



**АЛТАЙСКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ**

**X Всероссийская конференция с
международным участием
«Обработка пространственных данных в
задачах мониторинга природных и
антропогенных процессов» (SDM-2025),
посвященная памяти академика Ю.И. Шокина
26-29 августа, г. Белокуриха, Россия**



Валовая первичная продукция юга Западной Сибири по данным MODIS и реанализа с использованием методов машинного обучения

Н.В. Волков, А.А. Лагутин, Е.Ю. Мордвин

**ФГБОУ ВО «Алтайский государственный университет»
ФГБНУ «ФИЦ ИВТ»**

Введение

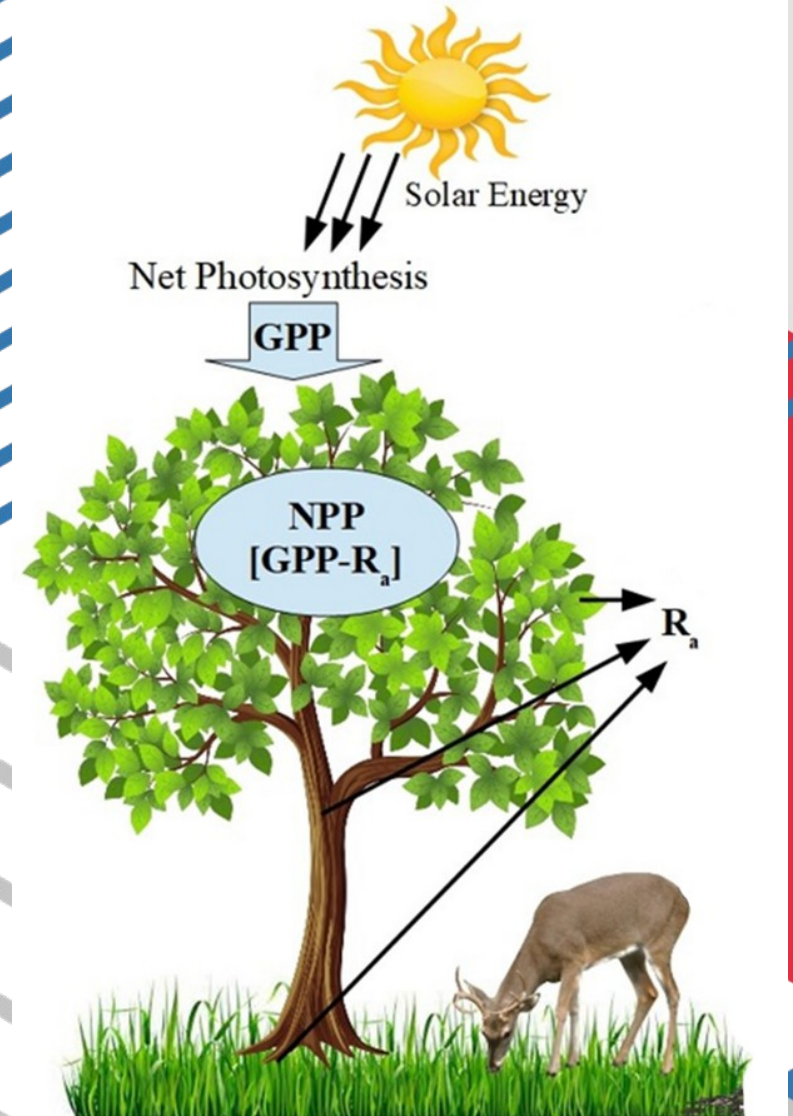
Первичная валовая продукция Gross Primary Production, GPP

GPP

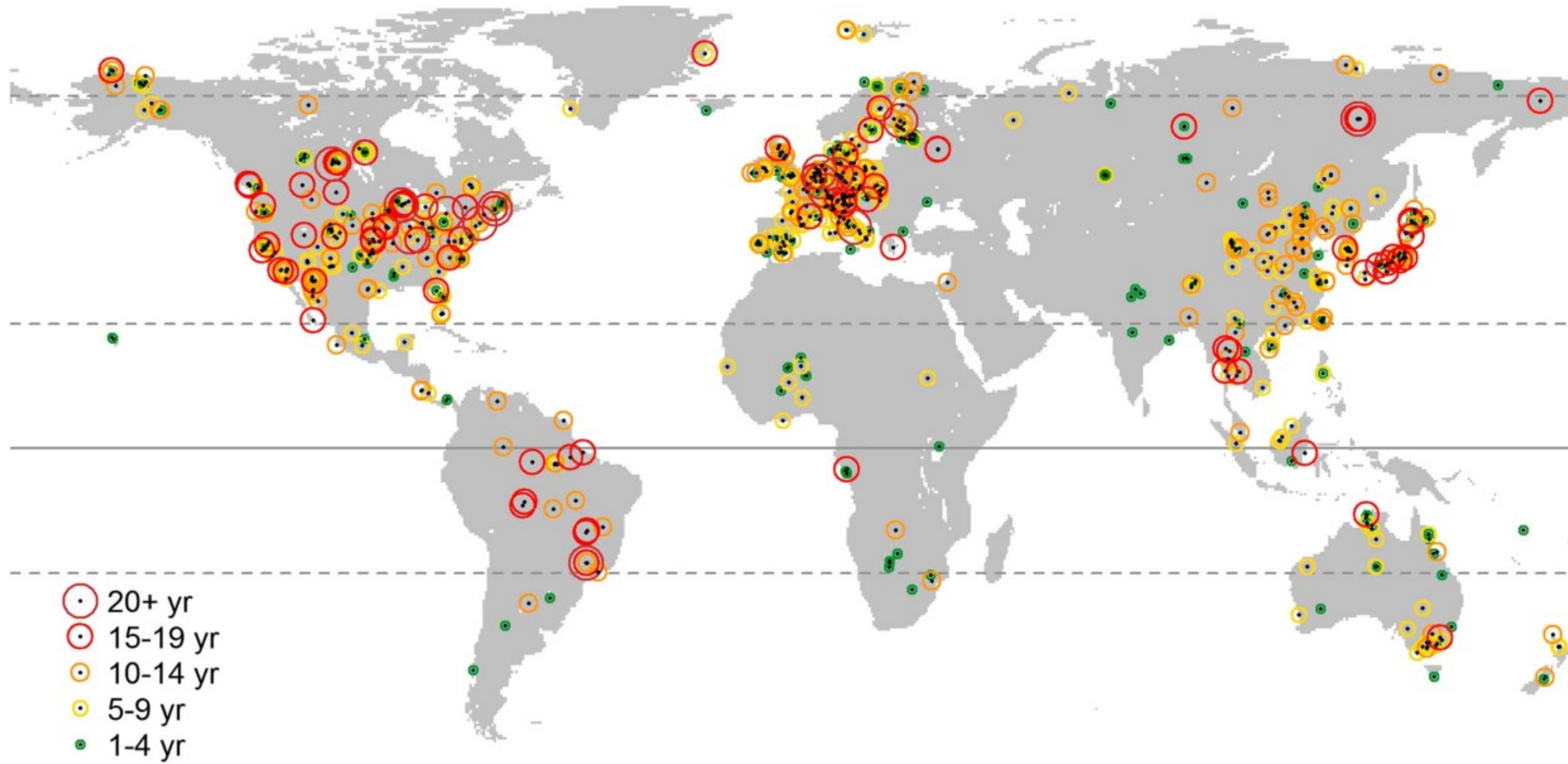
определяет общее количество органического вещества, синтезированного автотрофными организмами (прежде всего растения) из простых неорганических компонент. Основной компонентой переработки является атмосферный углекислый газ (CO_2).

