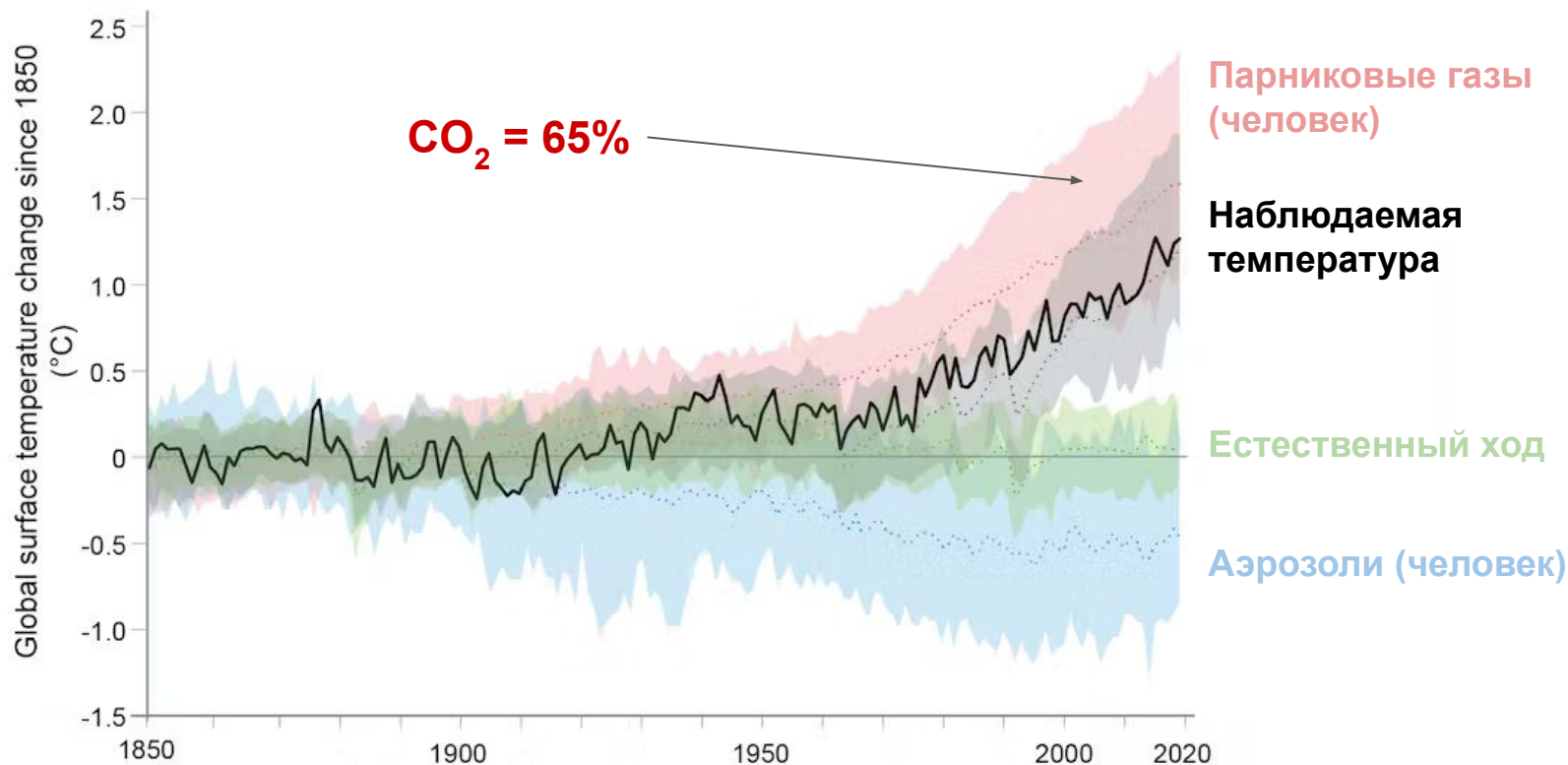


NorthFlux: ансамбль моделей машинного обучения для оценки потоков углерода в Северном полушарии по данным ДЗЗ

Розанов А.П.

Стажер-исследователь, ЛФКОС УрФУ
Магистрант, ФЦТ ИТМО

Глобальное изменение климата



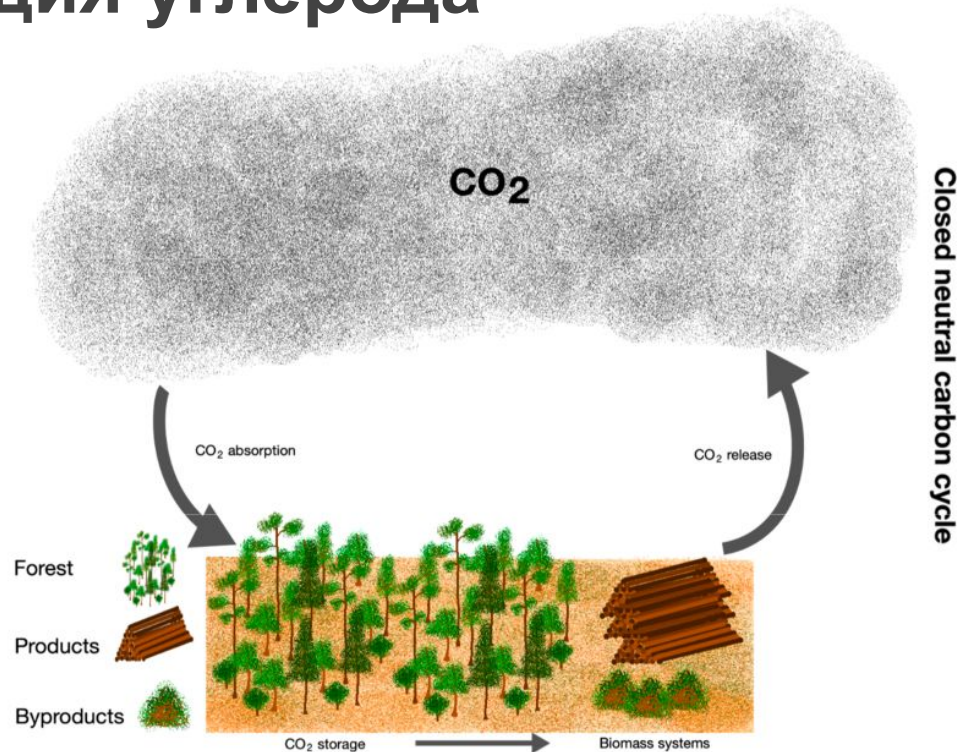
IPCC, 2022: *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegria, M. Craig, S. Langsdorf, S. Lössche, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:10.1017/9781009325844.

Секвестрация углерода



Одним из главных естественных мест **стока углекислого газа** являются наземные экосистемы.

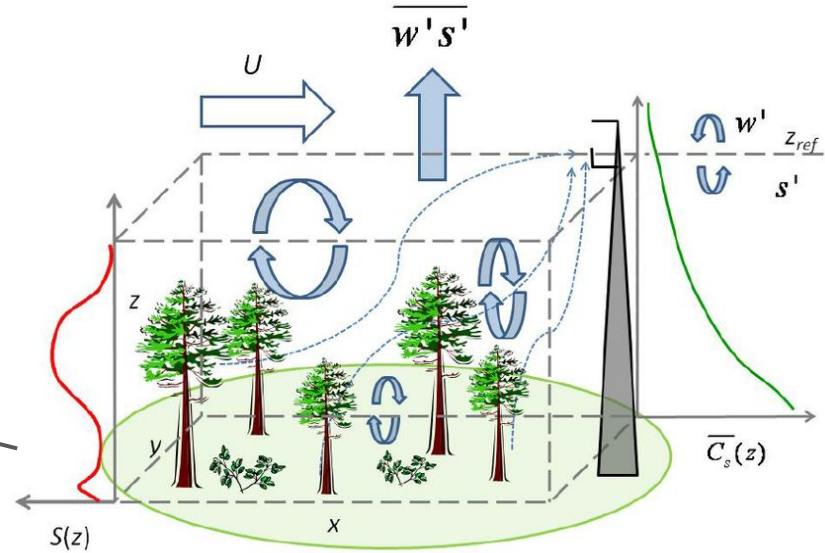
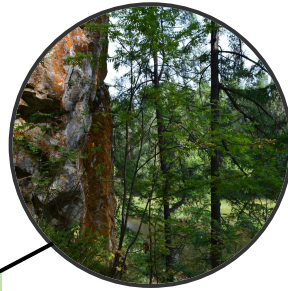
Утилизация **CO₂** с помощью секвестрации является перспективным методом борьбы с **глобальным изменением климата**.



