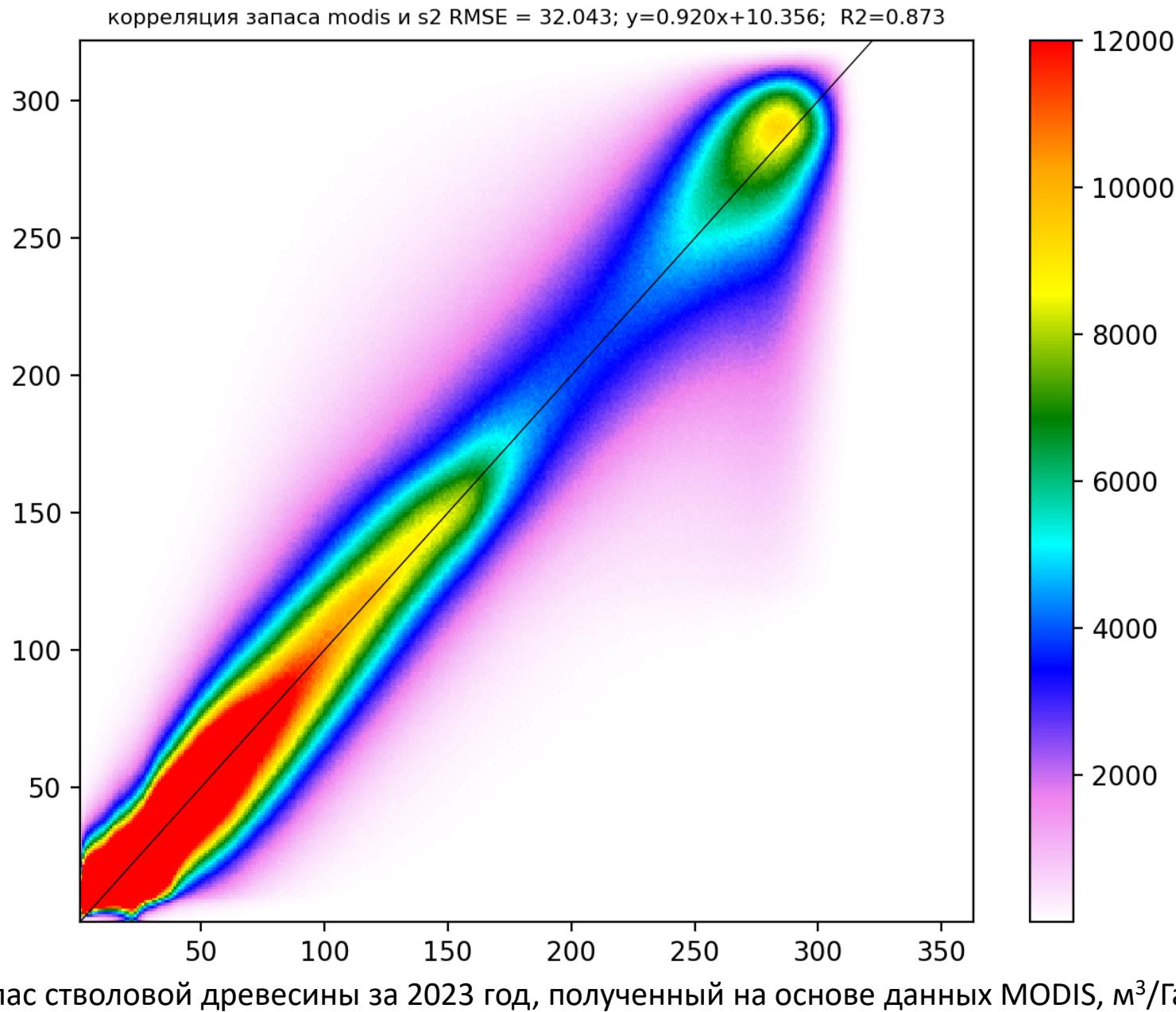
Интегральный запас стволовой древесины на территории России за 2023 год по определённой маске лесной территории для оценки полученной только для данных MODIS составляет 131,3 млрд м³

Оценки по данным Sentinel-2 за 2023 год на той же маске при соблюдении всех действий описанных выше суммарный интегральный запас составляет 130,0 млрд м³

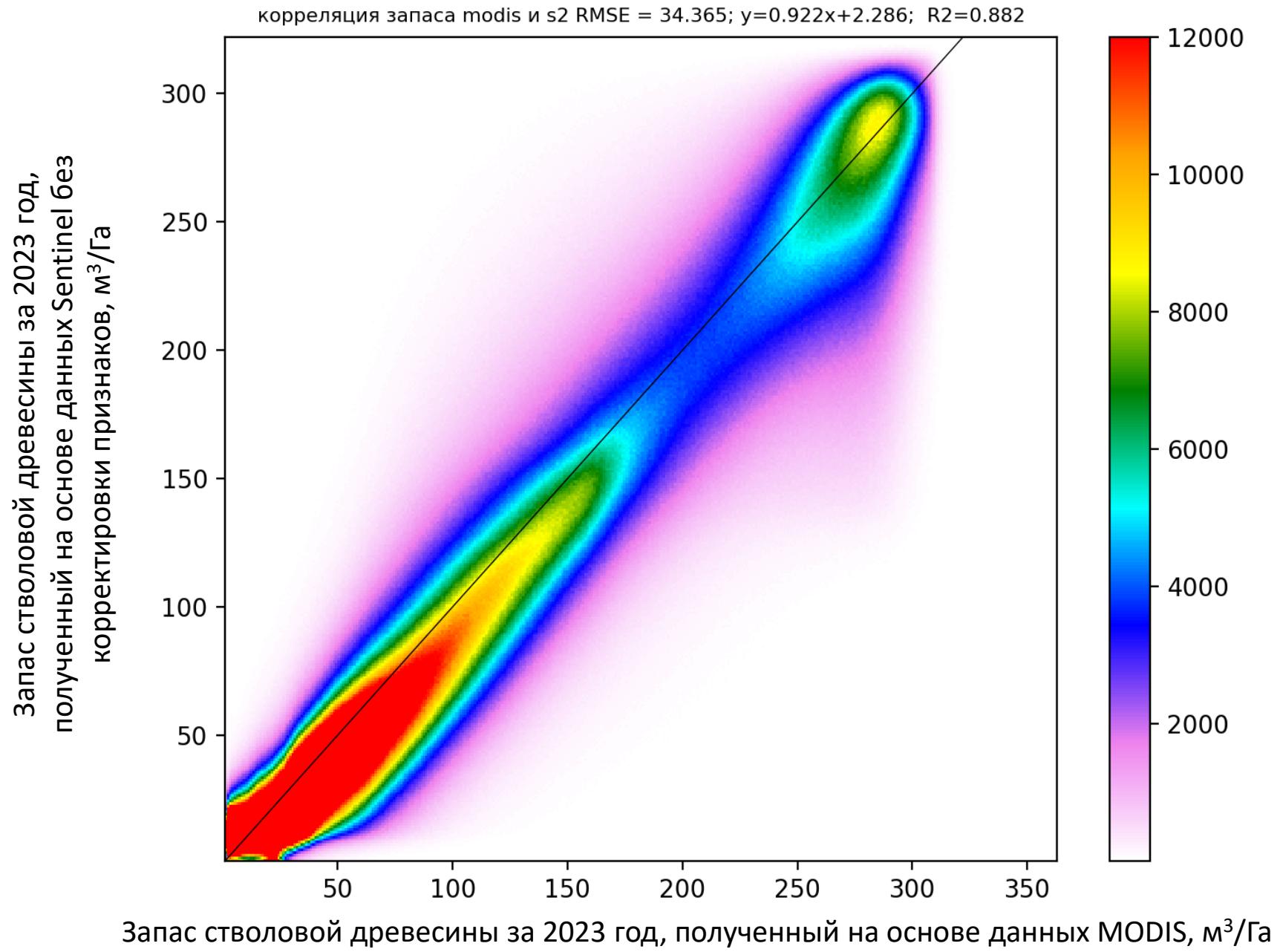
При отступлении от алгоритма указанного выше, а именно если не корректировать признаки для обучения модели суммарный интегральный запас составит 120,0 млрд м³

