

Всероссийская конференция с международным участием
«Обработка пространственных данных в задачах мониторинга
природных и антропогенных процессов» (SDM-2023)
22-25 августа 2023 г., Бердск, Россия



ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ МОДЕЛИРОВАНИЯ ПЕРВИЧНОЙ ВАЛОВОЙ ПРОДУКЦИИ ДЛЯ ТЕРРИТОРИИ АЛТАЙСКОГО КРАЯ И НОВОСИБИРСКОЙ ОБЛАСТИ



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

ФГБОУ ВО «Алтайский государственный университет»
ФИЦ ИВТ



ИВТ

WMO GREENHOUSE GAS BULLETIN

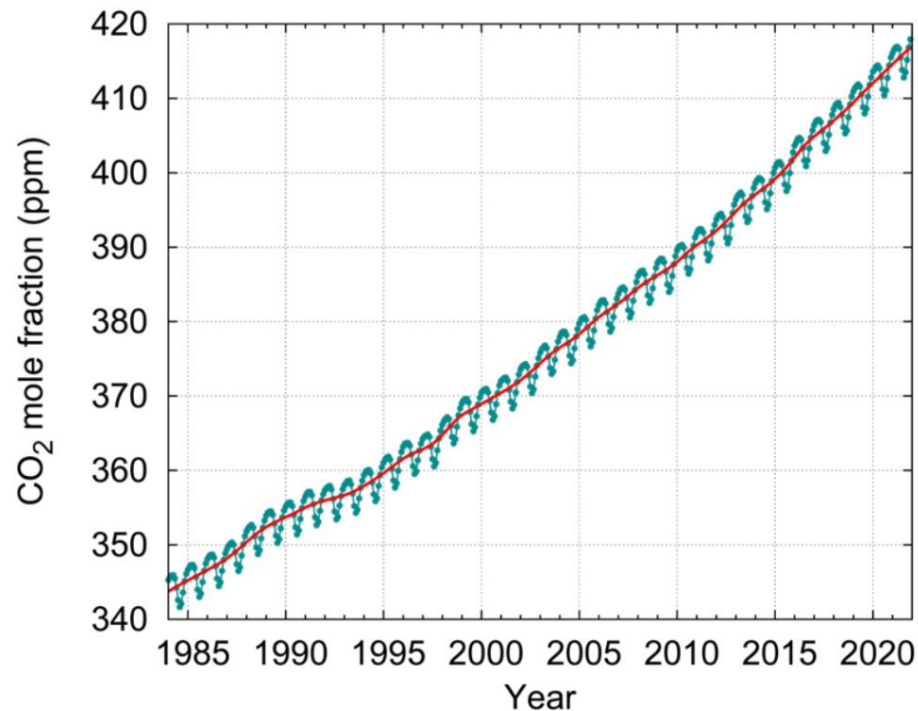
The State of Greenhouse Gases in the Atmosphere Based on
Global Observations through 2021

No. 18 | 26 October 2022

С начала индустриальной эры в 1750 г. содержание CO_2 в приземном слое атмосферы непрерывно увеличивалось с ~ 278 млн $^{-1}$.