Gorte, R.W. Carbon Sequestration in Forests; DIANE Publishing: Collingdale, PA, USA, 2009.

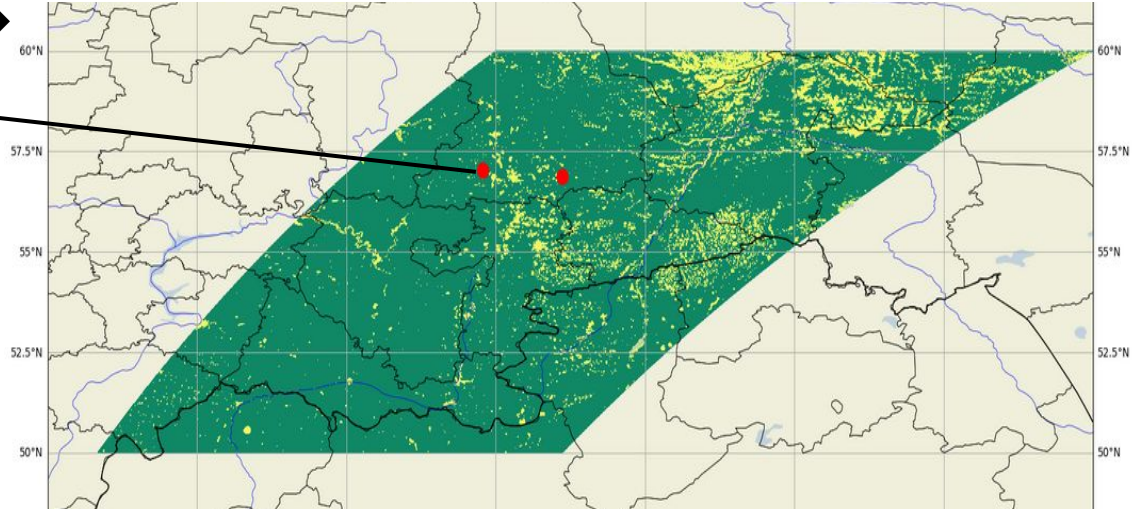
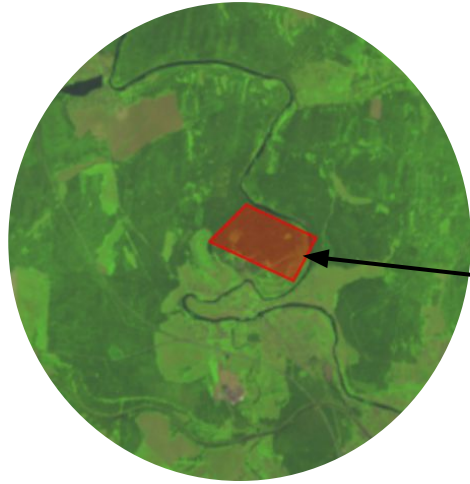
Методы измерений

Карбоновый полигон
“Урал-Карбон”



Метод турбулентных пульсаций позволяют измерить потоки CO_2 в радиусе 100-2000 м от вышки, поэтому он неприменим для **больших и труднодоступных территорий.**

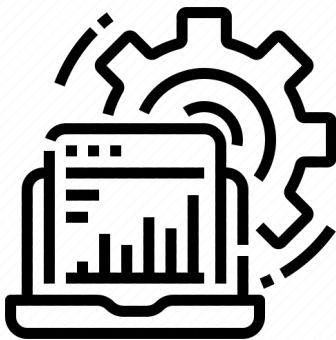
Методы измерений



Спутниковые методы позволяют получить информацию о состоянии атмосферы и подстилающей поверхности с **глобальным покрытием**. Однако с их помощью нельзя напрямую измерить потоки CO_2 , но можно создать **модель**.

Виды моделей стока углерода

Физическая
модель
Process-based



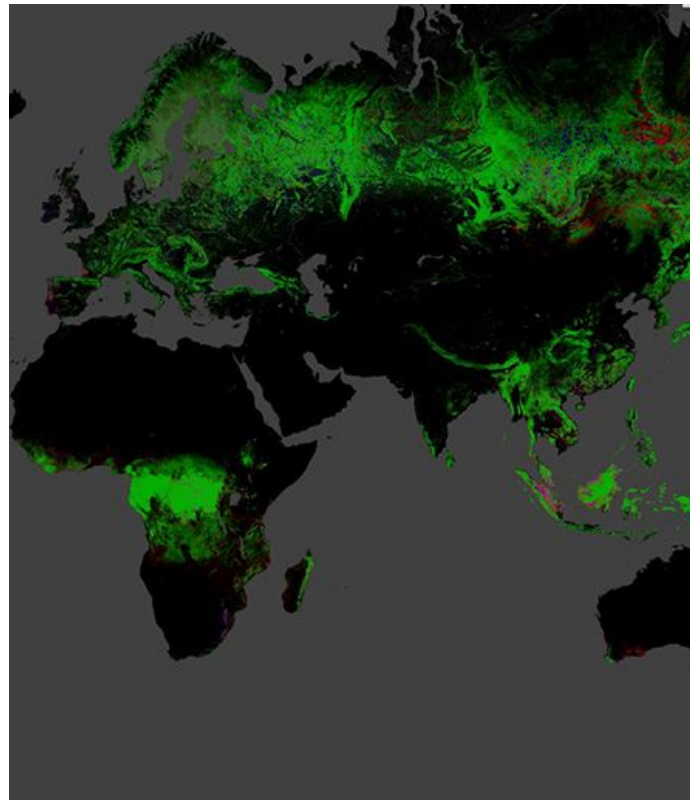
Модель на данных
Data-driven

Цель работы



Создание и обучение **моделей машинного** обучения для последующего получения оценки потоков углерода в Северном полушарии на регулярной пространственной и временной сетках.

Обученная модель должна обладать более высокой точностью оценки по сравнению с существующими аналогами.



MODIS

MOD09 CMG



Отражательная способность
поверхности в **7**
спектральных интервалах

+

**4 канала в тепловом ИК
диапазоне**



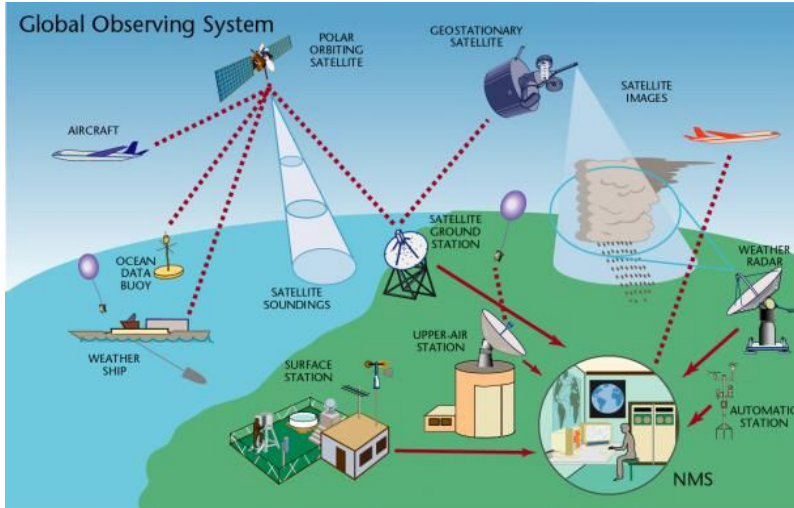
MCD12C1



IGBP (International
Geosphere-Biosphere
Programme) -
классификация
экосистем.