При дальнейших отступлениях происходит сжатие гистограммы, наиболее выраженно с правого края, но и с левого тоже происходит сдвиг.

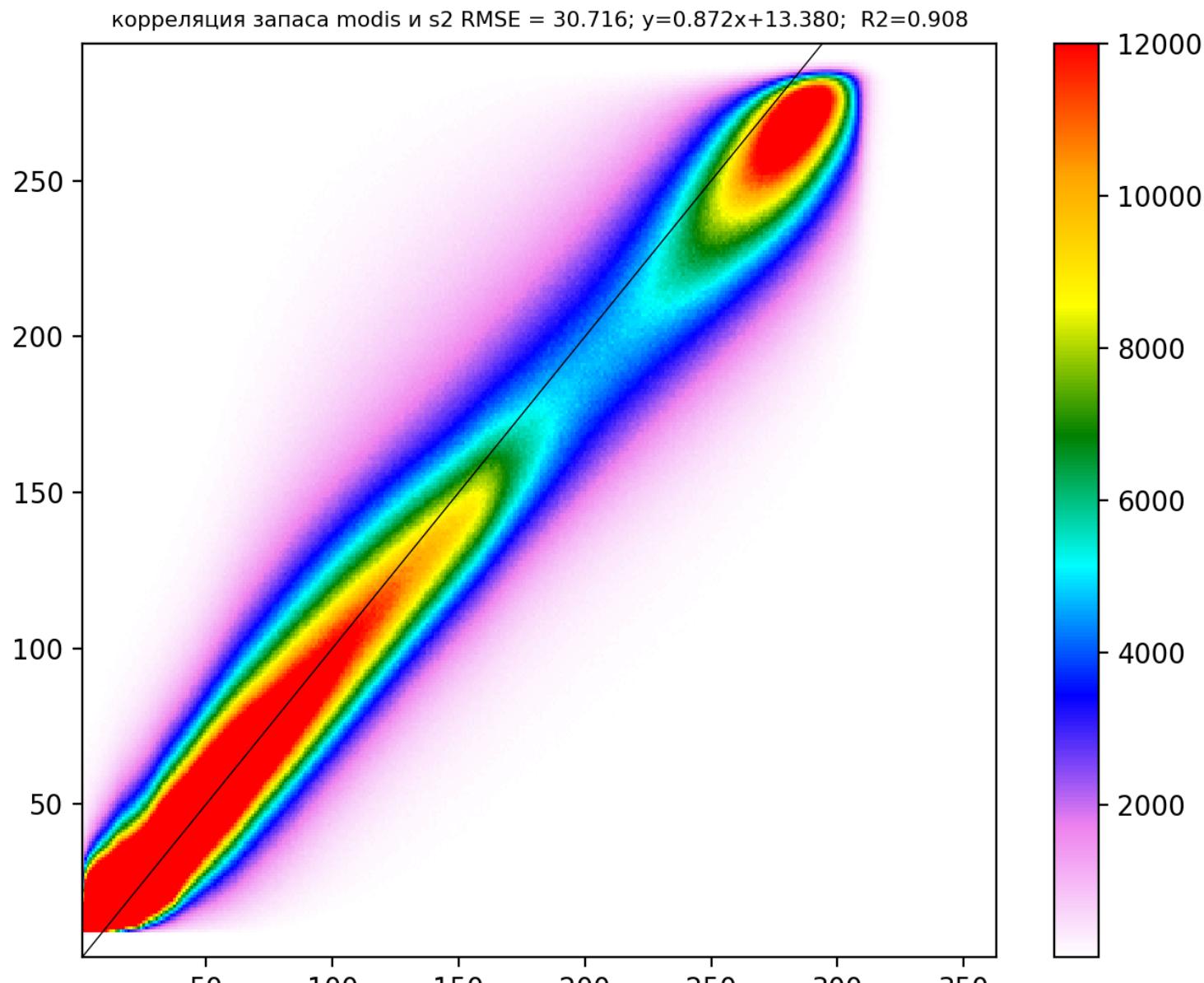
Запас стволовой древесины за 2023 год,
полученный на основе данных Sentinel, м³/Га



Запас стволовой древесины за 2023 год,
полученный на основе данных Sentinel без
корректировки признаков, м³/Га



Запас стволовой древесины за 2023 год,
полученный на основе данных Sentinel, если
использовать оценку по MODIS за 2023 год в
качестве опорной выборки, м³/Га

