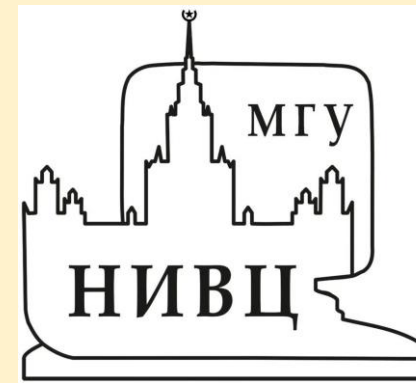




Реализация локальных моделей машинного обучения для аппроксимации коротковолновой радиации у поверхности земли в облачных условиях по геофизическим предикторам для Московского региона

Н.А. Петров¹ (nial.03@mail.ru), Н.Е. Чубарова¹, М.И. Варенцов^{1,2}

1 - Географический факультет МГУ им. М.В. Ломоносова

2 – Научно-исследовательский вычислительный центр МГУ им. М.В. Ломоносова

10-я Международная конференция по глубокому обучению в вычислительной физике (DLCP-2026)

г. Москва, НИИЯФ МГУ, 6 - 8 июля 2026 г.

Актуальность темы работы и её мотивация

Радиационный комплекс RAD-MSU (BSRN)



Фото и расположение
радиационного комплекса
RAD-MSU (BSRN)
(55.707 °N, 37.52 °E, H = 192 m)

Прямая коротковолновая
радиация



CHP1 Pyrheliometer

Рассеянная радиация
Суммарная радиация

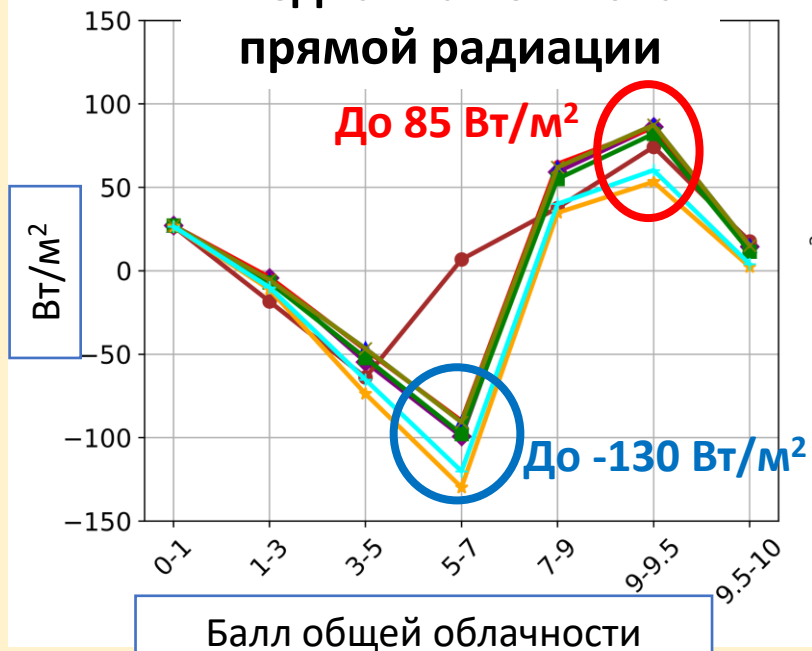


CMP21 Pyranometer

Источник: <https://www.kippzonen.com>

Модель esRad

Медианная ошибка
прямой радиации



Источник: [Петров, Чубарова, Физика
атмосферы и океана, 2026]

- Компоненты коротковолновой радиации являются **важными** составляющими **радиационного баланса**;
- В условиях **неоднородной облачности** потоки коротковолновой радиации **сильно трансформируются**;
- **Модели** восстановления коротковолновой радиации будут **полезны** для **восстановления пропусков приборов RAD-MSU на время нештатных ситуаций**;
- **Модели** позволят выявить наиболее **важные и физически обоснованные предикторы** компонентов **солнечной радиации**

Формулировка проблемы и цели работы

Общая проблема работы:

Восстановление и заполнение пропусков в данных измерений плотностей потоков коротковолновой радиации высокоточным радиационным комплексом RAD-MSU (BSRN) по известным предикторам

Общая формулировка цели:

Создание и тестирование локальных регрессионных моделей аппроксимации плотностей потоков прямой, рассеянной и суммарной коротковолновой радиации в облачных условиях в зависимости от геофизических и метеорологических предикторов с использованием классических подходов машинного обучения для метеорологической обсерватории (МО) МГУ.