Сегодня существует два основных подхода для получения оценок GPP: проведение наземных мониторинговых измерений и использование данных аэрокосмического зондирования.

10.1007/s00704-024-05158-4



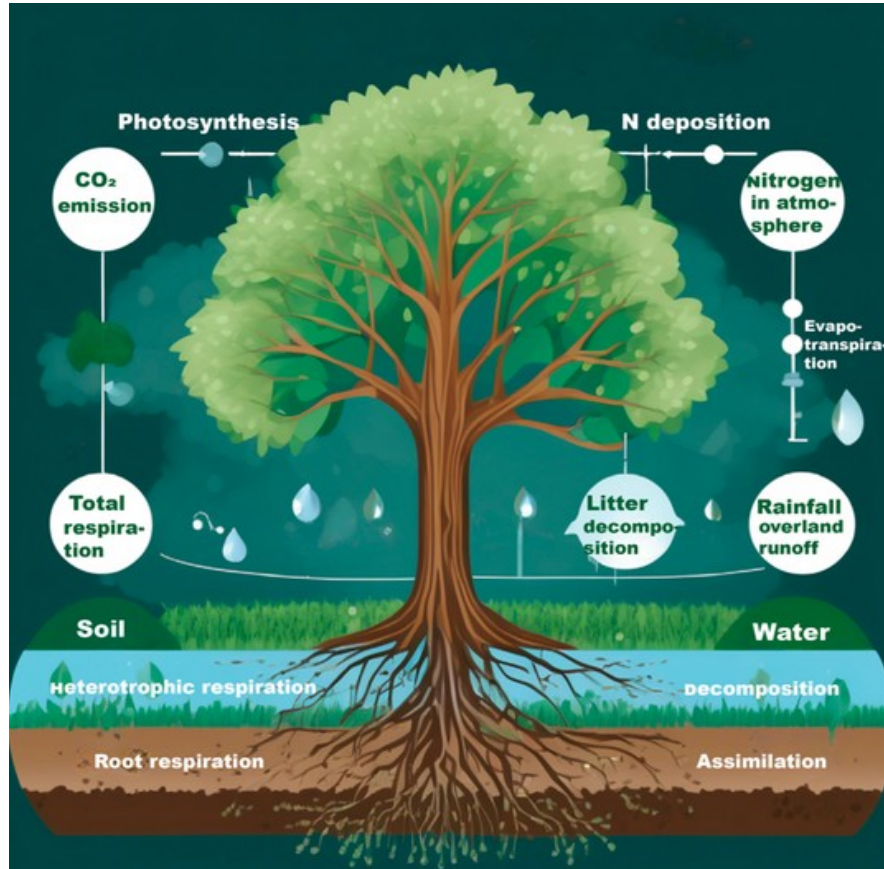
GPP. Глобальный проект FLUXNET



<https://doi.org/10.1002/2016JG003576>

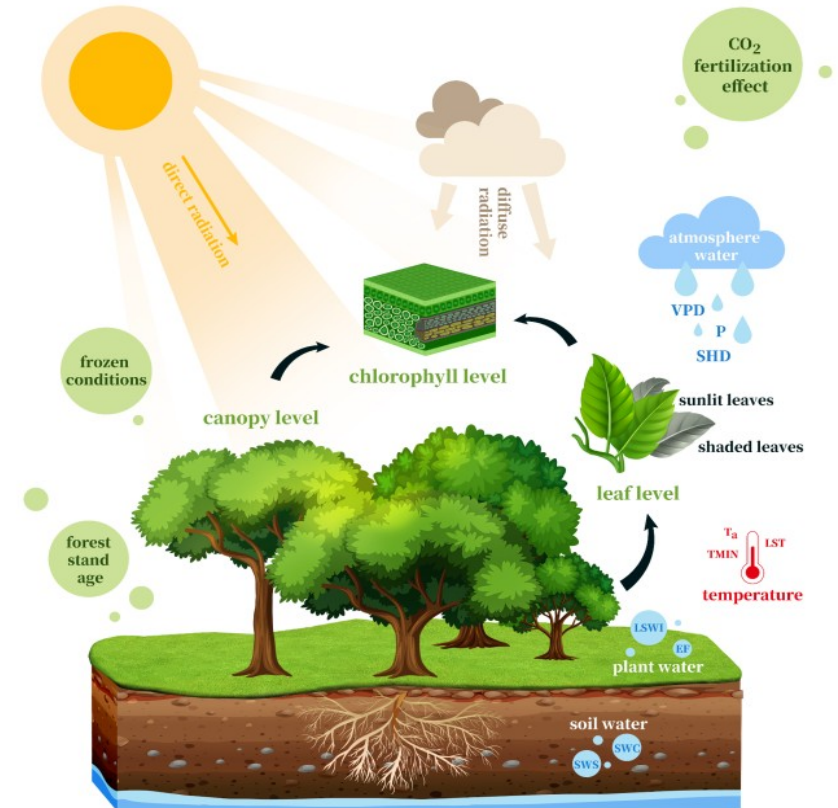
Подходы к оценкам GPP

Биогеохимические модели (BGCM)