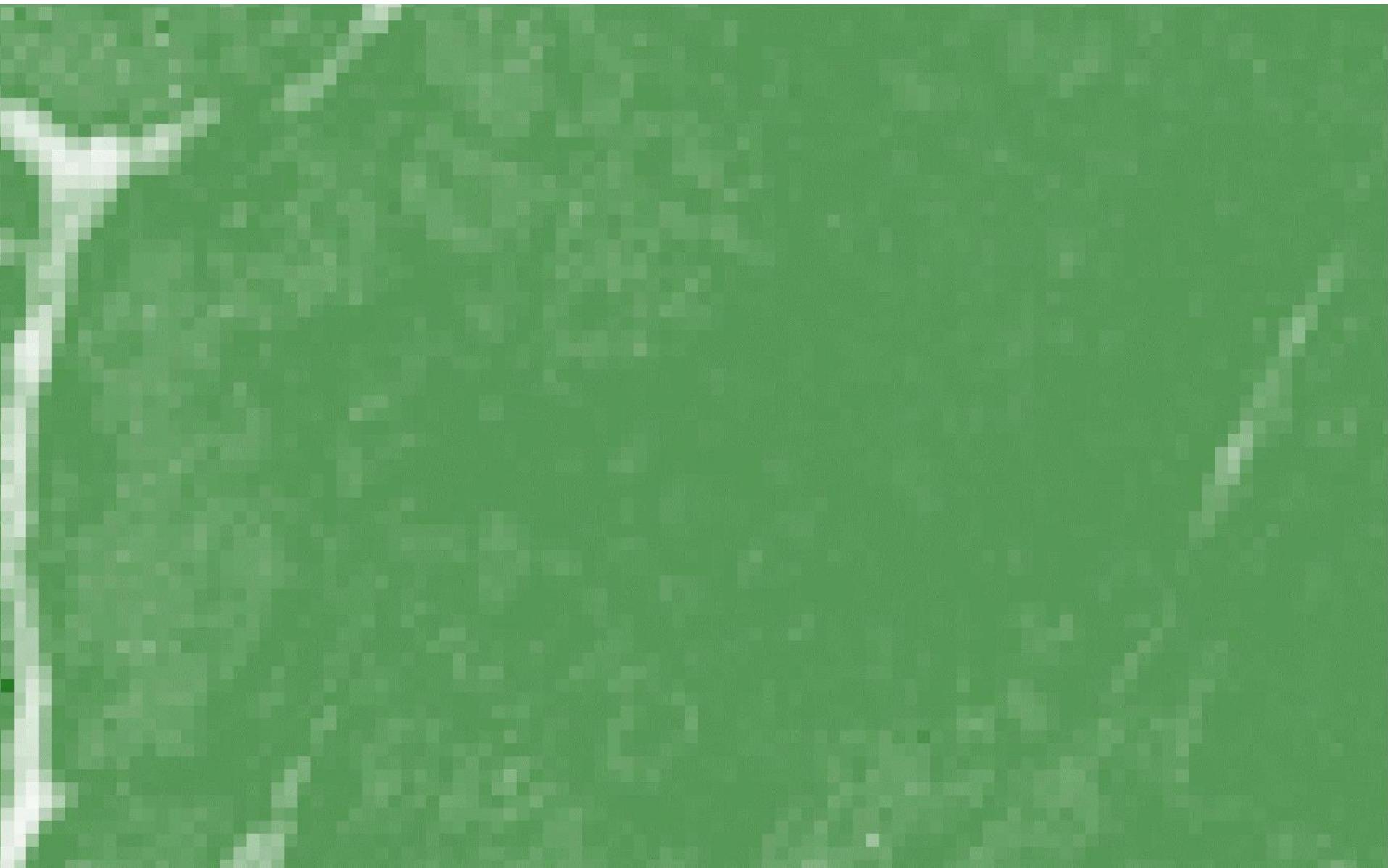


Запас стволовой древесины за 2023 год, полученный на основе данных MODIS, м³/Га

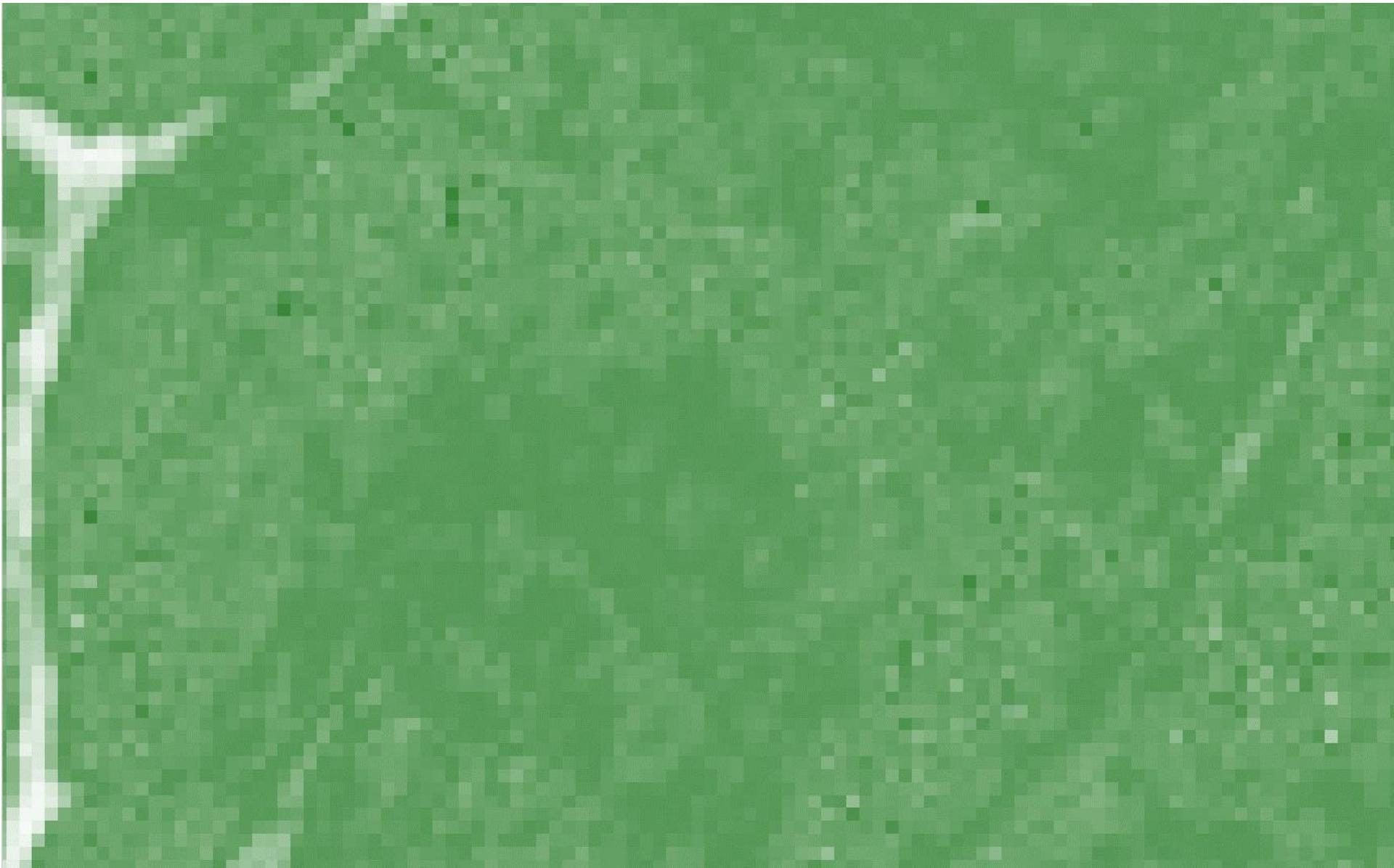