Постановка задач в терминах машинного обучения

Класс задачи МО



С учителем (Supervised)

Вид задачи МО



Регрессия

Целевые переменные



Прямая, рассеянная,
суммарная радиация

Временной диапазон и
дискретность данных



С августа 2021 по июль 2024 гг.
Дискретность: 1 час

Географический регион



МО МГУ (55.7 с.ш., 37.5 в.д.)

Фильтрация данных



Высота Солнца > 10°, продолжительность
солнечного сияния < 60 мин/час, балл
облачности > 0

Выбор метеорологических и геофизических предикторов

Физически обоснованный выбор различных предикторов

Астрономические
параметры

Синус высоты Солнца, внеатмосферная
солнечная радиация на ВГА: расчёты по
астрономическим формулам

Метеорологические
предикторы

Атмосферное давление на у.м., нижний и общий
балл облачности: МО МГУ

Водосодержание облаков, влагосодержание
атмосферы: ERA-5

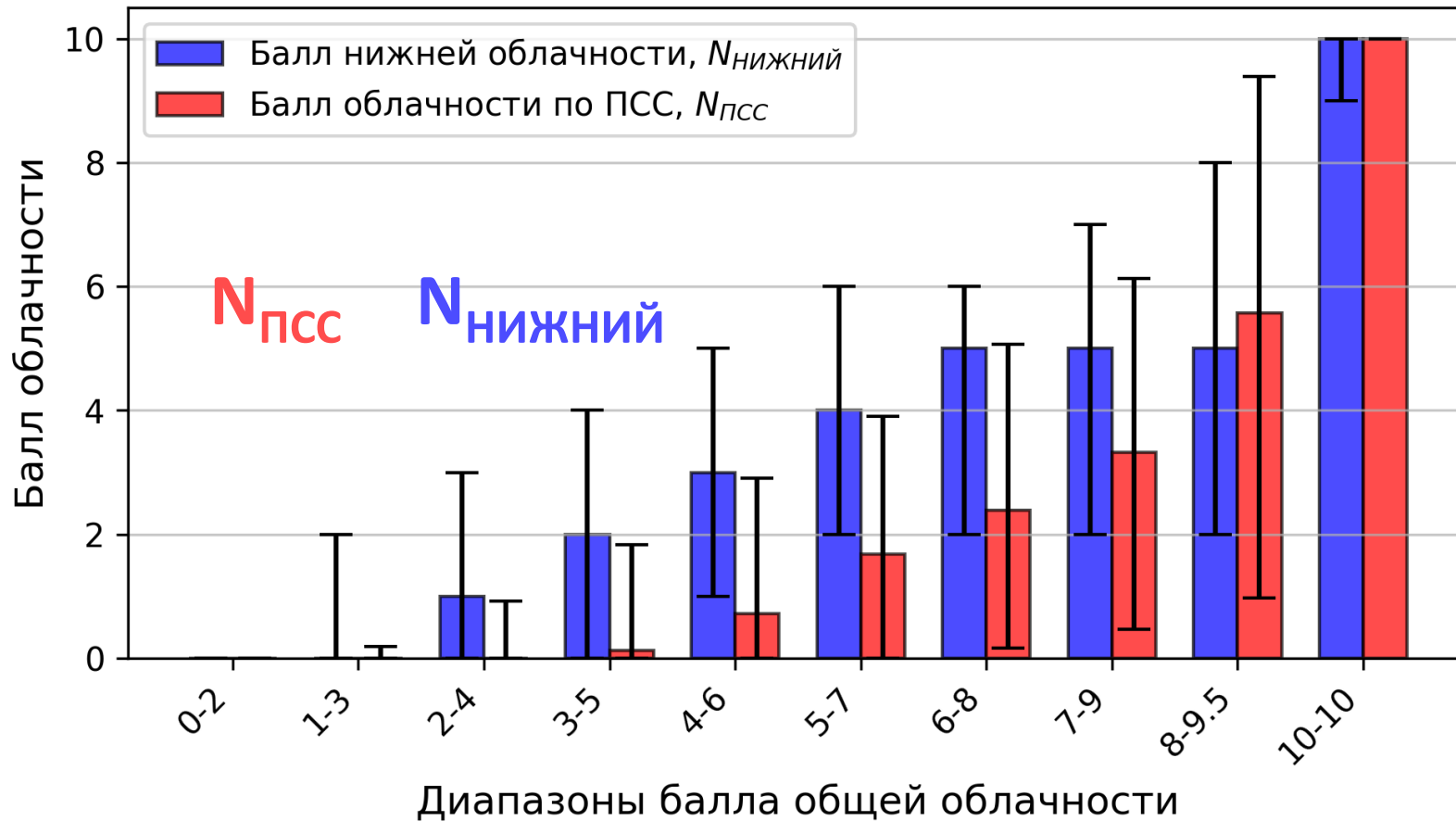