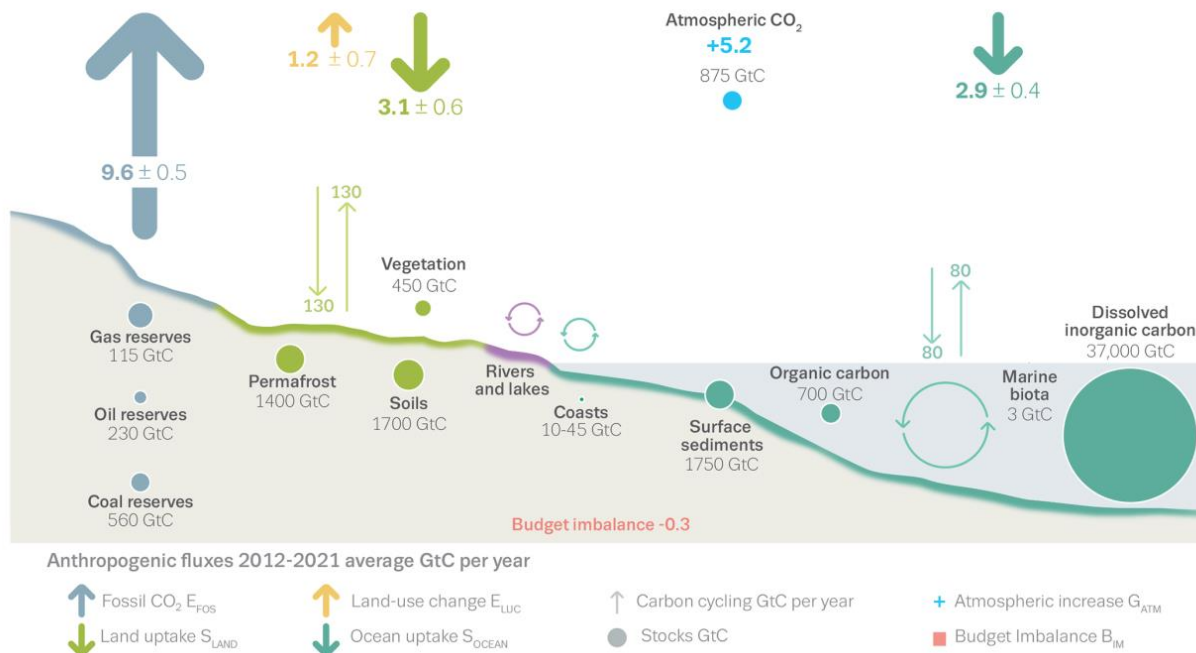


В 2021 г. содержание CO_2 достигло новых исторически максимальных значений $\sim 415.7 \pm 0.2$ млн $^{-1}$.



Скорость роста
эмиссии CO₂ в
постковидный период
вернулась к значениям
до 2019 г. и в 2021 г.
составила
~36.3±1.8 ГтCO₂/год.

The global carbon cycle



Рост эмиссии CO₂ обуславливает дисбаланс глобального карбонового бюджета, составляющий -0.3 ГтC/год



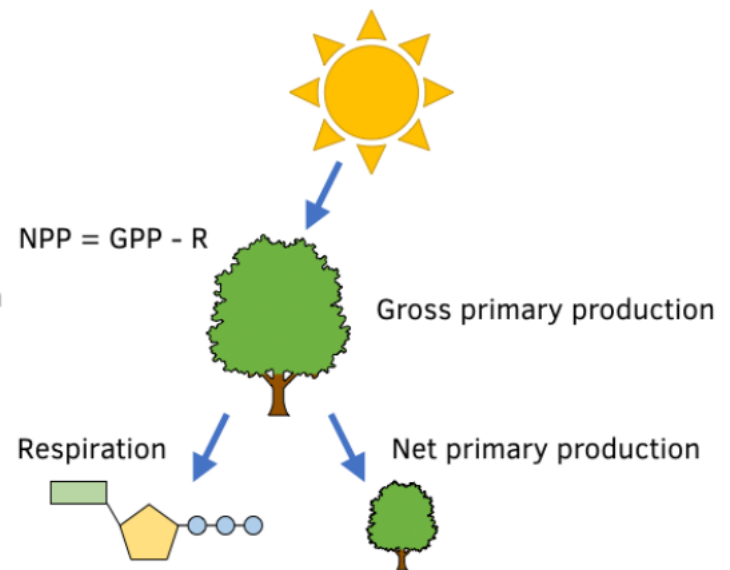
Первичная валовая продукция



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).