Снимок высокого разрешения



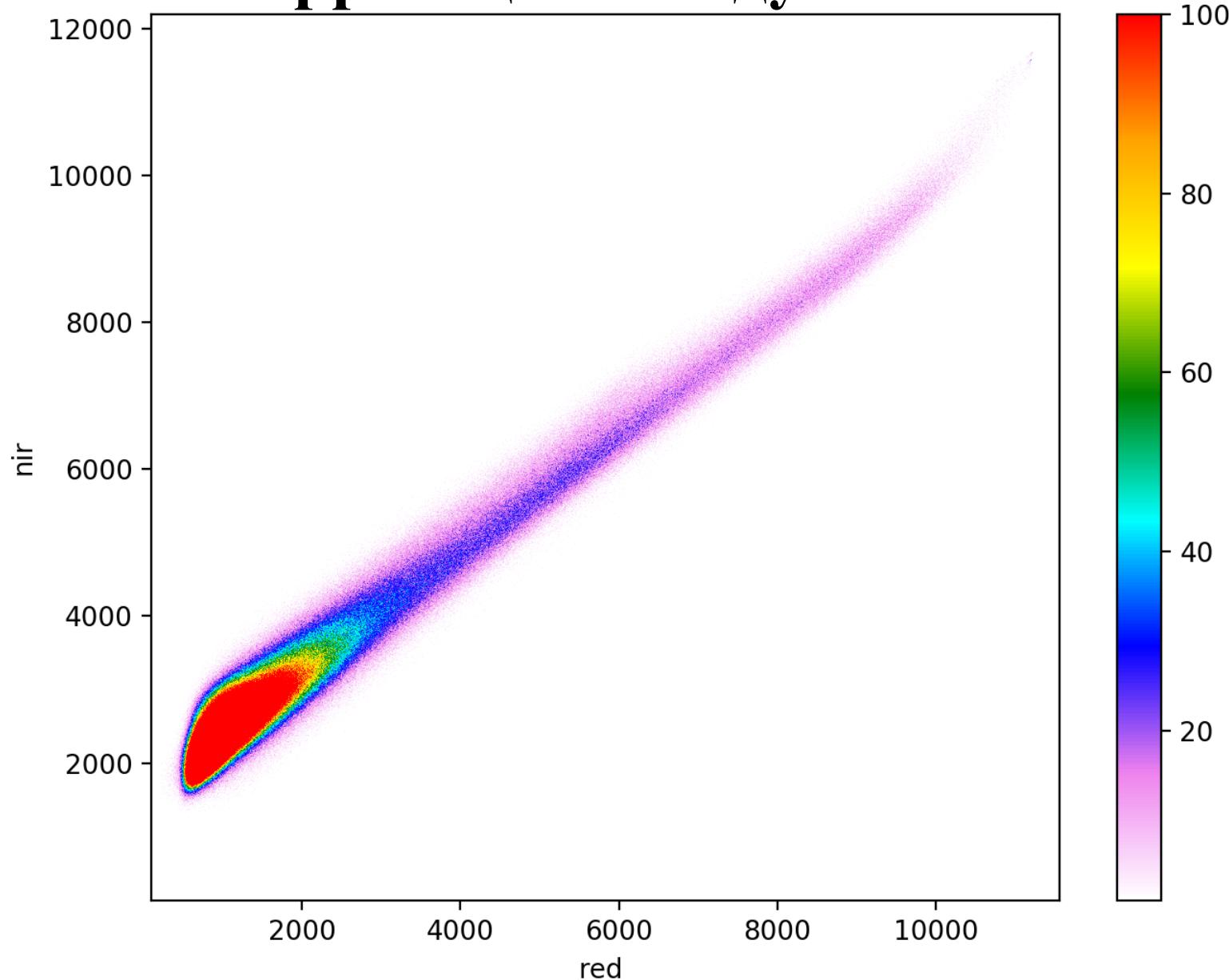
Оценка запаса по S-2 за 2023 год



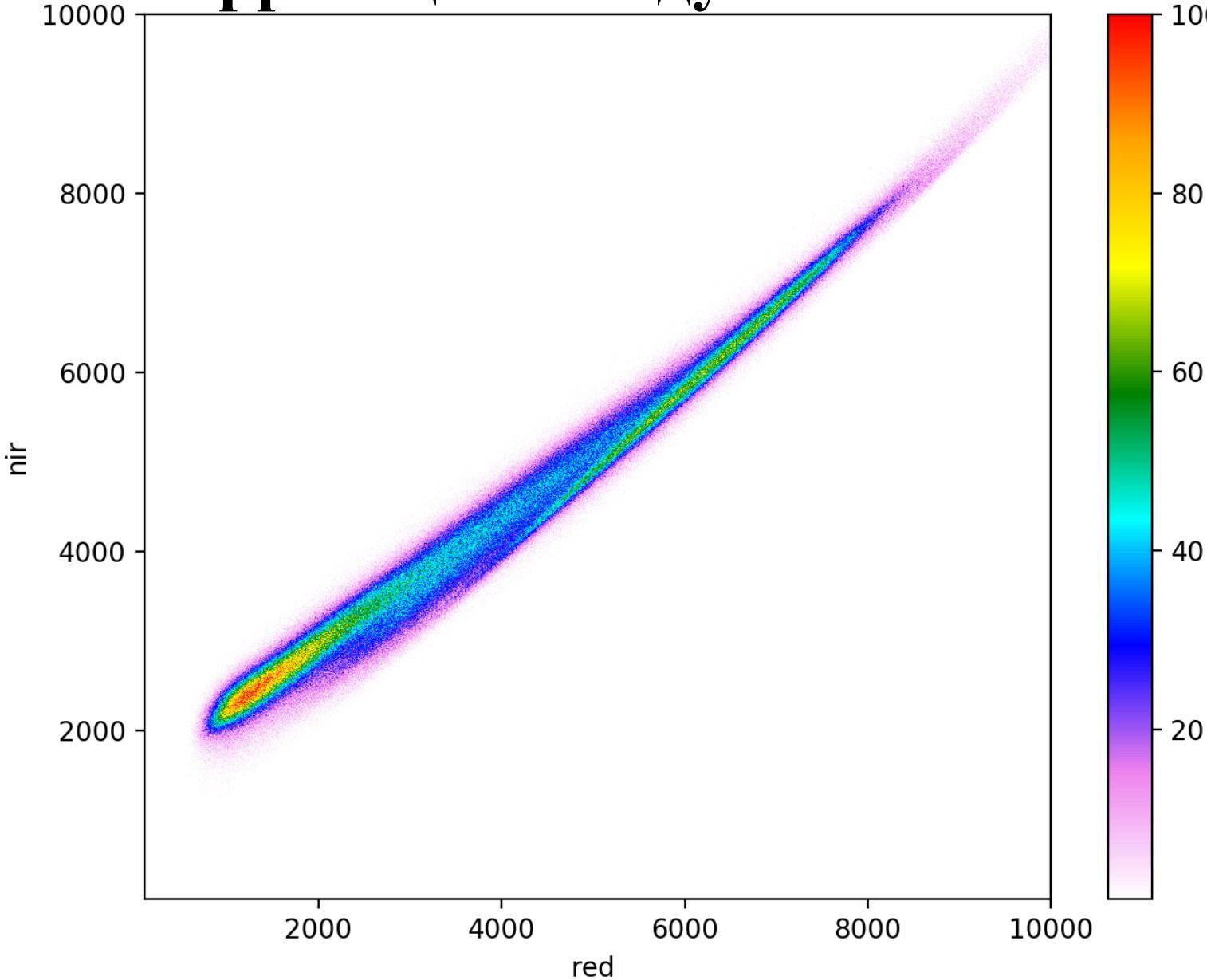