10.3390/f16040661

Модели переноса излучения и его усвоения растениями (LULC)



10.1016/j.agrformet.2022.108905



Цель и задачи

Цель и задачи

Цель:

расчеты GPP для территории юга Западной Сибири с использованием данных спектрорадиометра MODIS, реанализа ERA5, цифровой модели рельефа SRTM и методов машинного обучения.

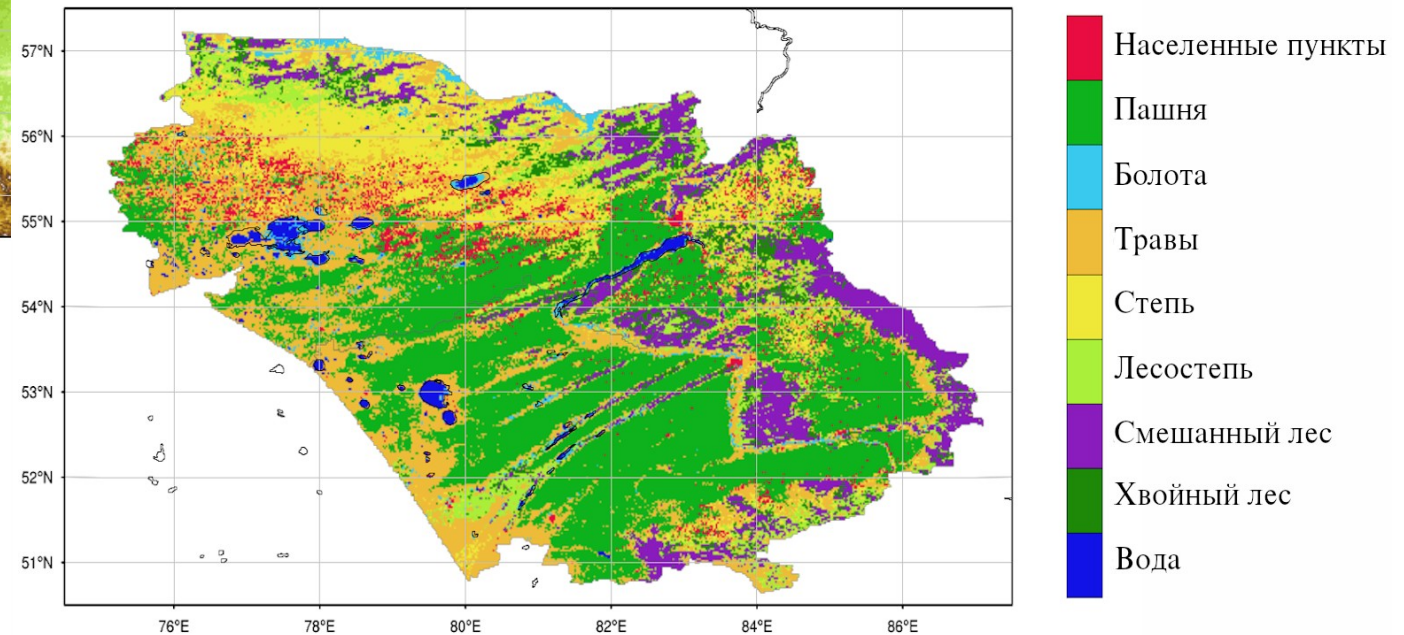
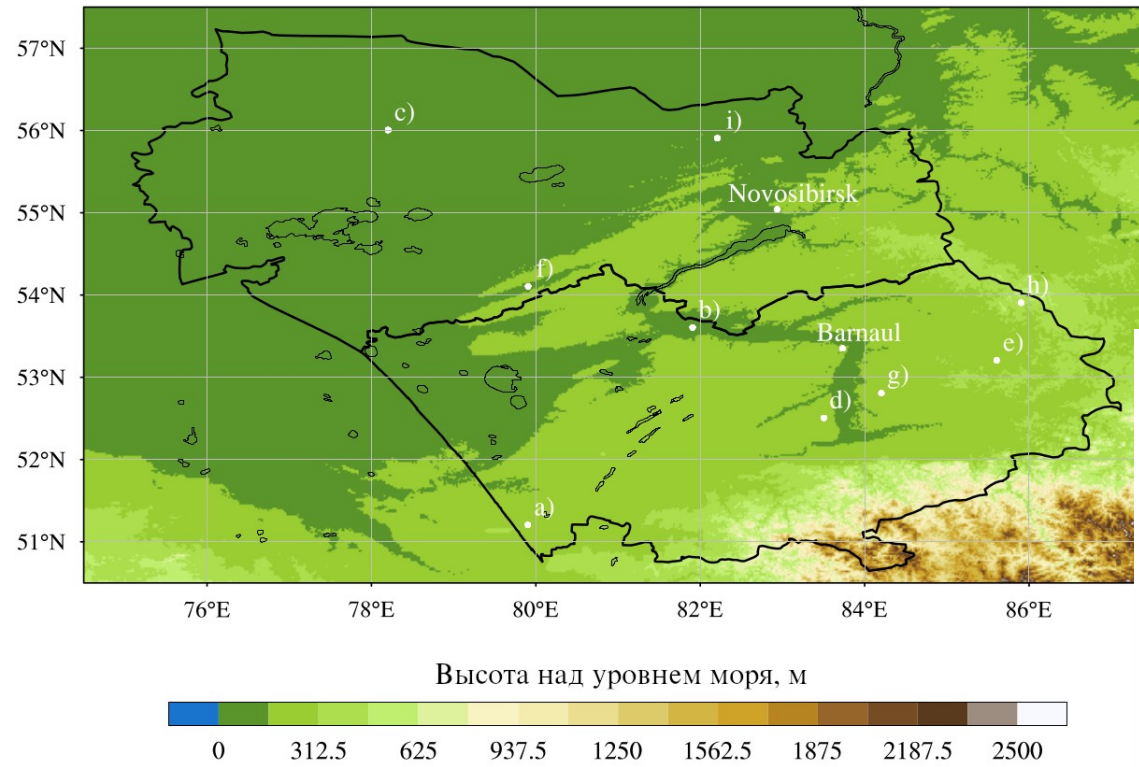
Решаемые задачи:



Data & Experiments Design



Область исследования



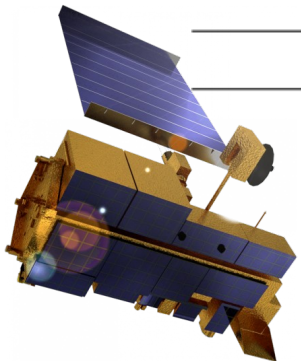
Тип растительности (продукт MCD12Q1) по классификации UMD университета Мэрилэнда.

Данные

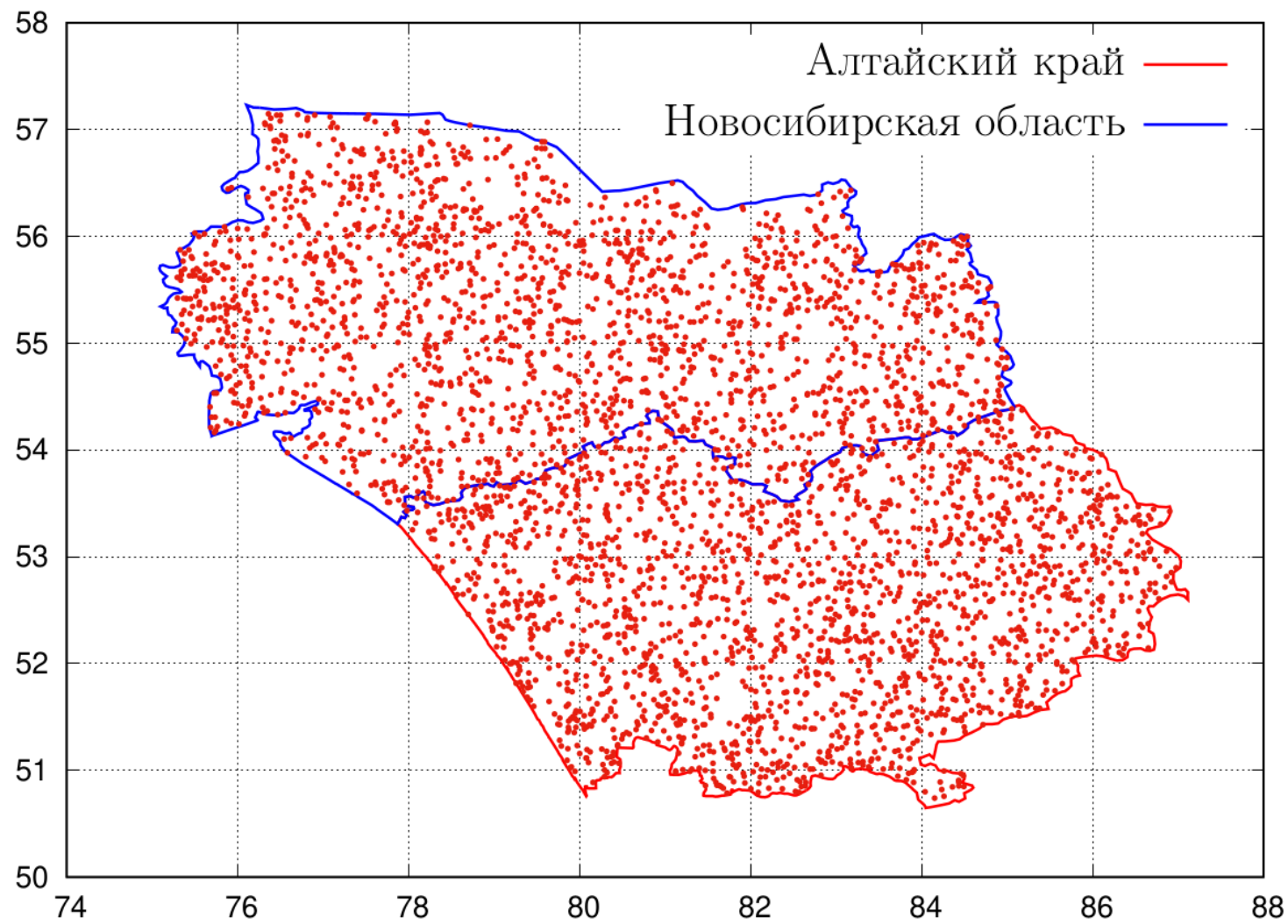
Категория	Параметр	Пространственное разрешение	Временное разрешение	Источник данных
Физиология растительности	EVI	500 м	8 дней	MODIS (MOD09A1)
	NDVI	500 м	8 дней	MODIS (MOD09A1)
	LAI	500 м	8 дней	MOD15A2H
Подстилающая поверхность	LSWI	500 м	8 дней	MODIS (MOD09A1)
	Height	90 м	—	SRTM
	LULC	500 м	1 год	MCD12Q1
	LST	1 км	8 дней	MOD11A2
	FPAR	500 м	8 дней	MOD15A2H
Параметры окружающей среды	TPR	0.25°	1 час	ERA5
	$T_{\min}$	0.25°	1 час	ERA5
	$T_{\max}$	0.25°	1 час	ERA5
	ET	500 м	8 дней	MOD16A2
	PET	500 м	8 дней	MOD16A2
	GPP	500 м	8 дней	MODIS (MOD17A2H)



era5



Выборки данных



Период: май-сентябрь 2024 г.