Геофизические
предикторы

Продолжительность солнечного сияния
(ПСС), альbedo поверхности: RAD-MSU

Аэрозольная оптическая толщина: CAMS

Метрика «временного» балла облачности

Медианные баллы облачности по диапазонам



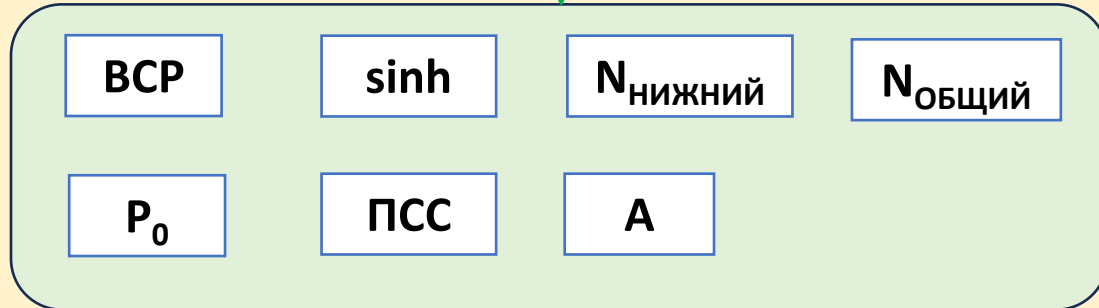
ПСС – продолжительность
солнечного сияния

$$N_{\text{ПСС}} = 10 * \left(1 - \frac{\text{ПСС(мин)}}{60}\right)$$

Балл по ПСС ($N_{\text{ПСС}}$) –
характеристика
относительного времени
закрытости солнечного
диска – по эргодичности
соответствует нижнему
баллу плотных облаков:

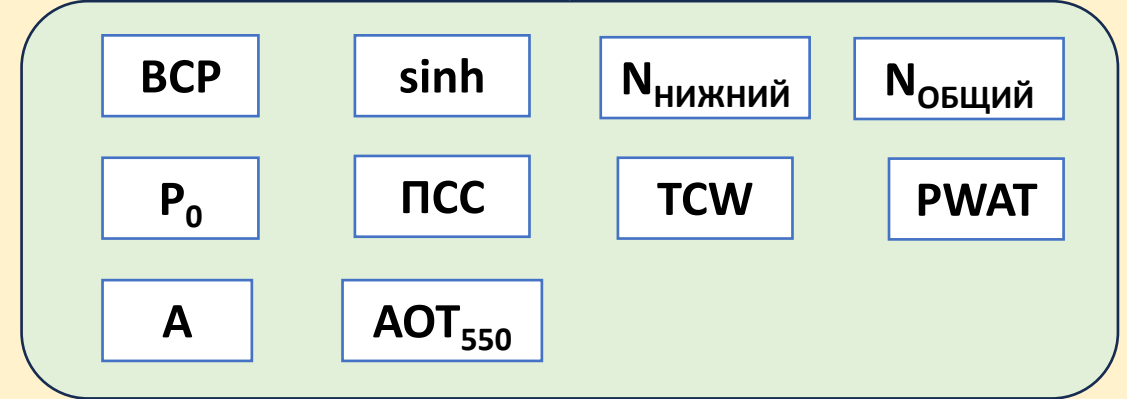
Формирование баз данных предикторов

Без данных реанализов



BCP – внеатмосферная солнечная радиация
sinh – синус высоты Солнца
 $N_{\text{нижний}}$ и **$N_{\text{общий}}$** – нижний и общий баллы облачности
 P_0 – атмосферное давление МО МГУ
ПСС – продолжительность солнечного сияния
А – альbedo поверхности