ERA 5

Климатические данные представлены:



<https://public.wmo.int/en/programmes/global-observing-system>

1. *Компонентами ветра;*
2. *Температурой воздуха на 2 м ;*
3. *Точкой росы на 2 м;*
4. *Температурой пов-ти;*
5. *Уровнем осадков;*
6. *Уровнем испарения;*
7. *Давлением на пов-ти;*
8. *Балансом солнечной энергии на поверхности;*
9. *Покрытиями верхнего и нижнего ярусов растительности.*

FLUXNET



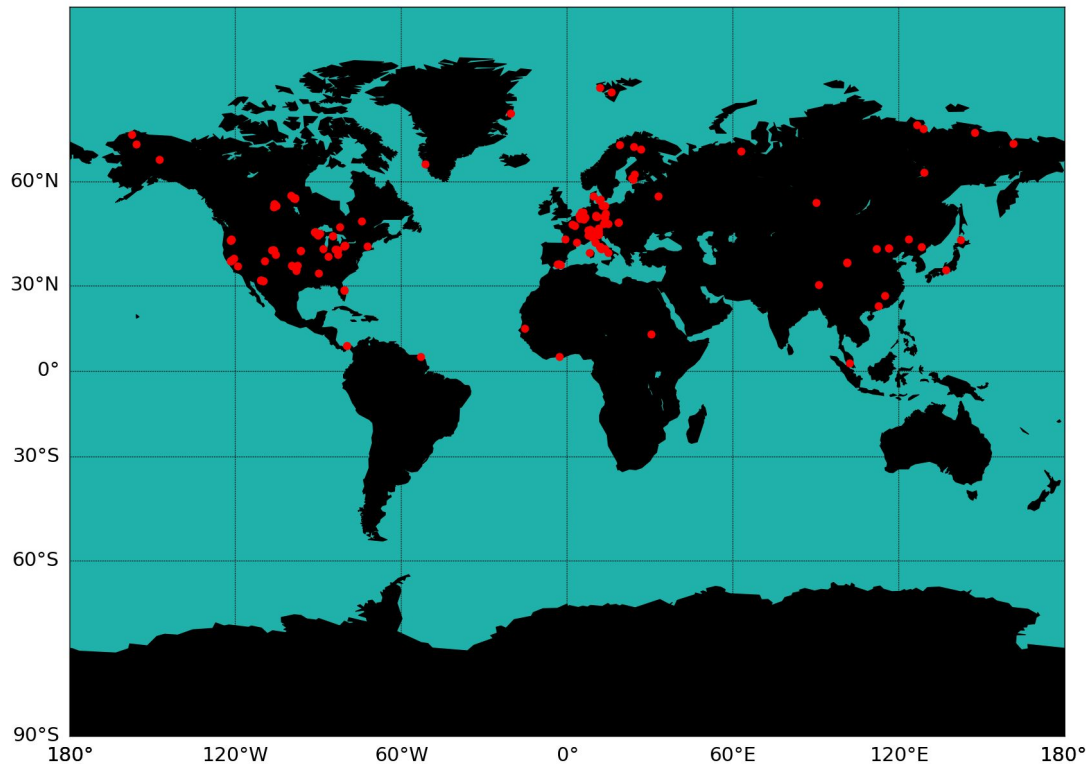
$$NEE = GPP - R_a - R_h$$

NEE (баланс CO₂)

GPP (первичный
валовый продукт)

R_a (автотрофное
экосистемное дыхание)

R_h (гетеротрофное
экосистемное дыхание)

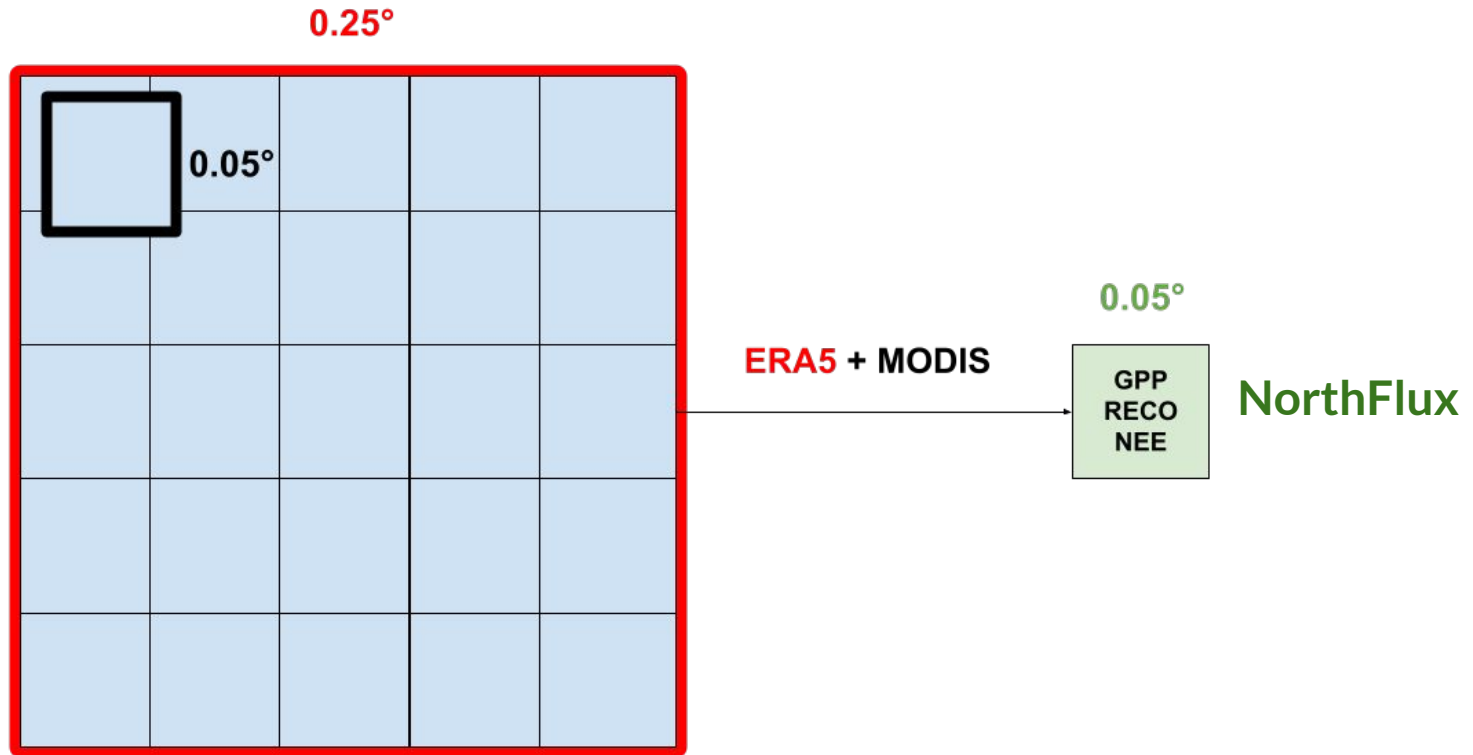


Отобранные станции **FLUXNET**

Структура задачи



Разрешение итогового продукта



Моделирование

ANN

RF

AdaBoost

XGBoost

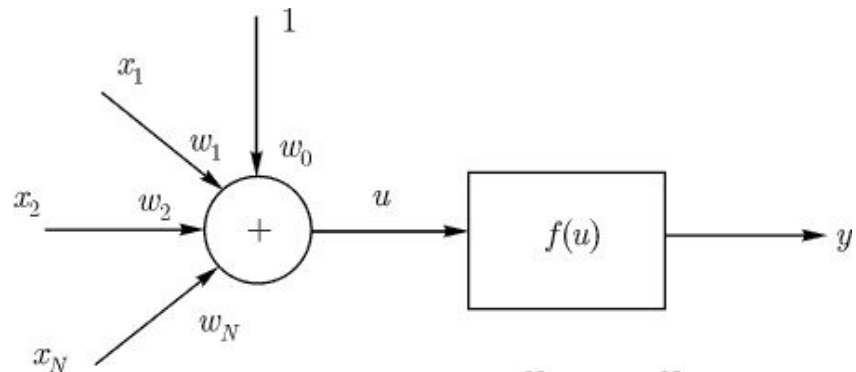
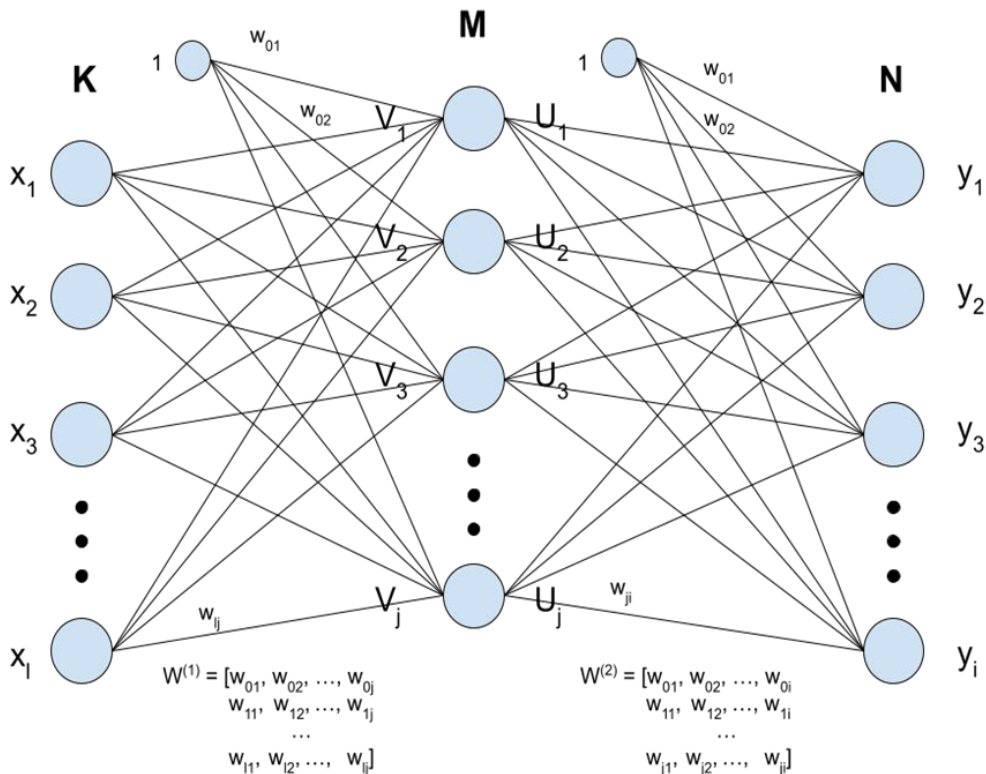
(искусственная
нейронная
сеть)