Объем выборки 5000 точек (80% обучающая выборка, 20% - тестовая). Из анализа кривых обучения (Learning curve).

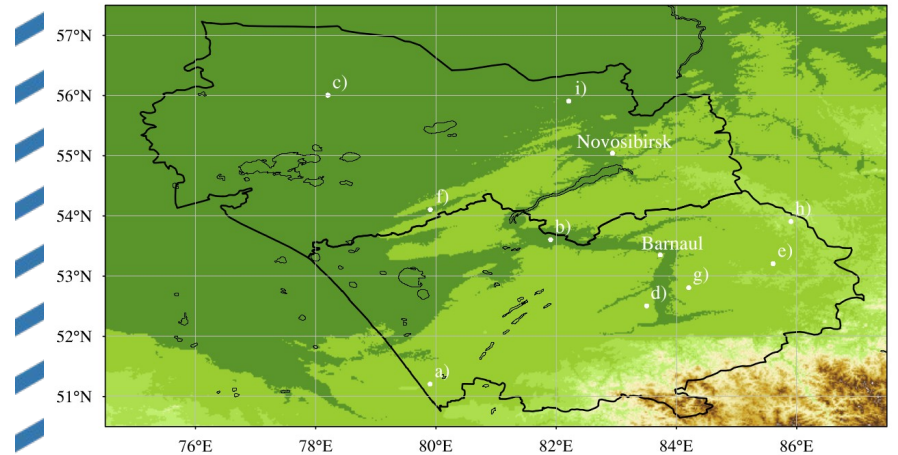
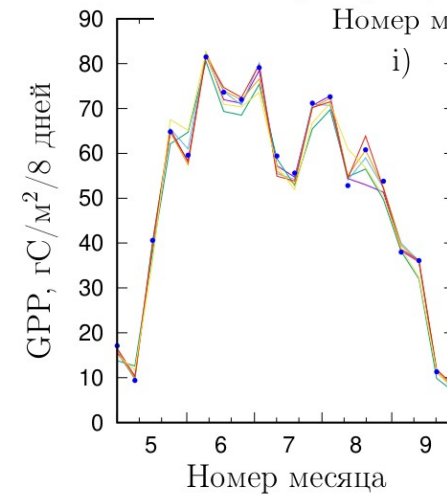
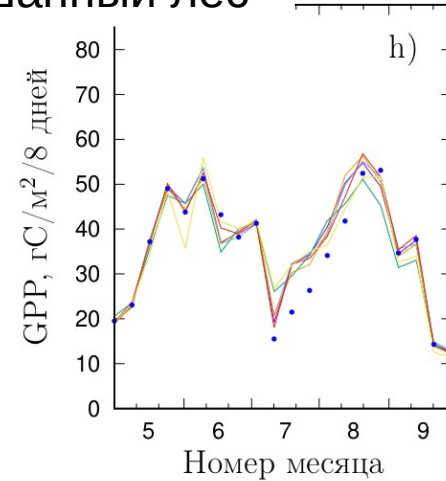
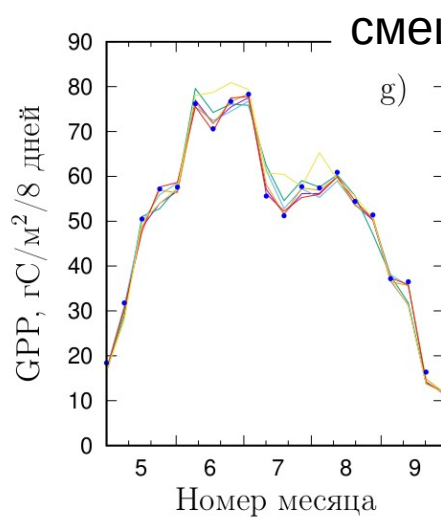
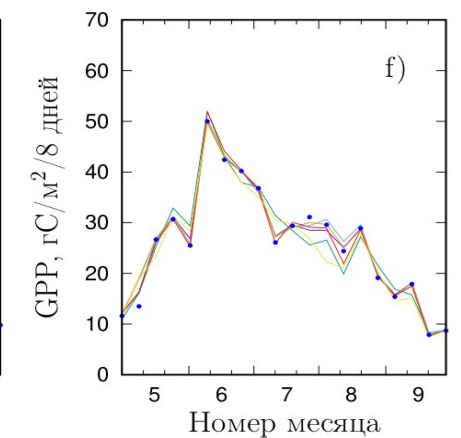
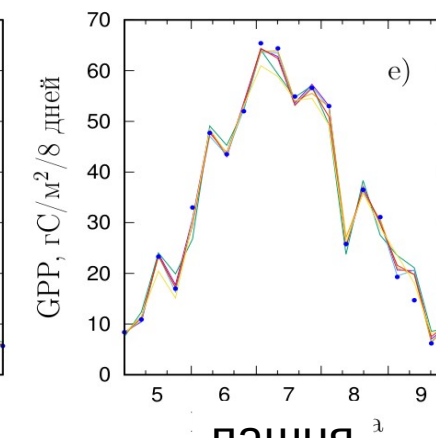
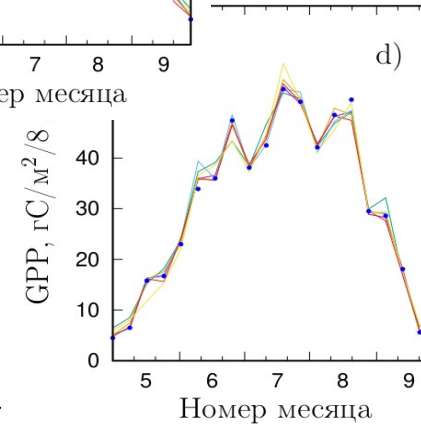
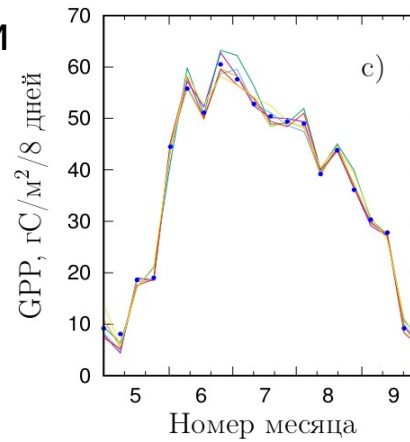
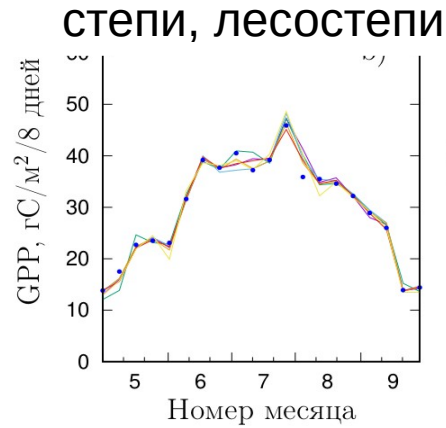
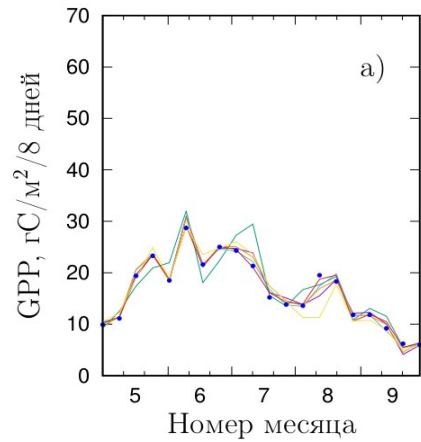
10 блоков для кроссвалидации.