GPP = Gross Primary Production
R = Respiratory Losses to the Environment



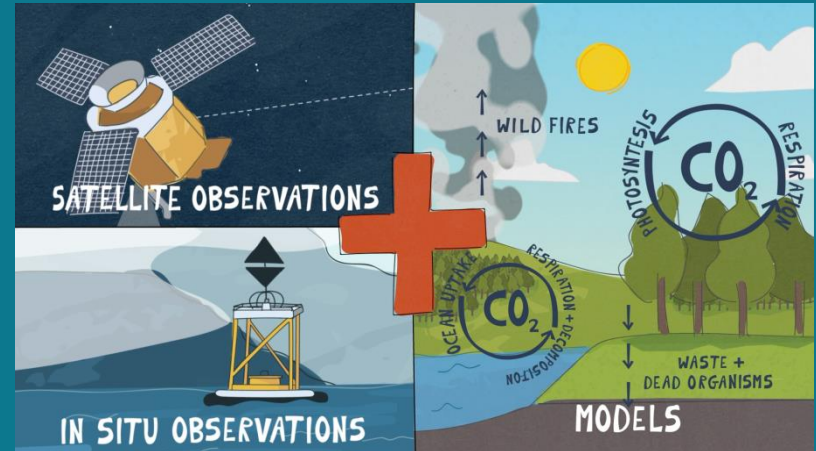
<https://studymind.co.uk/notes/net-primary-production/>



Карбоновый бюджет



Баланс CO₂



Методы машинного обучения



Цель и задачи



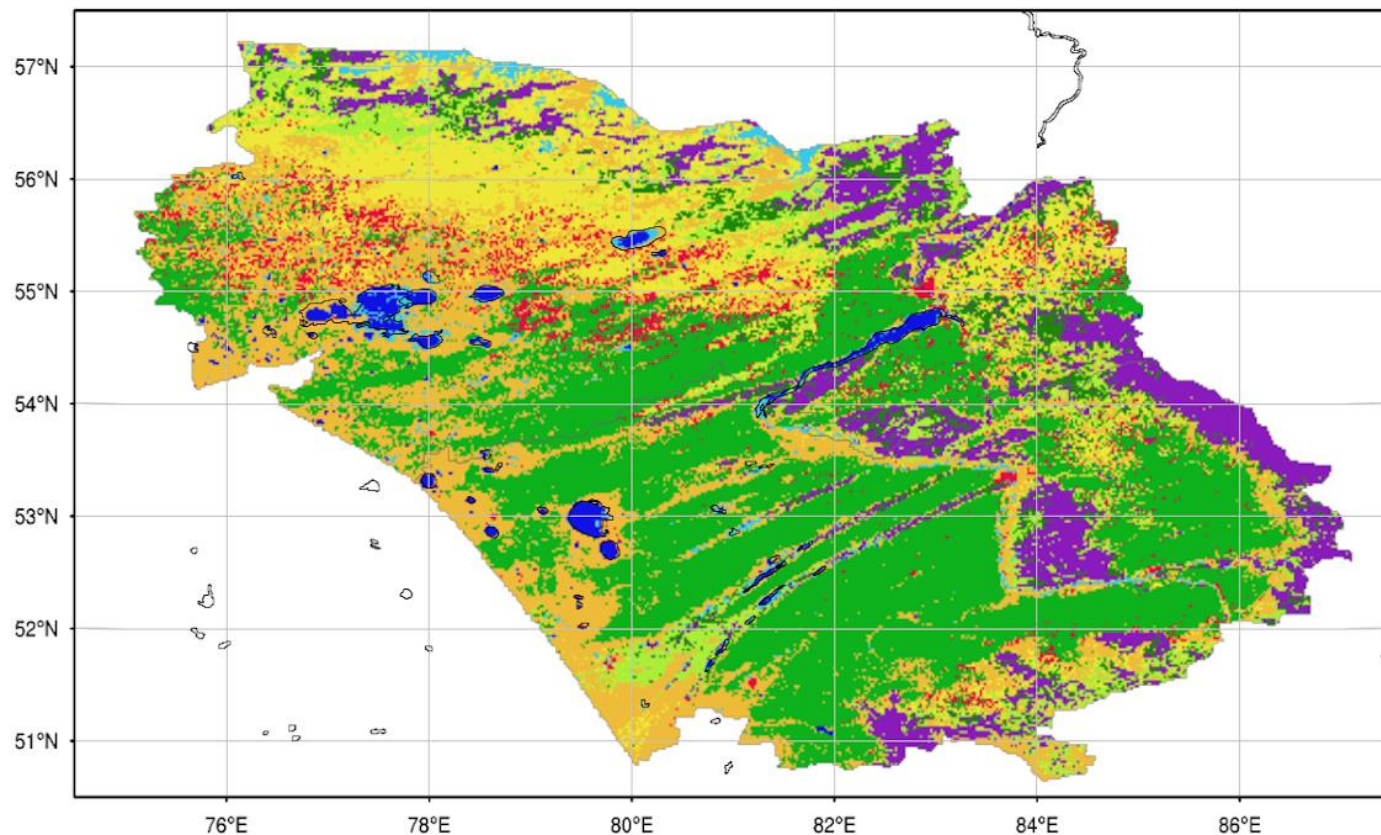
Цель работы - обучение древовидной регрессионной модели Cubist для получения количественных оценок первичной валовой продукции для территории Алтайского края и Новосибирской области.