(Случайный
лес)

(Адаптивный
бустинг)

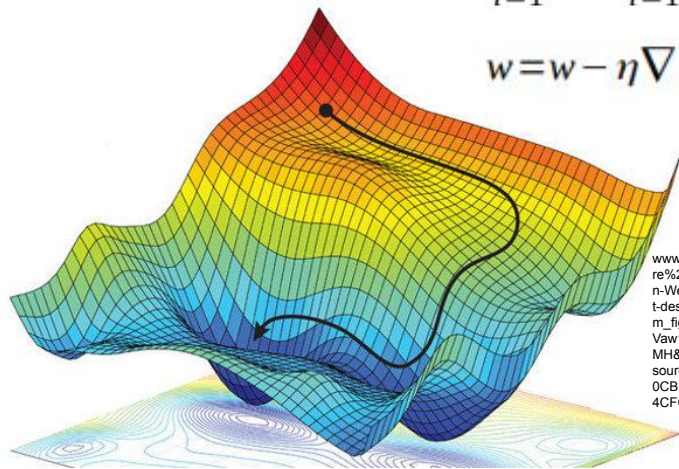
(Экстремальный
градиентный
бустинг)

Искусственная Нейронная Сеть (ANN)



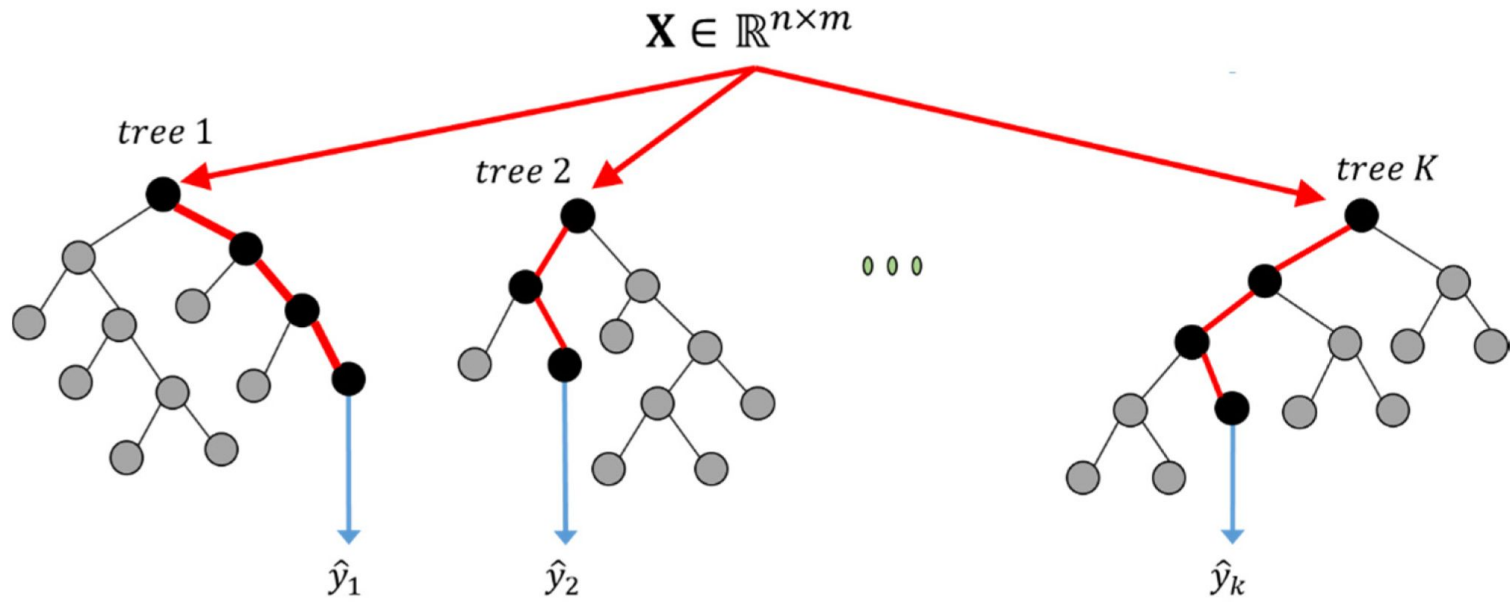
$$E(w) = \frac{1}{2} \sum_{i=1}^N E_i = \frac{1}{2} \sum_{i=1}^N (y_i - d_i)^2$$

$$w = w - \eta \nabla E(w)$$



www.researchgate.net/figure/figure.png?figid=1325142728&psig=AOvVaw1J7zPv53b8q6Wv35WwheMH&ust=1683102397076000&source=images&cd=vfe&ved=0CBMQjhxqFwoTCMl9pWb1v4CFQAAAAAdAAAAABAK

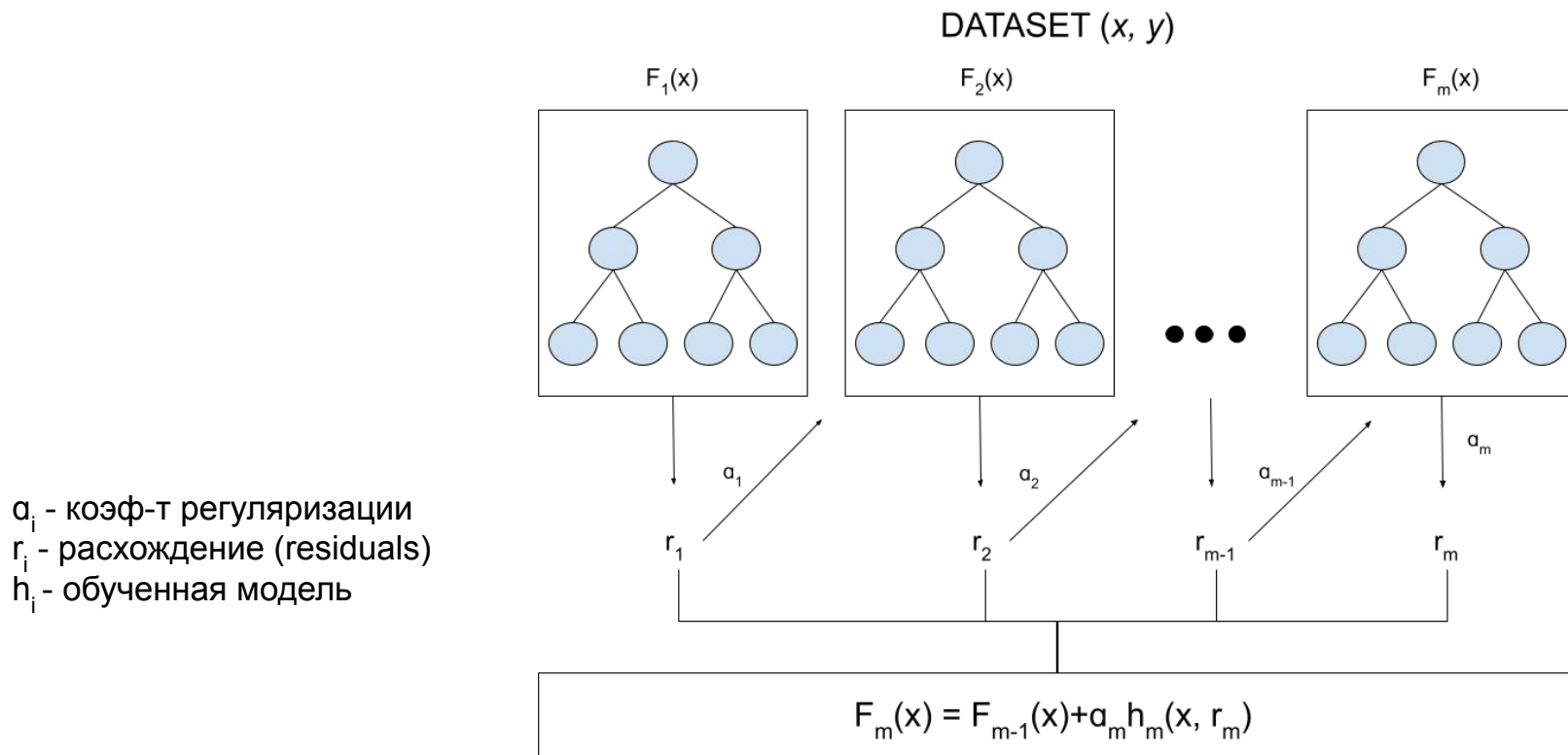
Случайный лес (RF)