Оптимальные значения гиперпараметров определялись из анализов кривых проверок (Validation curve), а также с использованием рандомизированного поиска.

Валидация моделей



Сезонный ход GPP



пашня



Метрики регрессии

Модель	MAE	RMSE	RPD	R ²	Bias
RF	2.021	3.054	5.016	0.960	0.142
ADA	3.869	4.836	3.326	0.910	-1.066
GBDT	1.804	3.096	5.106	0.962	0.179
XGB	1.741	2.636	5.984	0.972	0.061
LGBM	1.493	2.260	6.900	0.979	0.042
MLP	2.638	3.766	4.066	0.940	-0.826

$$MAE = \frac{1}{m} \sum_{i=1}^m |\hat{y}_i - y_i|$$

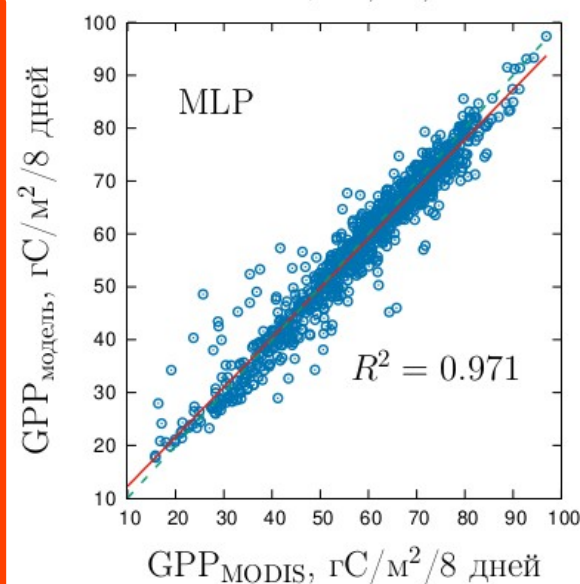
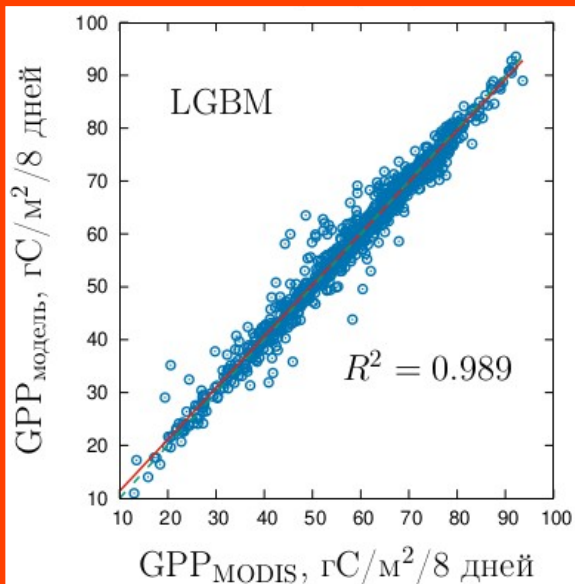
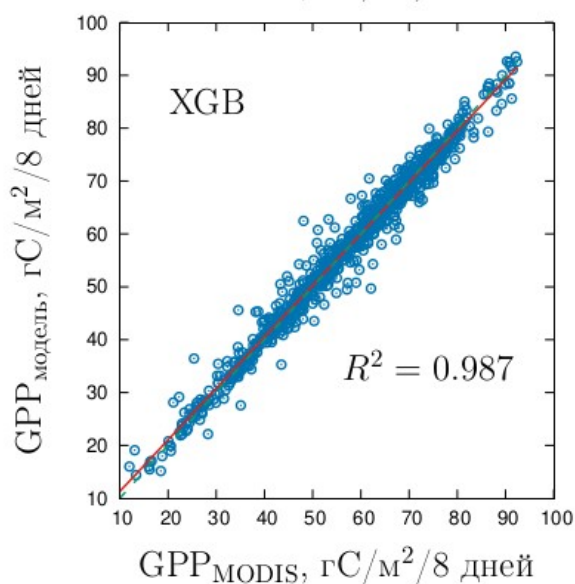
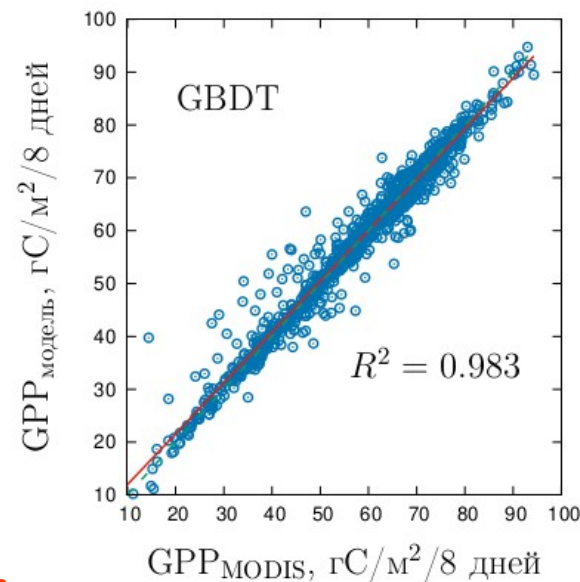
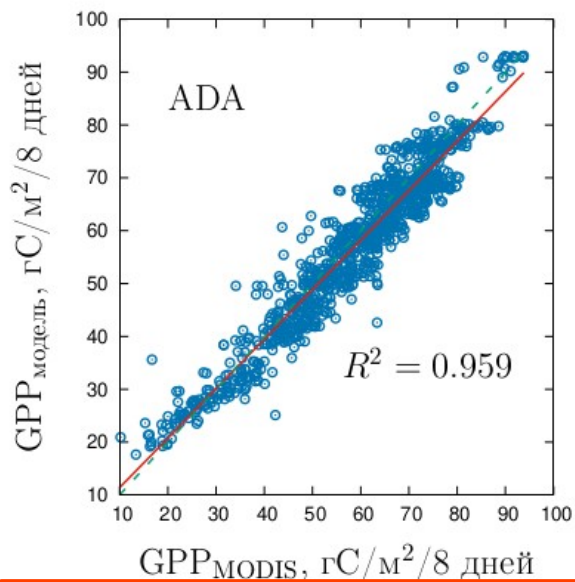
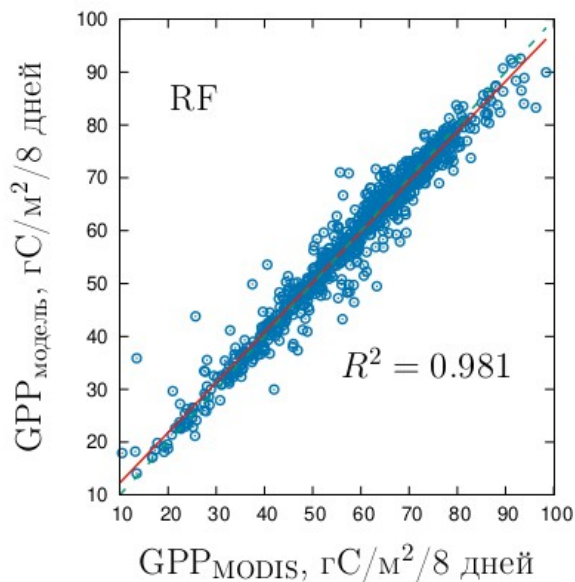
$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2}$$

$$RPD = \frac{SD}{RMSE}$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$bias = \frac{1}{m} \sum_{i=1}^m \hat{y}_i - y_i$$