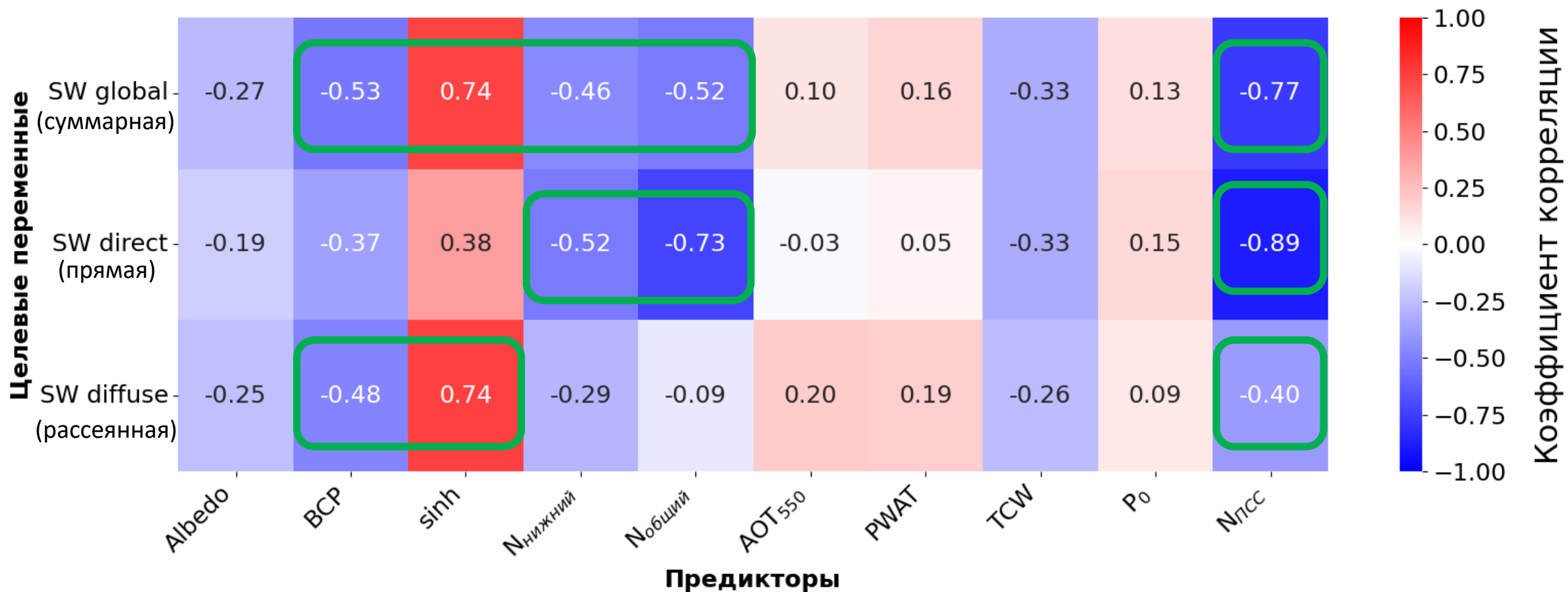
С данными реанализов



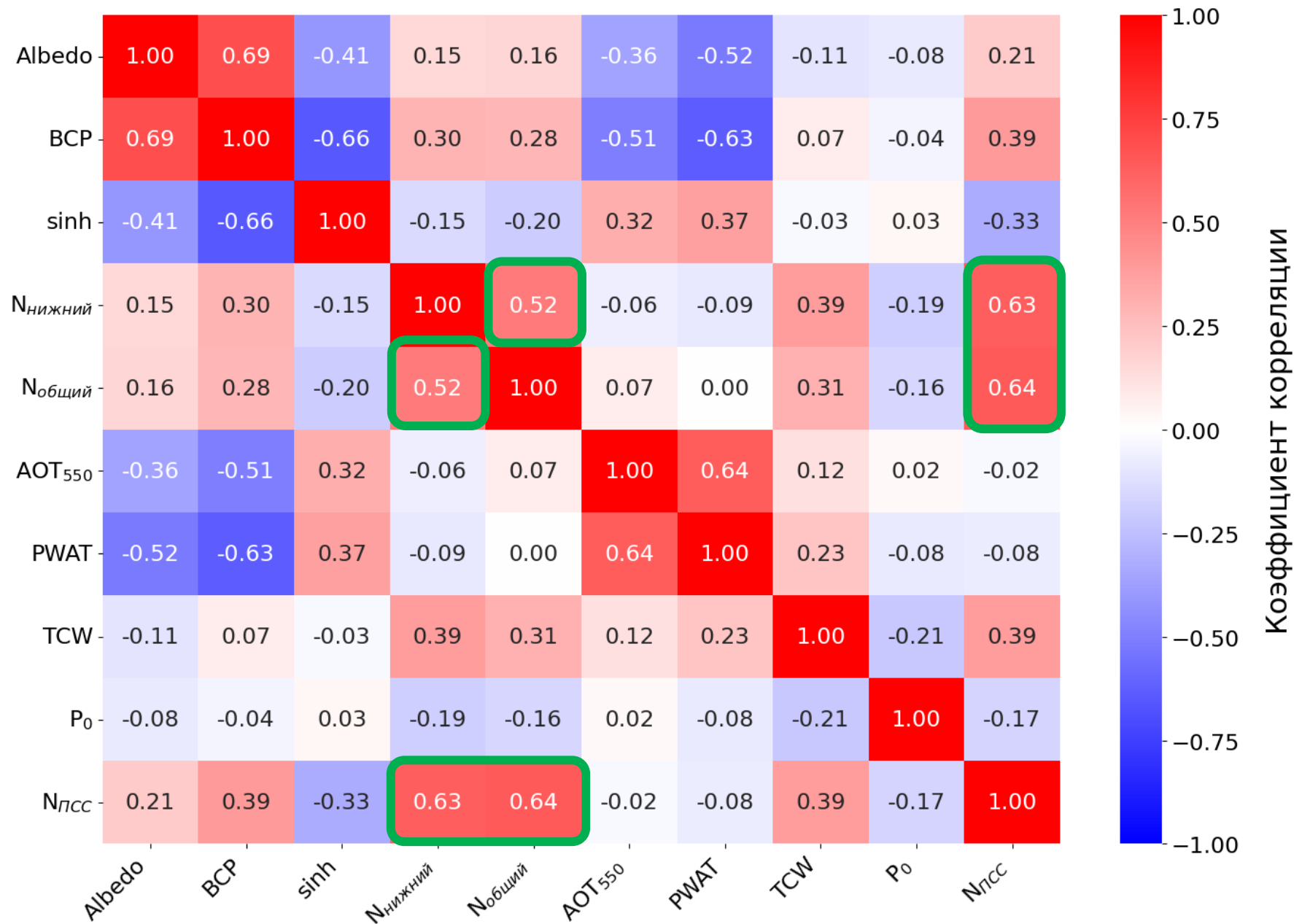
TCW – интегральное водосодержание облачности
PWAT – интегральное влагосодержание атмосферы
 AOT_{550} – аэрозольная оптическая толщина на длине волны 550 нм

Разведочный анализ данных (EDA)