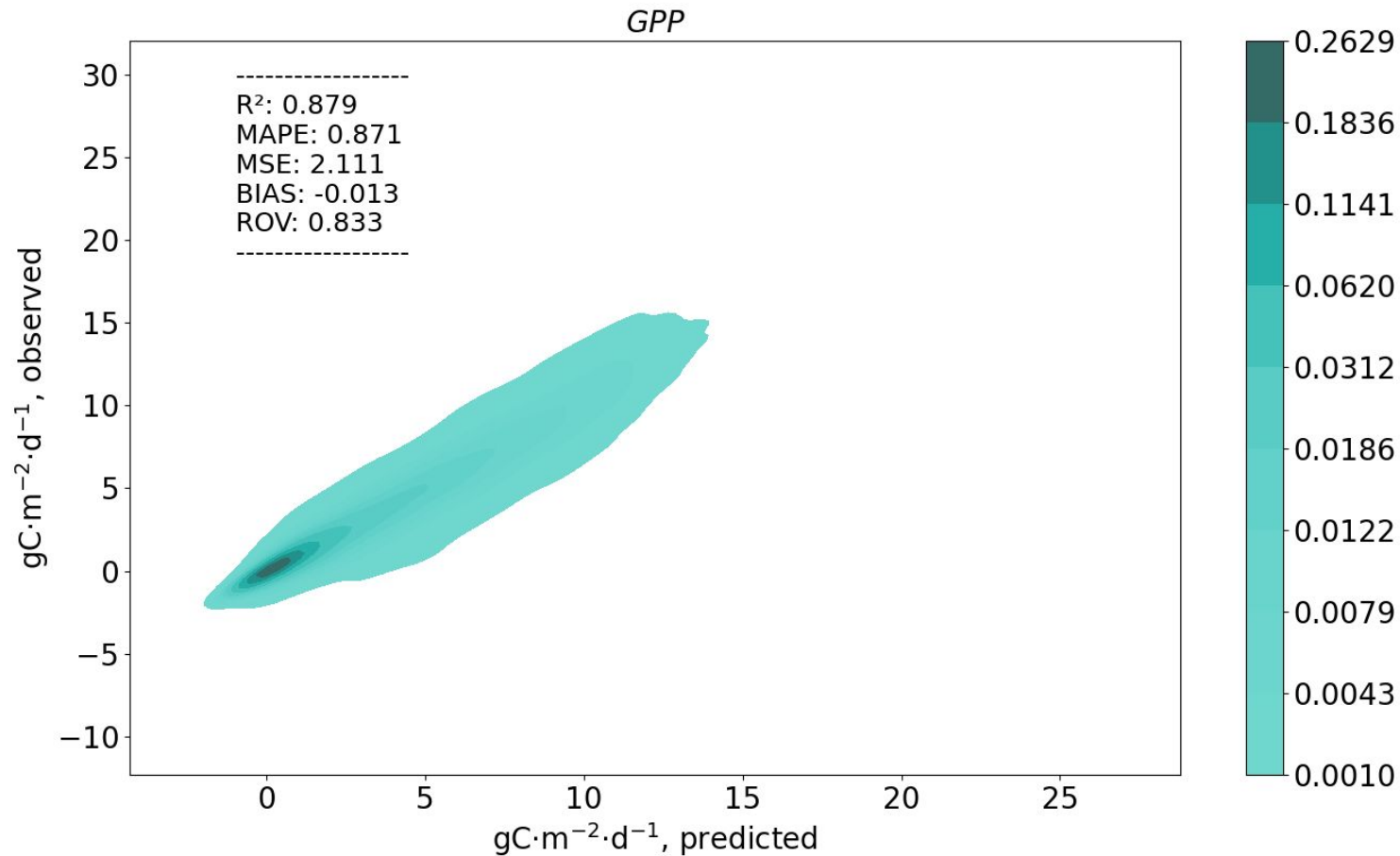
Aldrich, C. Process Variable Importance Analysis by Use of Random Forests in a Shapley Regression Framework. *Minerals* **2020**, *10*, 420. <https://doi.org/10.3390/min10050420>

$$\hat{y} = \frac{1}{K} \sum_{k=1}^K \hat{y}_k$$

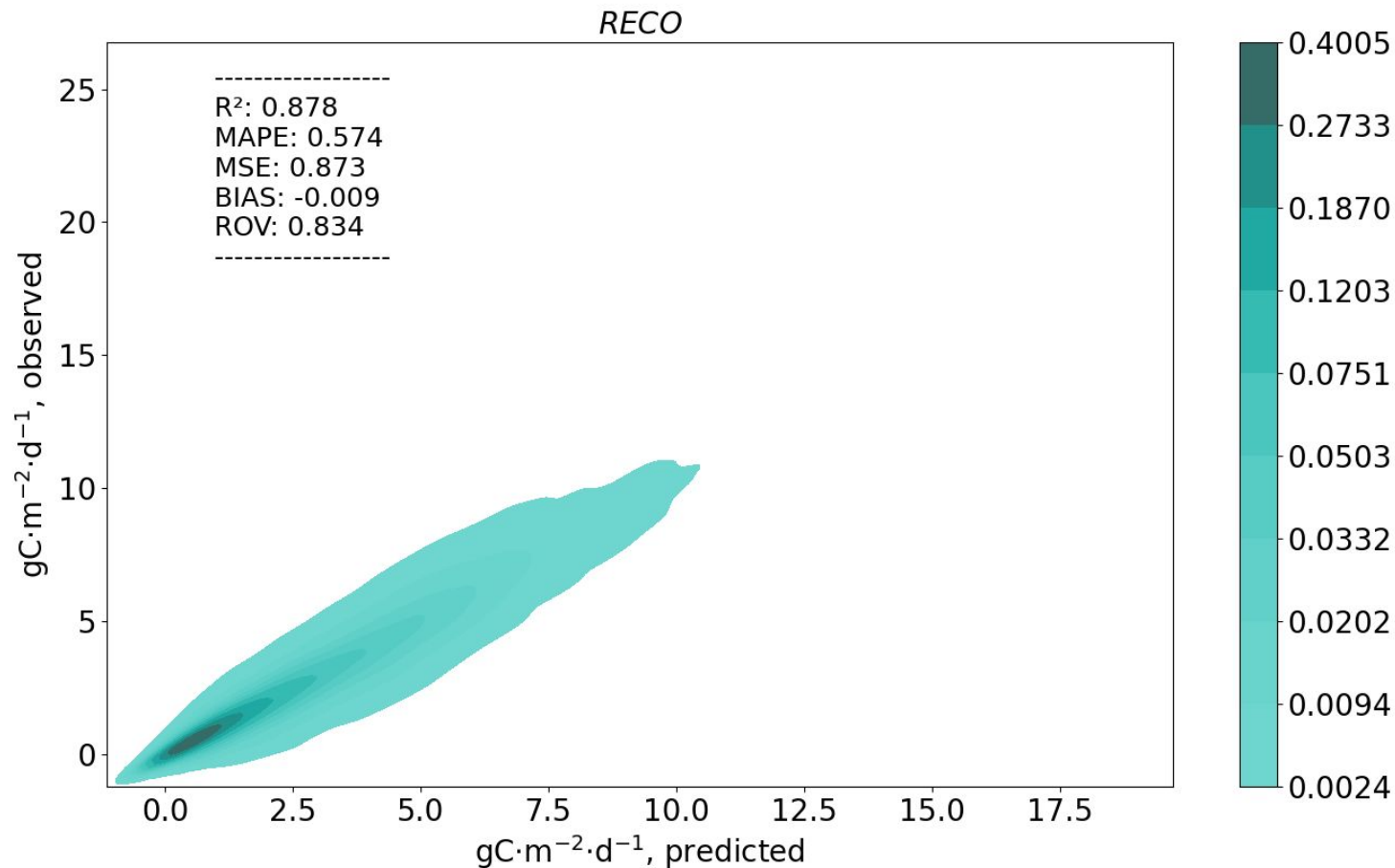
Бустинг (XGBoost и AdaBoost)