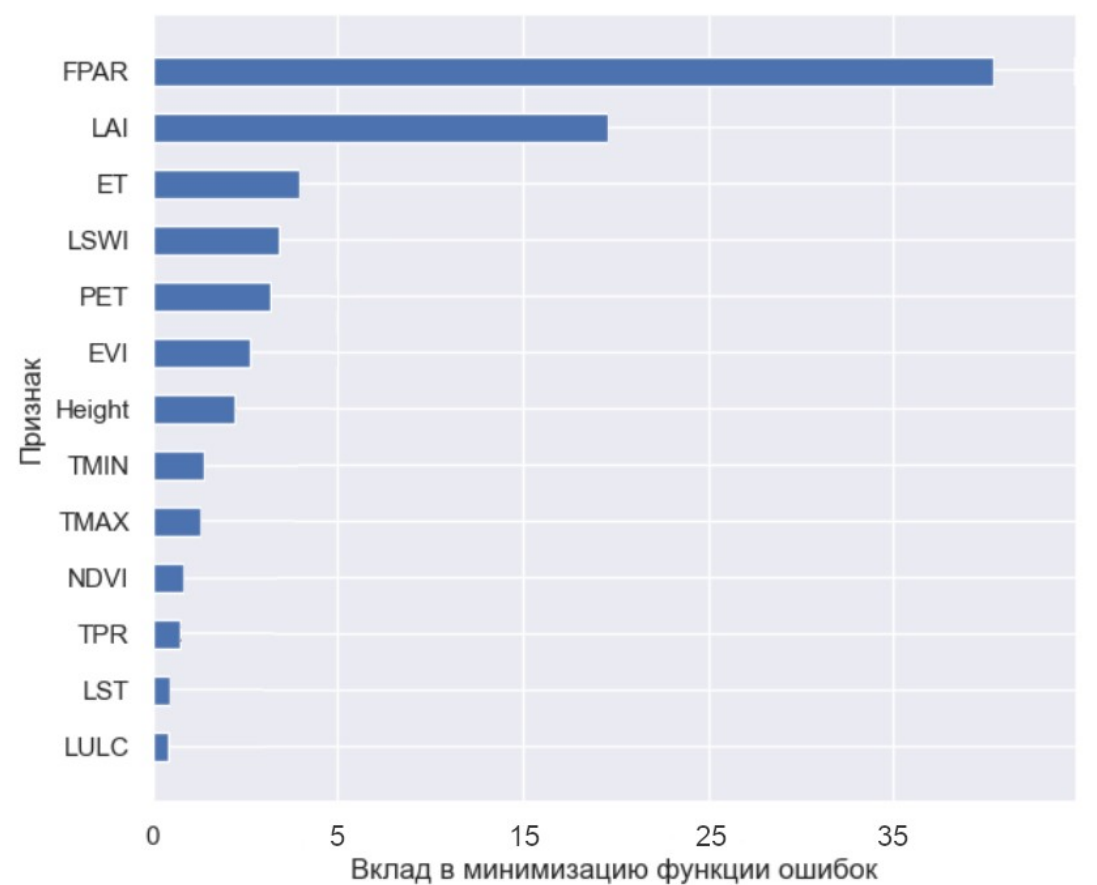
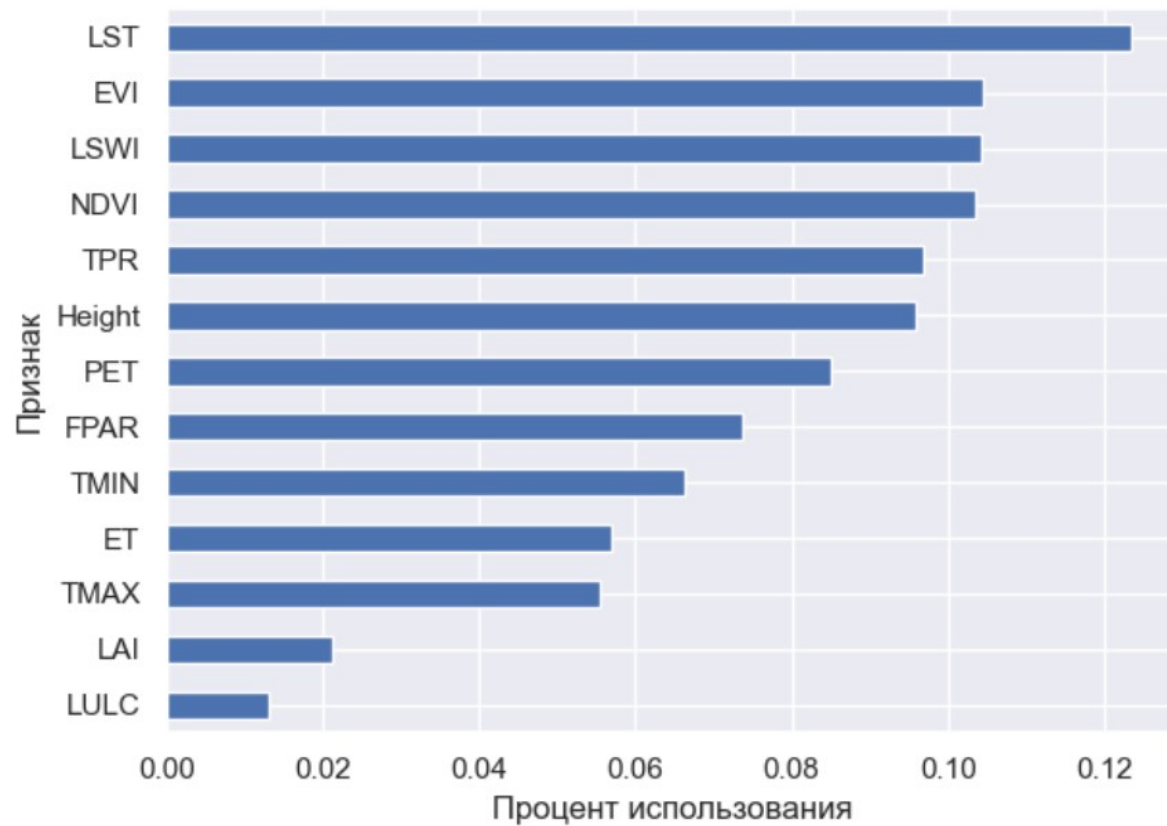
Коэффициент детерминации



Влияние признаков на GPR



Влияние признаков на GPR



Результаты и выводы

Результаты и выводы

- С использованием данных спектрорадиометра MODIS, реанализа ERA5, цифровой модели рельефа SRTM и методов машинного обучения проведены расчеты GPP для территории юга Западной Сибири. Для получения значений GPP использовались шесть методов: случайный лес, Adaboost, XGBoost, LightGBM, деревья решений с градиентным бустингом и нейросетевая модель на основе многослойного перцептрона.
- В результате анализа метрик регрессии установлено, что наилучшими показателями по качеству предсказаний значений GPP обладает модель LightGBM.
- Для модели LightGBM проведено исследование влияния признаков на величину GPP. Показано, что признаки, которые чаще всего использовались в модели для принятия решения о разделении дерева в большинстве случаев вносят меньший вклад в минимизацию функционала потерь.
- Наиболее важными признаками, вклад которых в получаемые оценки GPP составляет более 80 % являются : FPAR, LAI, ET, LSWI, PET, EVI.

Спасибо за внимание!

Questions & answers
volkov@theory.asu.ru



Работа выполнена при поддержке ФИЦ ИВТ (проект «Методы и технологии сбора, хранения и обработки пространственных данных и их временных серий для решения фундаментальных и прикладных задач исследования природных и антропогенных процессов»).