Решаемые задачи:

- Подготовка обучающей и тестовых выборок с использованием данных продуктов радиометра MODIS/Terra, а также результатов реанализа ERA5 за 2021-2022 гг.
- Обучение древовидной регрессионной модели Cubist, получение деревьев решений.
- Валидация полученной модели.

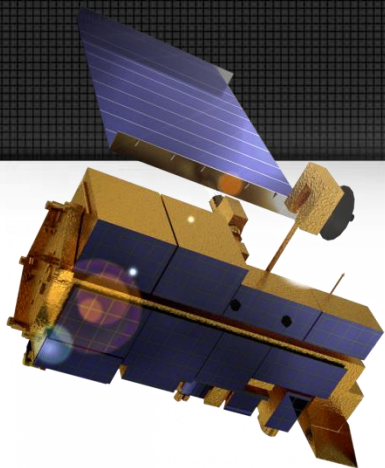


Исследуемые регионы



Тип растительности (продукт MCD12Q1) по классификации UMD университета Мэрилэнда, используемый в обучающей и тестовой выборках модели.

Исходные данные



Переменная выборки	Продукт	Пространственное разрешение	Временное разрешение
GPP	MODIS MOD17A2	500 м	8 дней
EVI	MODIS MOD13A1	500 м	16 дней
LST	MODIS MOD11A2	1 км	8 дней
LCT	MODIS MCD12Q1	500 м	1 год
SW	ERA5 SW	0.25°	1 час
VPD	ERA5 T2M и D2M	0.25°	1 час

Обучающая
выборка

Тестовая
выборка

2021 г.

2022 г.

750000 пикселей

era5



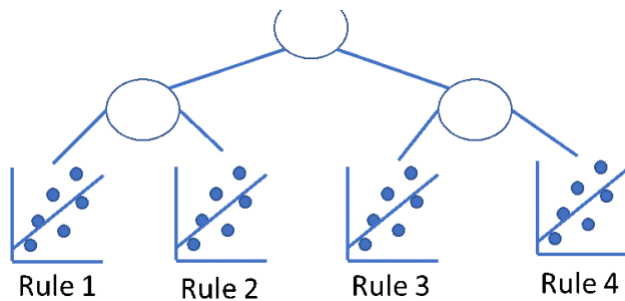
Copernicus
Europe's eyes on Earth



Модель Cubist



Вход – вектор значений целевой переменной (GPP) + вектора переменных предикторов (EVI, LST, LCT, SW, VPD).
Результат – дерево решений, представленное в виде системы правил.