Оценка запаса по S-2 за 2022 год



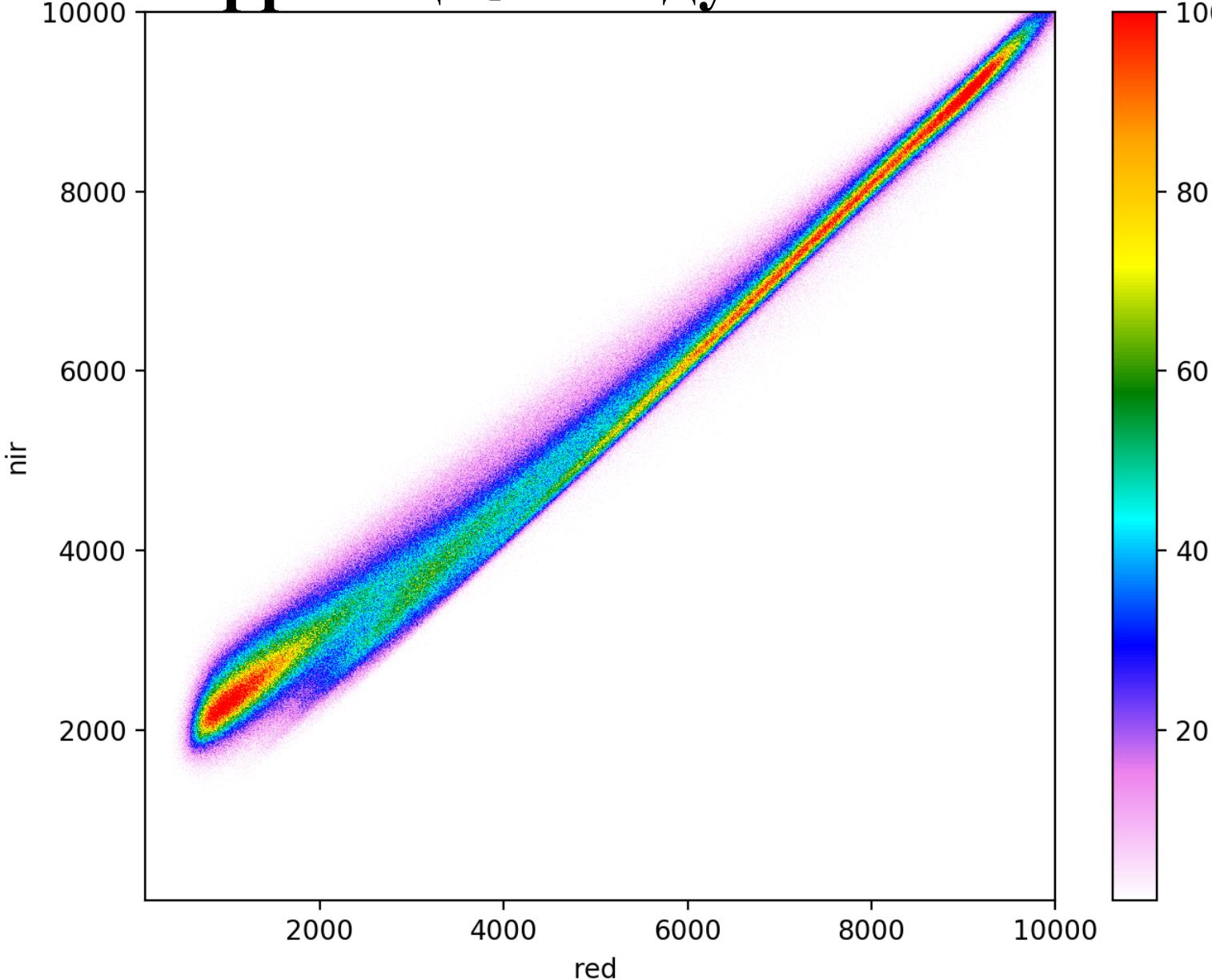
Корреляция между red и nir S-2



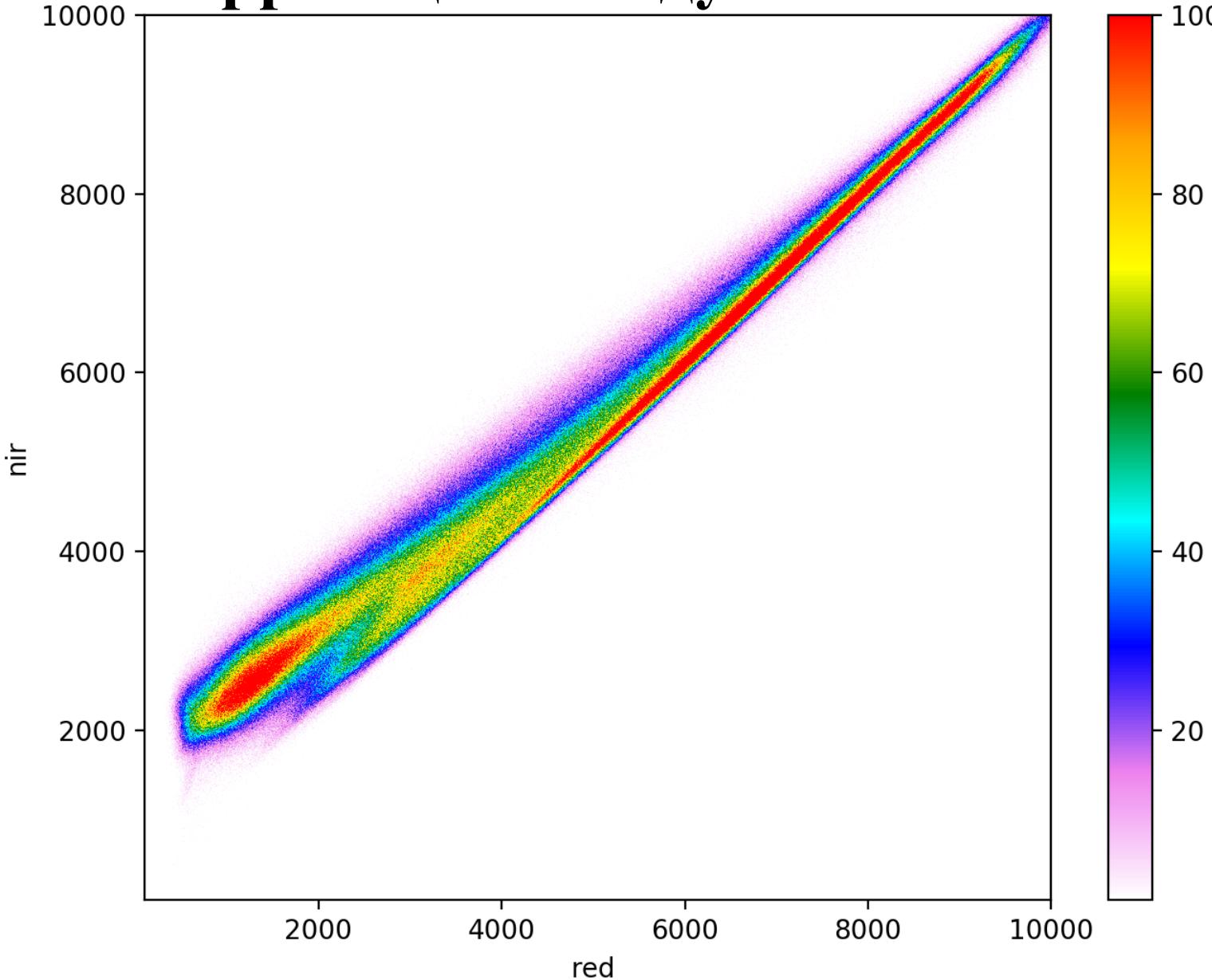