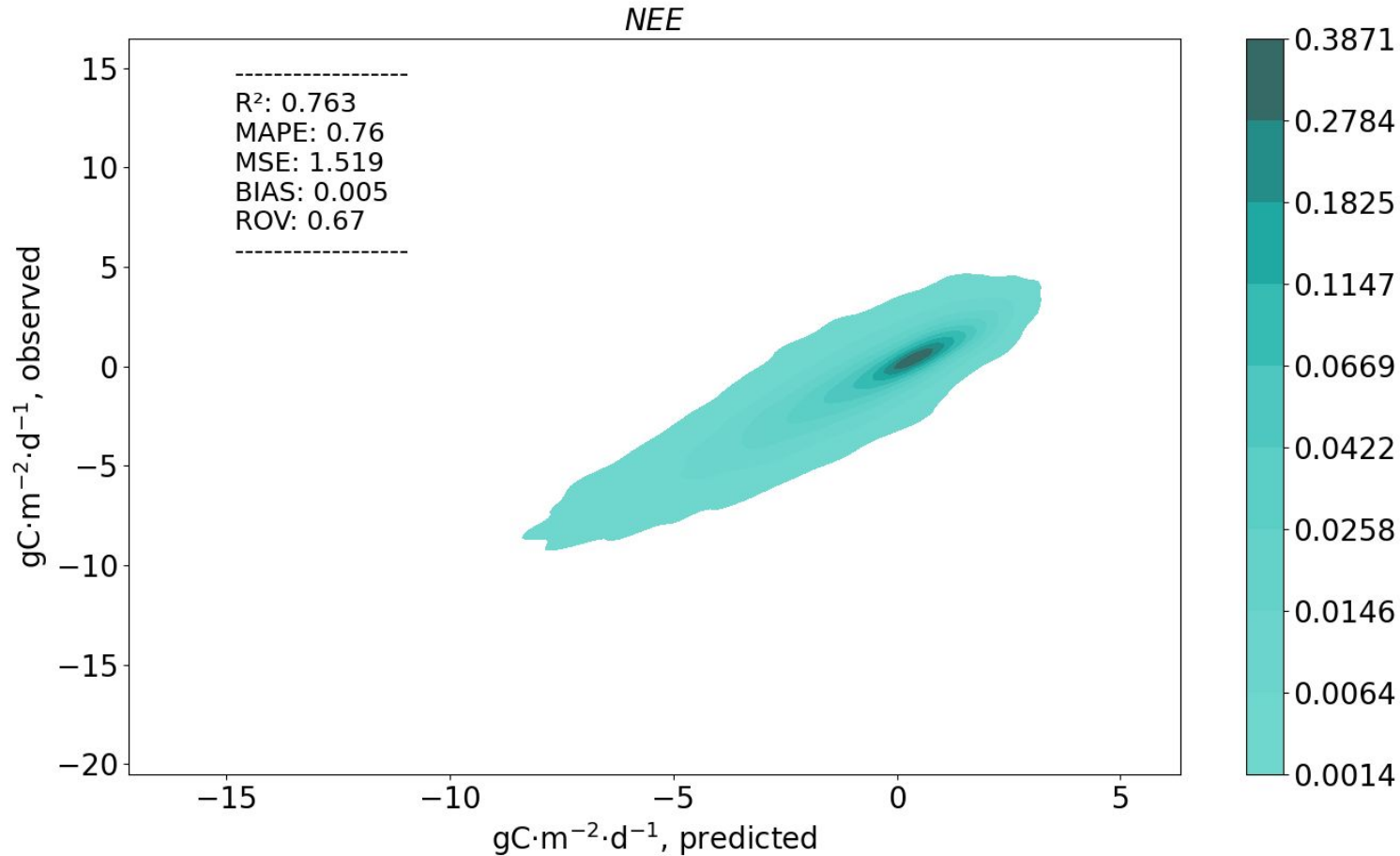
Ансамблевая оценка



Ансамблевая оценка



Ансамблевая оценка



Результаты моделирования на тестовом наборе данных с учетом типа экосистемы

		GPP	RECO	NEE
Forests	• ENF	ANN	Ensemble	ANN
	• EBF	Ensemble	AdaBoost	Ensemble
	• DBF	Ensemble	Ensemble	ANN
	• MF	ANN	Ensemble	ANN
Aquatic	• Water Bodies	Ensemble	AdaBoost	Ensemble
	• Wetlands	Ensemble	Ensemble	Ensemble
Un/sparsely-vegetated	• Barren	RF	RF	RF
	• CNVM	Ensemble	Ensemble	Ensemble
	• Urban	Ensemble	AdaBoost	RF
Others	• WoodSav	Ensemble	Ensemble	Ensemble
	• Sav	Ensemble	Ensemble	Ensemble
	• Grass	Ensemble	AdaBoost	Ensemble
	• Crop	Ensemble	Ensemble	Ensemble
	• OpenShr	Ensemble	Ensemble	RF

Валидация

FLUXCOM (2020)
1d, 0.0833°

QC > 0

Model	GPP _{MSE}	RECO _{MSE}	NEE _{MSE}
NorthFlux	8.89	2.36	4.47
FLUXCOM	13.06	3.77	5.49
	46.93%	59.43%	22.63%

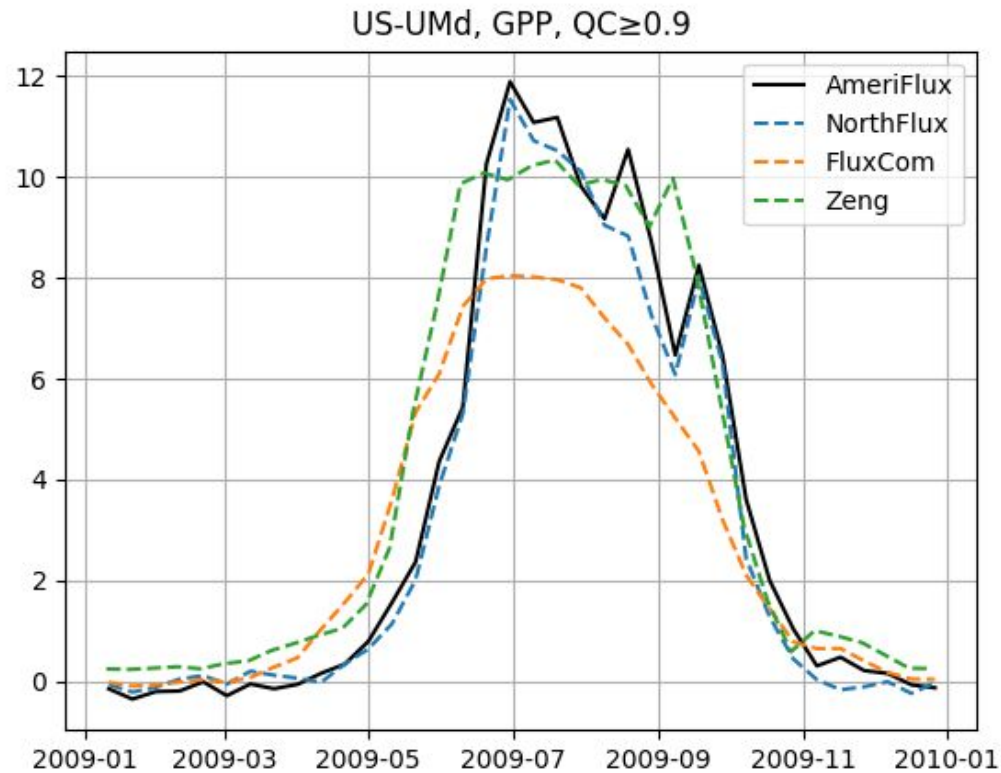
QC ≥ 0.9

Model	GPP _{MSE}	RECO _{MSE}	NEE _{MSE}
NorthFlux	11.11	3.06	4.03
FLUXCOM	14.44	4.45	5.3
	29.97%	45.51%	31.28%