Rule 1: [128063 cases, mean 1.3183997, range 0.008749999 to 5.11, est err 0.3267910]

if SW <= 128.4723 then

$$GPP = 1.9625043 + 0.0126 SW + 4.3 EVI + 0.59 VPD - 0.012 LST$$

Rule 2: [25740 cases, mean 1.6294678, range 0.105 to 7.49875, est err 0.4497935]

if EVI <= 0.3518 LCT > 9 VPD <= 0.6379011 SW > 194.5455 SW <= 220.6714 then

$$GPP = -9.9489539 + 6.7 EVI - 0.0085 SW + 0.042 LST - 0.66 VPD - 0.02 LCT$$

...

Rule 24: [54712 cases, mean 8.1323910, range 0.5599999 to 14.00875, est err 0.8467416]

if EVI <= 0.5633 LST > 298.9207 LCT > 8 VPD <= 1.244966 SW > 260.9348 then

$$GPP = 47.647412 + 12.2 EVI - 0.136 LST - 0.0093 SW - 1.26 VPD + 0.08 LCT$$

Rule 25: [4716 cases, mean 8.1611423, range 1.0675 to 14.44625, est err 1.4273475]

if LST <= 302.87 LCT > 9 VPD > 0.8216063 VPD <= 0.9621247 SW > 237.1799 SW <= 245.9033 then

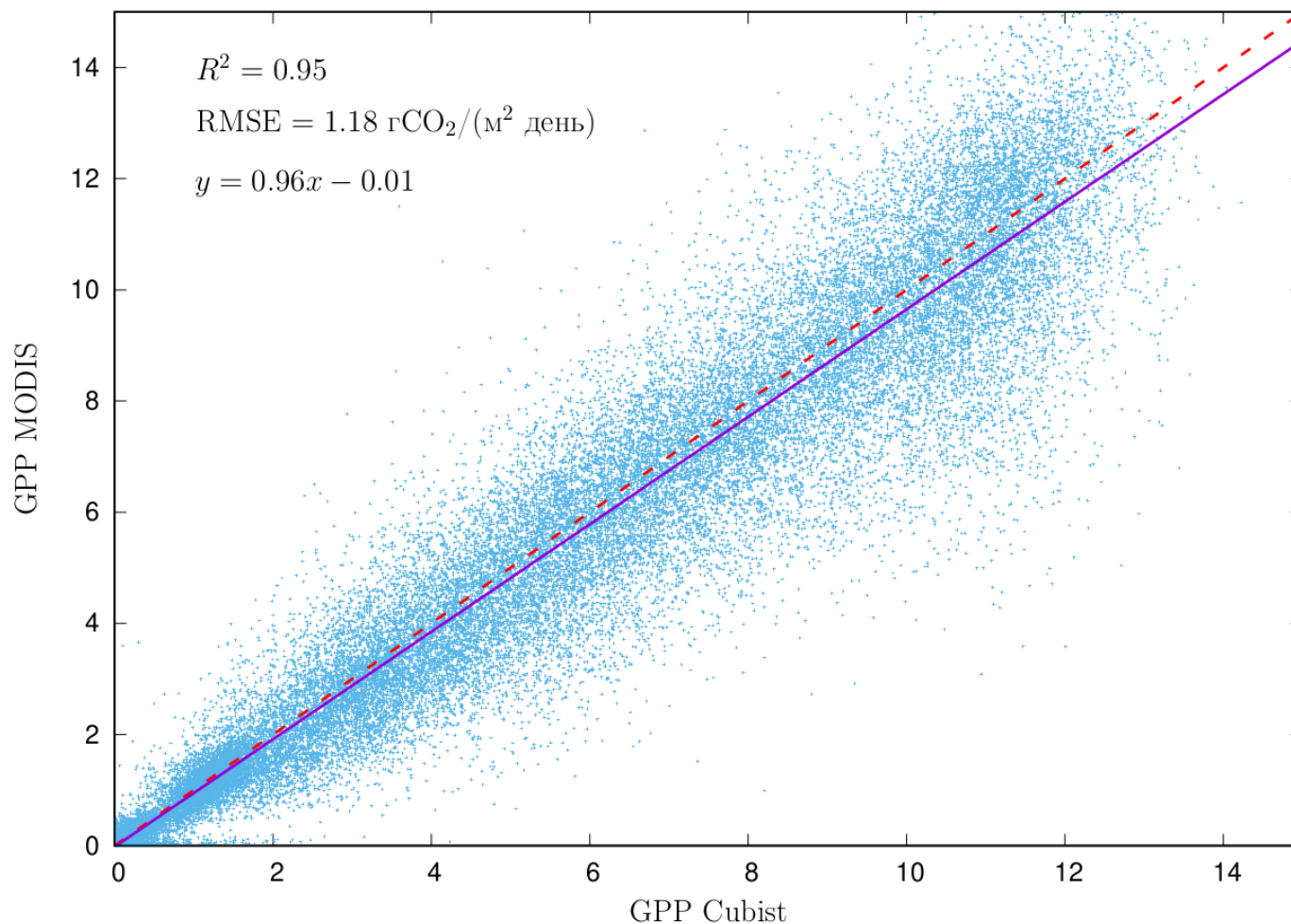
$$GPP = -68.3962626 + 0.154 SW - 7.37 VPD + 13.5 EVI + 0.129 LST + 0.07 LCT$$