Корреляция между red и nir MODIS



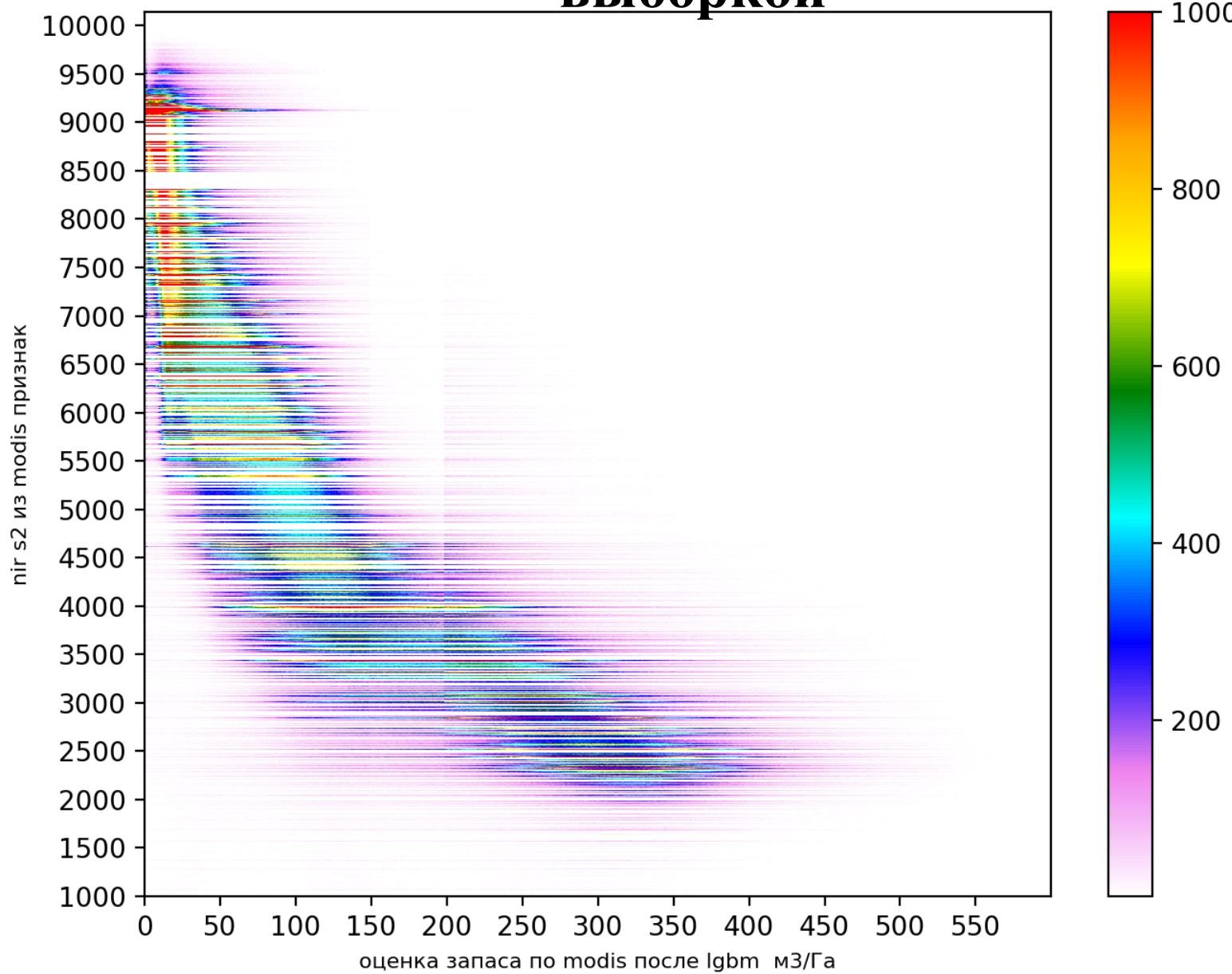
Корреляция между red и nir S-2 40м