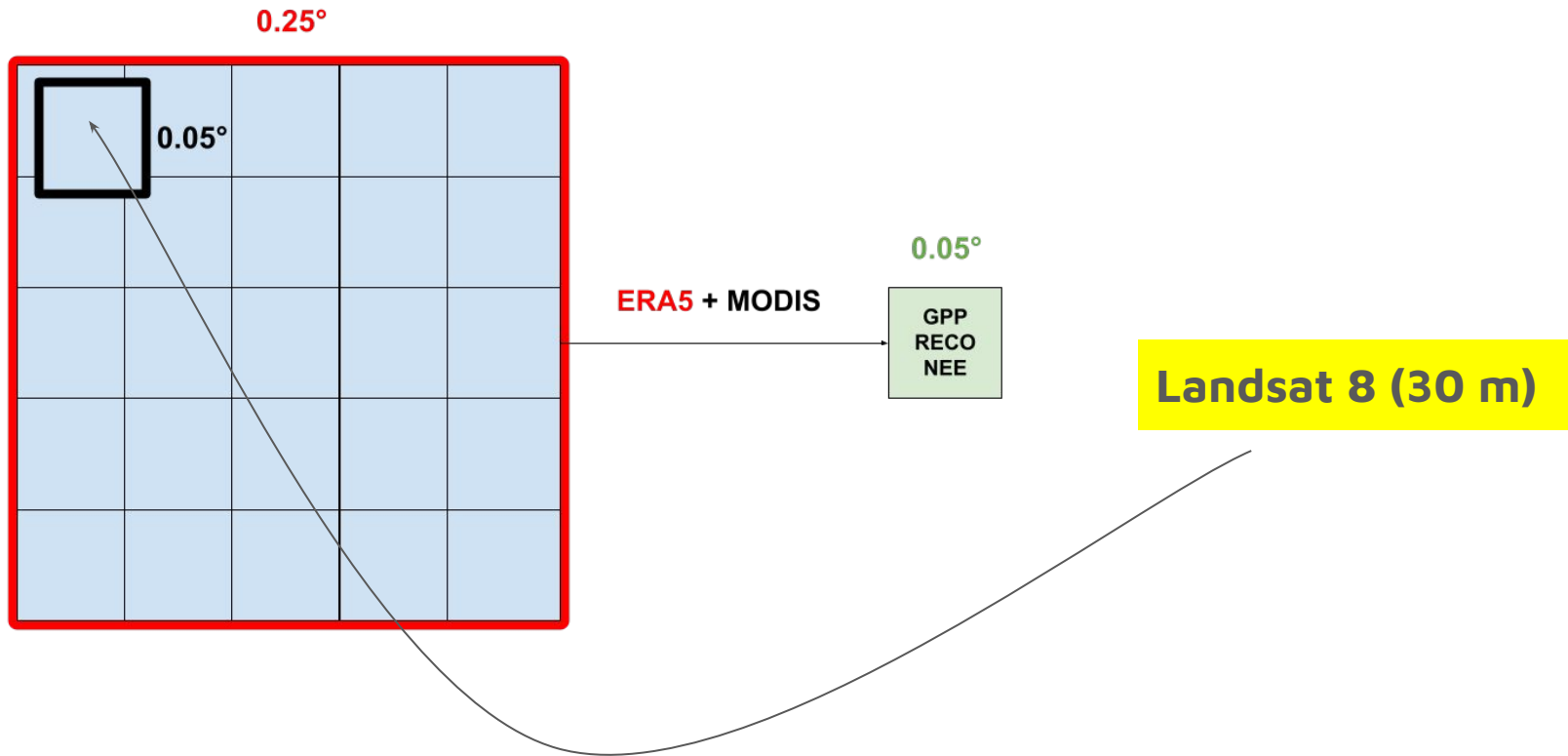
Zeng (2020)
10d, 0.1°

Model	GPP _{MSE}	RECO _{MSE}	NEE _{MSE}
NorthFlux	7.22	2.38	2.87
Zeng	9.29	3.62	3.24
	28.77%	52.45%	12.83%

Валидация

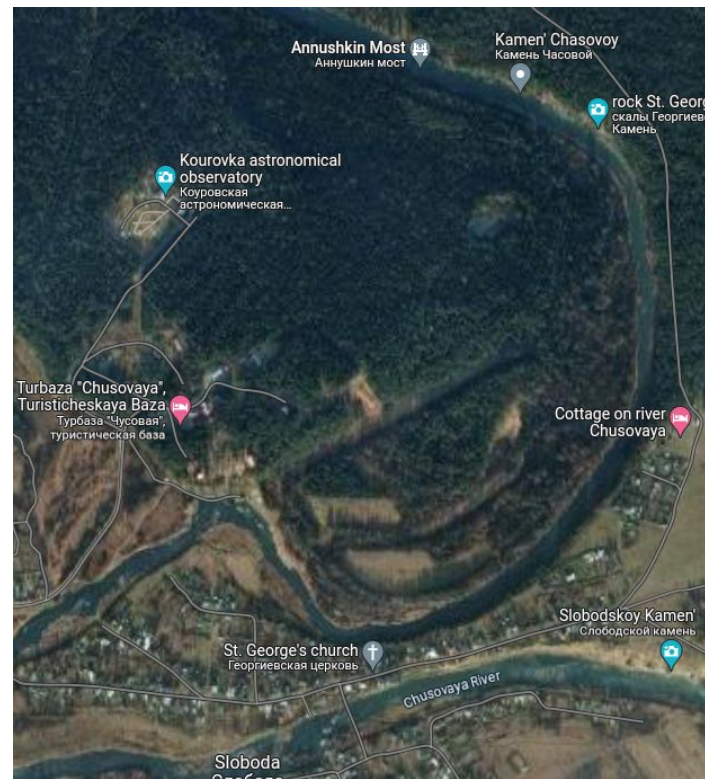
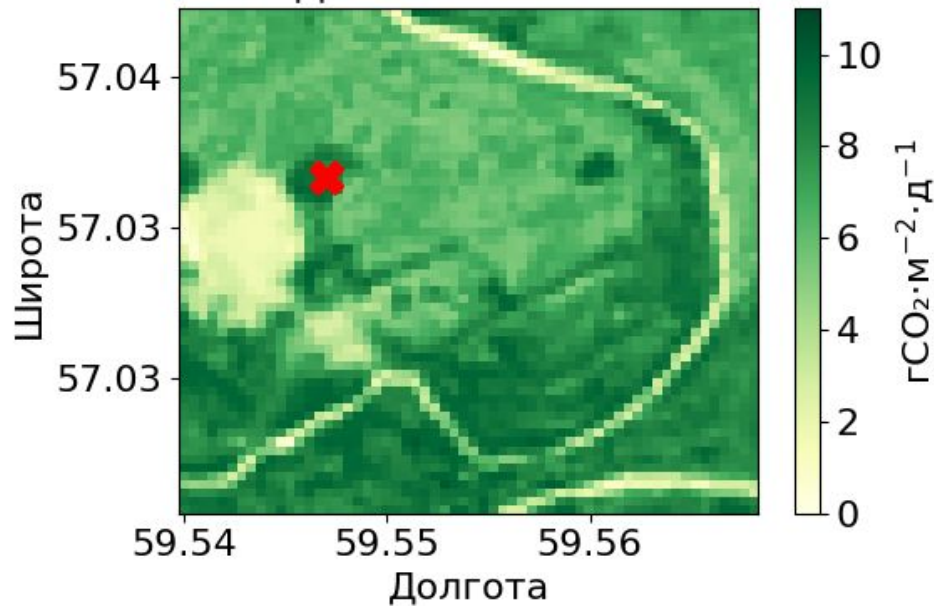