Валидация модели



Отложенная выборка (hold-out split валидация)

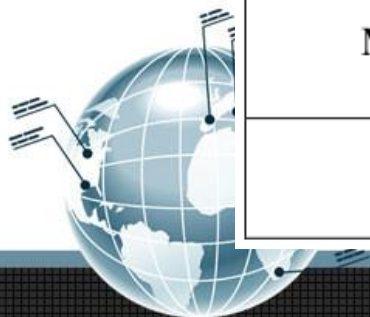


Валидация модели



Метрики регрессии

Метрика	Определение	Значение (без размерностей)
MSE	$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	1.45
RMSE	$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	1.18
MAE	$\text{MAE} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	0.83
MAPE	$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right $	0.37
MSLE	$\text{MSLE} = \frac{1}{n} \sum_{i=1}^n \{\ln(1 + y_i) - \ln(1 + \hat{y}_i)\}^2$	0.0035
R^2	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	0.95



Валидация модели



Кросс-валидация по типам растительности

Тип растительности	R^2	RMSE, гCO ₂ /(м ² день)	Объем выборки (количество пикселей)
Хвойный лес	0.87	1.34	7500
Лиственный лес	0.87	1.73	25500
Смешанный лес	0.90	1.42	81000
Кустарники	0.86	1.58	145000
Травы	0.88	1.18	165000
Сельскохозяйственные культуры	0.89	1.19	255000



Результаты и выводы



1. С использованием обучающей выборки, включающей данные продуктов радиометра MODIS спутника Terra, а также результаты реанализа ERA5 для территории Алтайского края и Новосибирской области за 2021 г., сформировано дерево решений модели Cubist для получения количественных оценок первичной валовой продукции регионов.
2. Проведена валидация модели двумя методами: валидация на основе отложенной выборки и кросс-валидация по типам растительности. Результаты hold-out split валидации дают значения метрик регрессии $R^2 = 0.95$ и $RMSE = 1.18 \text{ гСО}_2/(\text{м}^2 \text{ день})$. Результаты анализа в случае кросс-валидации по типам растительности дают значения коэффициента детерминации $R^2 > 0.85$.



Полученные результаты позволяют сделать вывод о высоком качестве обучения модели Cubist и о возможности ее применения для моделирования первичной валовой продукции регионов.

Спасибо за внимание!

Вопросы и комментарии
volkov@theory.asu.ru



Работа выполнена при поддержке ФИЦ ИВТ (проект “Методы и технологии обработки временных серий спутниковых данных для оценки региональных источников парниковых газов”).