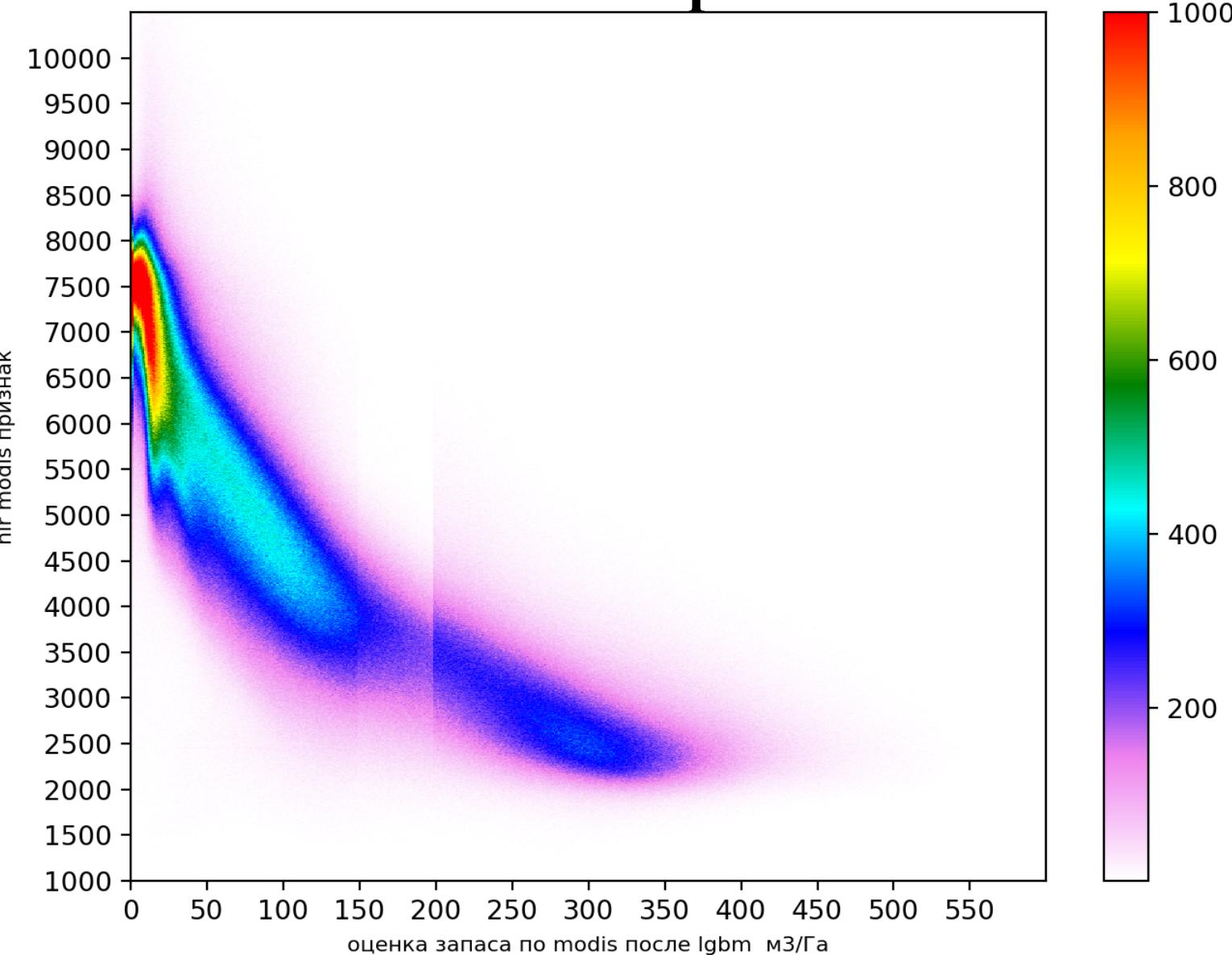
Корреляция между red и nir S-2 80м



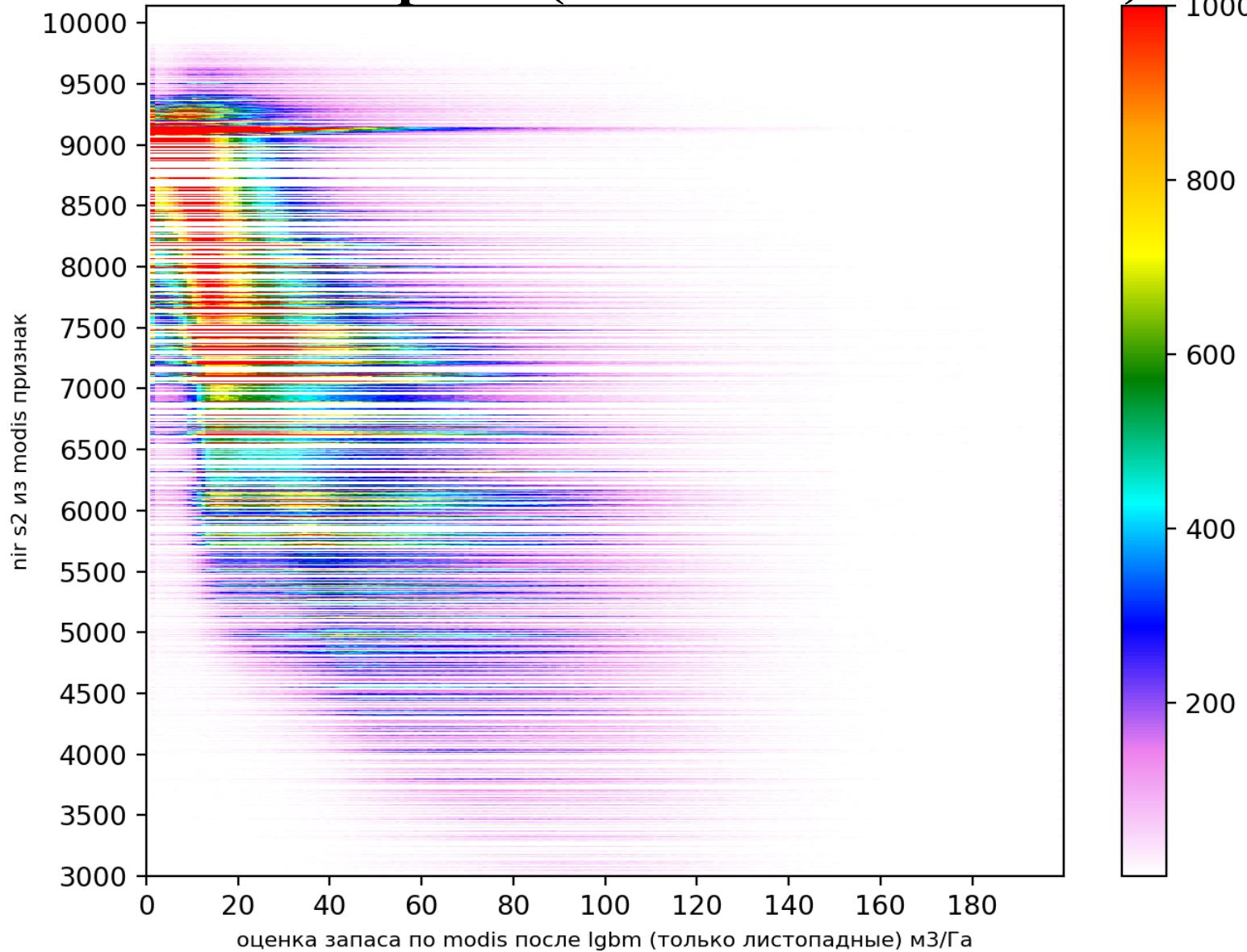
Корреляция между nir S-2 и обучающей выборкой