NorthFlux+Landsat 8



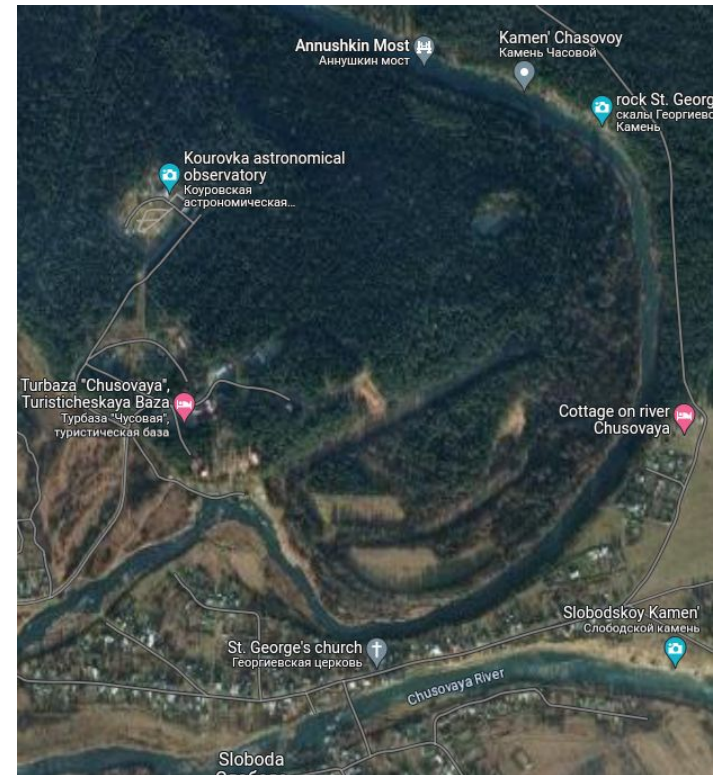
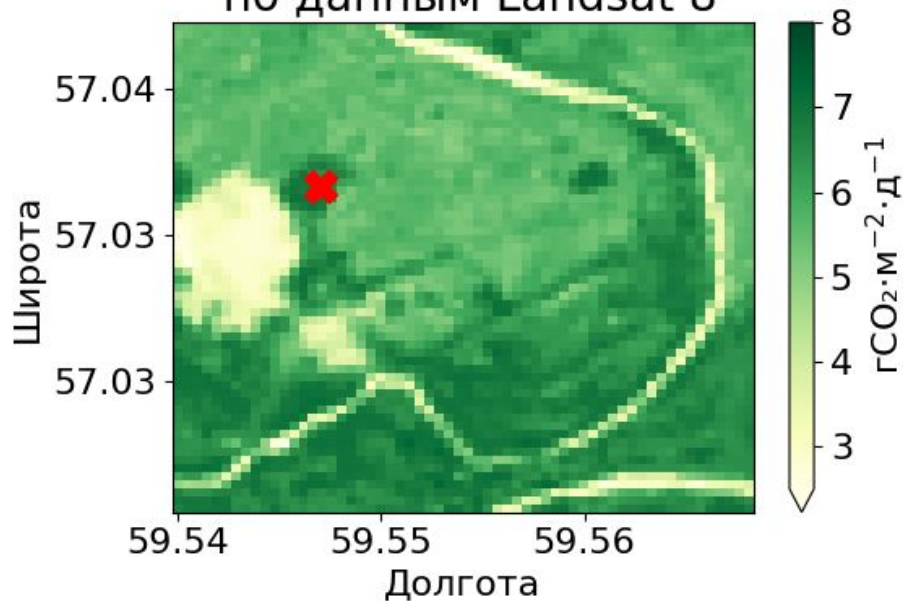
NorthFlux+Landsat 8

Коуровка
Первичный валовый продукт
по данным Landsat 8

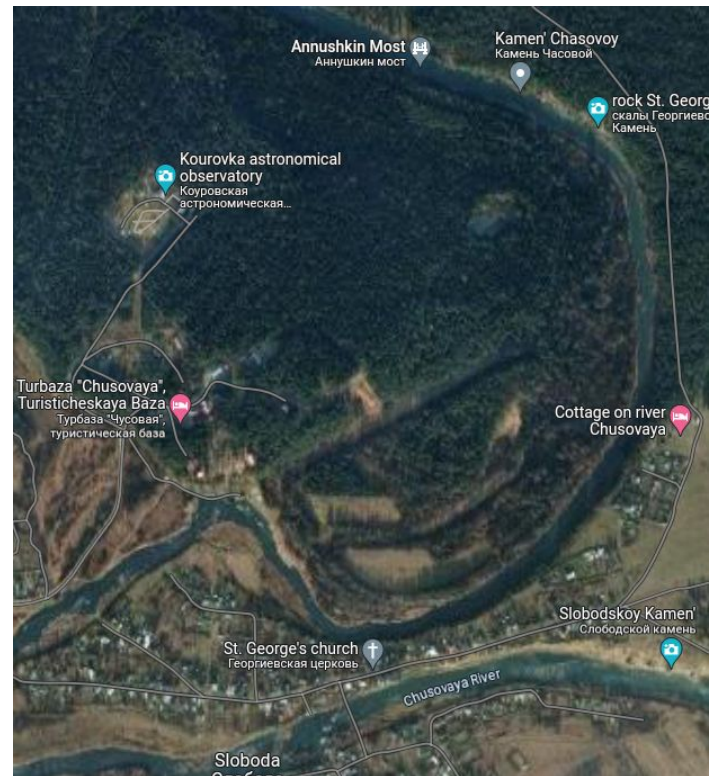
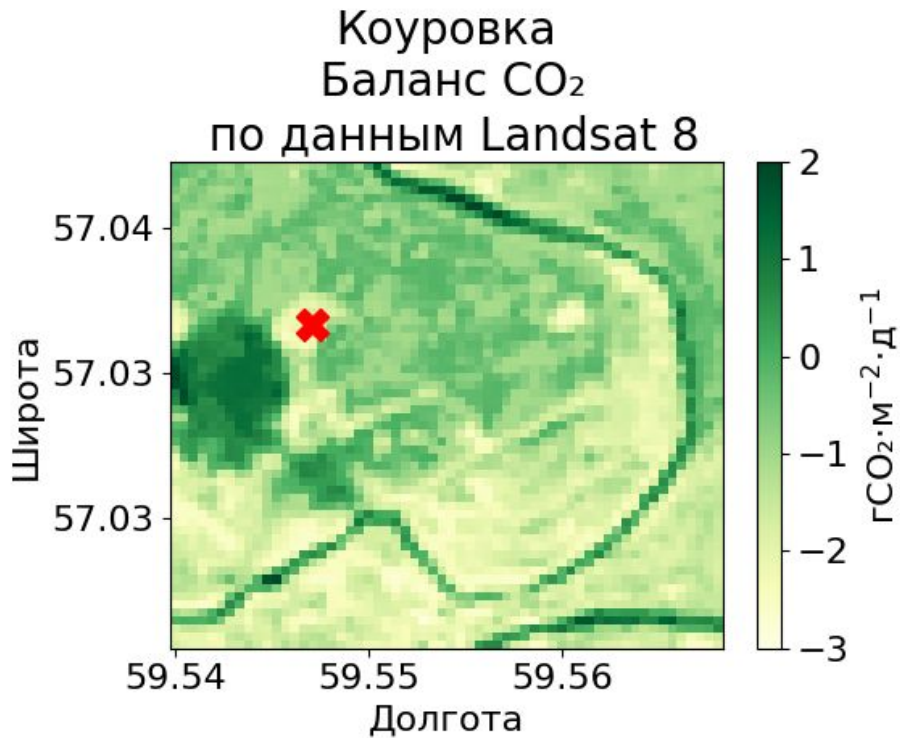


NorthFlux+Landsat 8

Коуровка
Экосистемное дыхание
по данным Landsat 8



NorthFlux+Landsat 8



Заключение

- Обученные модели демонстрируют высокую точность оценки потоков углерода на тестовом наборе данных, однако ансамблевое предсказание имеет ошибку ниже, чем предсказания отдельных моделей;
- Созданный ансамбль моделей на валидационном наборе данных обладает более высокой точностью, чем существующие аналоги, что свидетельствует о возможности его использования в различных отраслях (включая климатическое моделирование);
- Предварительные результаты исследования также свидетельствуют о возможности получения оценок потоков CO_2 высокого разрешения с помощью моделей обученных на данных с низким разрешением.
- Следующим этапом работы является завершение генерации данных и сопоставление оценок разного разрешения с реальными данными.



Спасибо за внимание!

Контакты