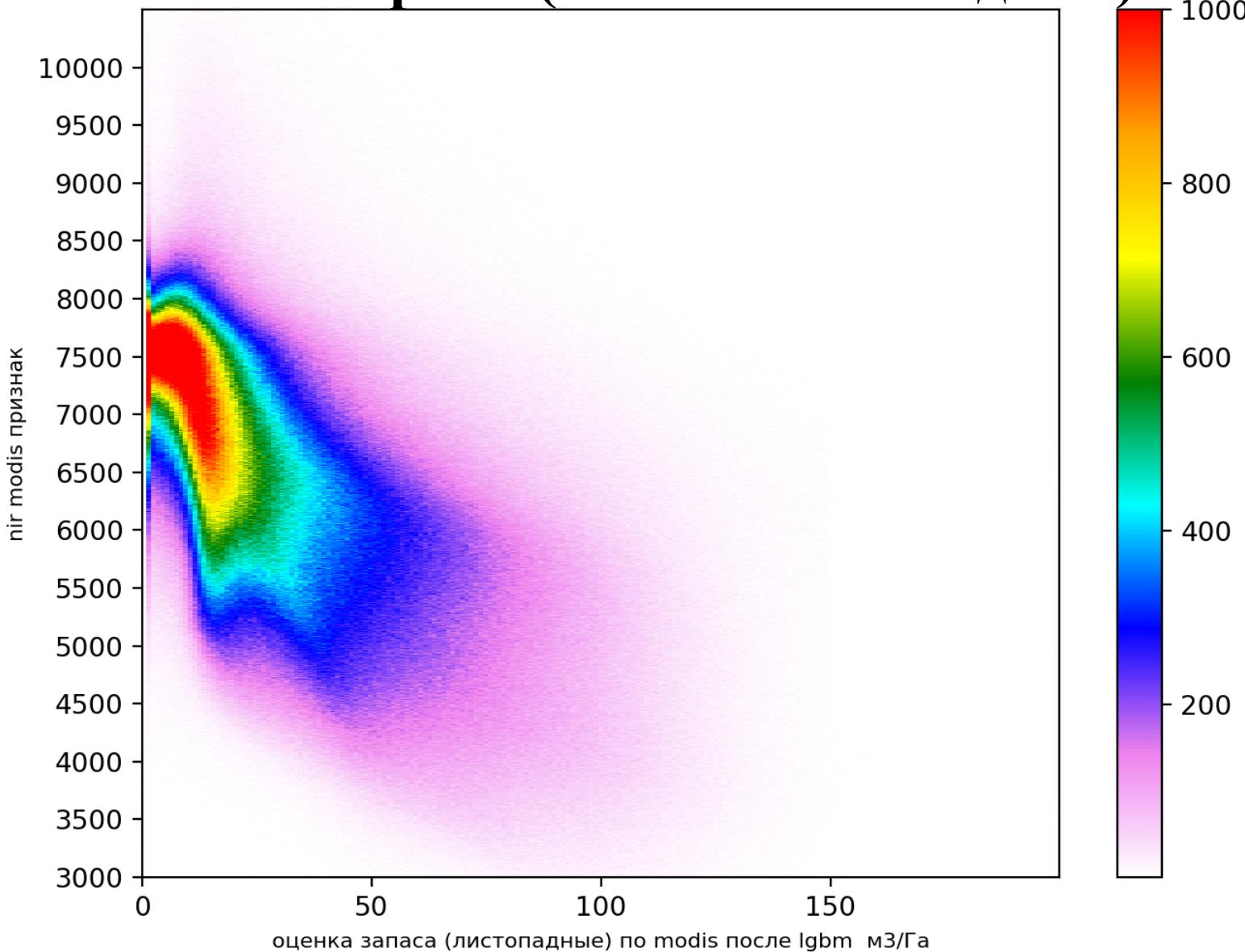
Корреляция между nir MODIS и обучающей выборкой



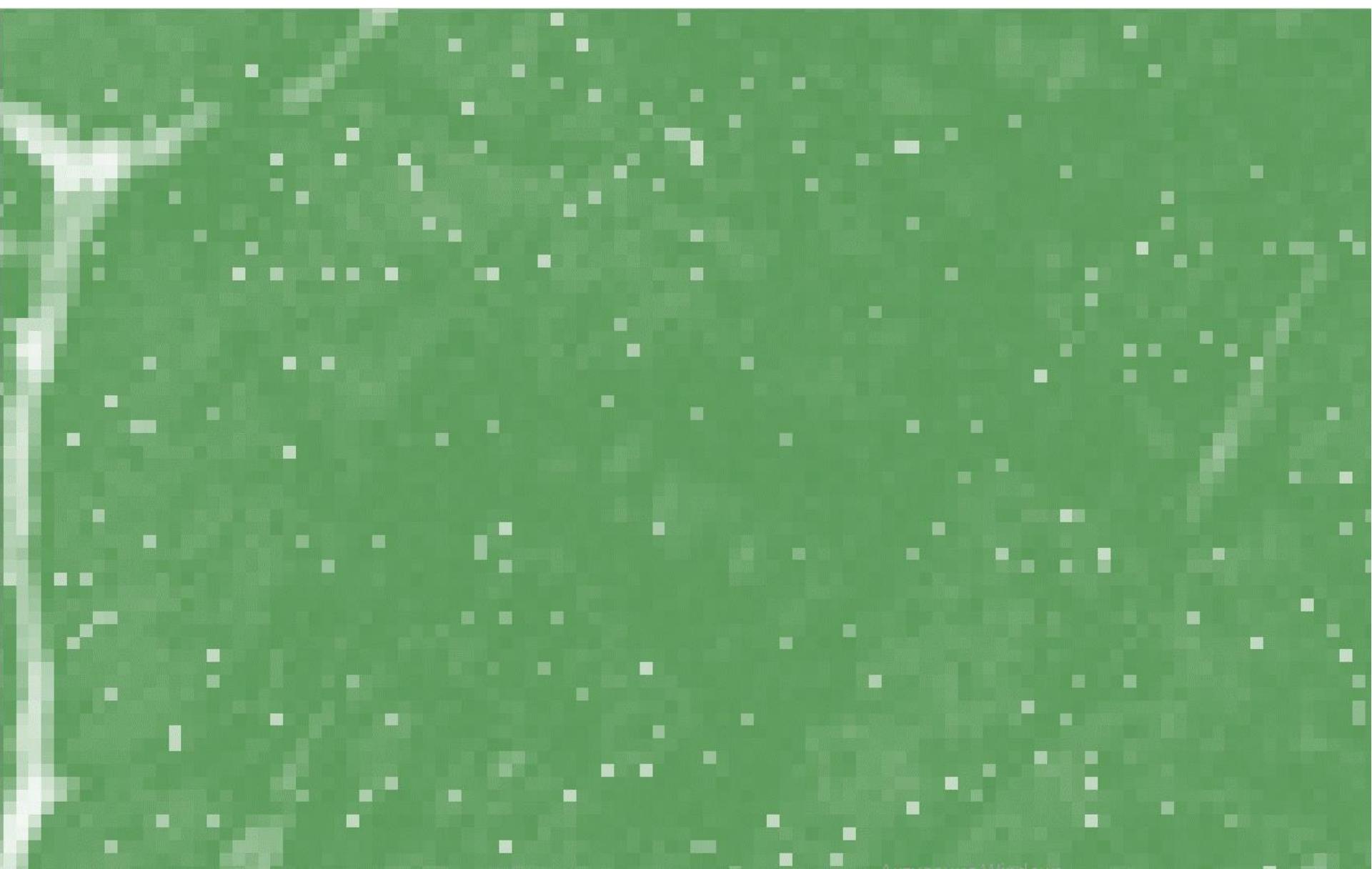
Корреляция между nir S-2 и обучающей выборкой (только листопадные)



Корреляция между nir MODIS и обучающей выборкой (только листвопадные)



Оценка запаса отдельно для листопадных по S-2 за 2022 год



Выводы

Описанная в работе методика позволяет перенести оценку запаса стволовой древесины на данные более высокого разрешения, минимизировав при этом погрешности связанные с ограничениями методов машинного обучения. А также минимизировать погрешности на этапах переноса выборки, подготовки модели, подготовки признаков. Тем не менее в продукте могут встречаться локальные значительные погрешности и сейчас идёт работа по их устранению.

Спасибо за внимание.