Корреляции между предикторами и целевой переменной



Разведочный анализ данных (EDA)



Корреляции между предикторами:

- Нижний и общий балл облачности (**0.52**)
- Нижний балл и балл по ПСС (**0.63**)
- Общий балл и балл по ПСС (**0.64**)

Идея преобразований:

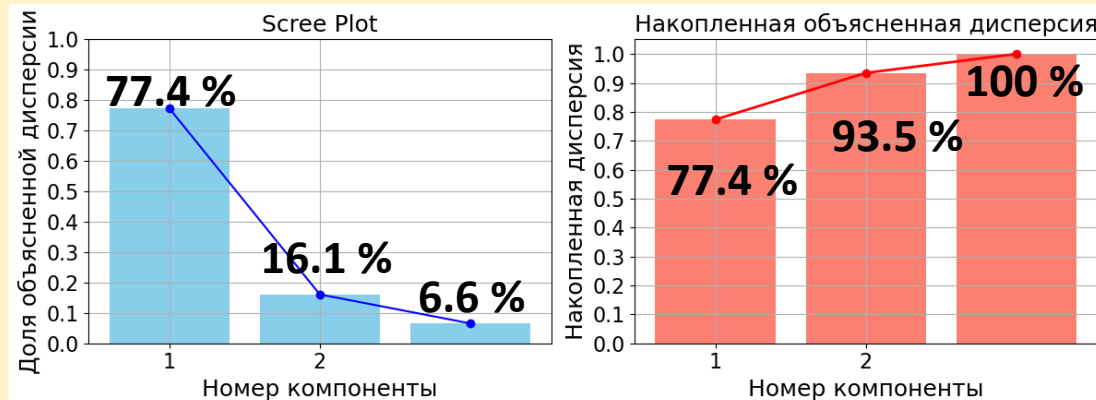
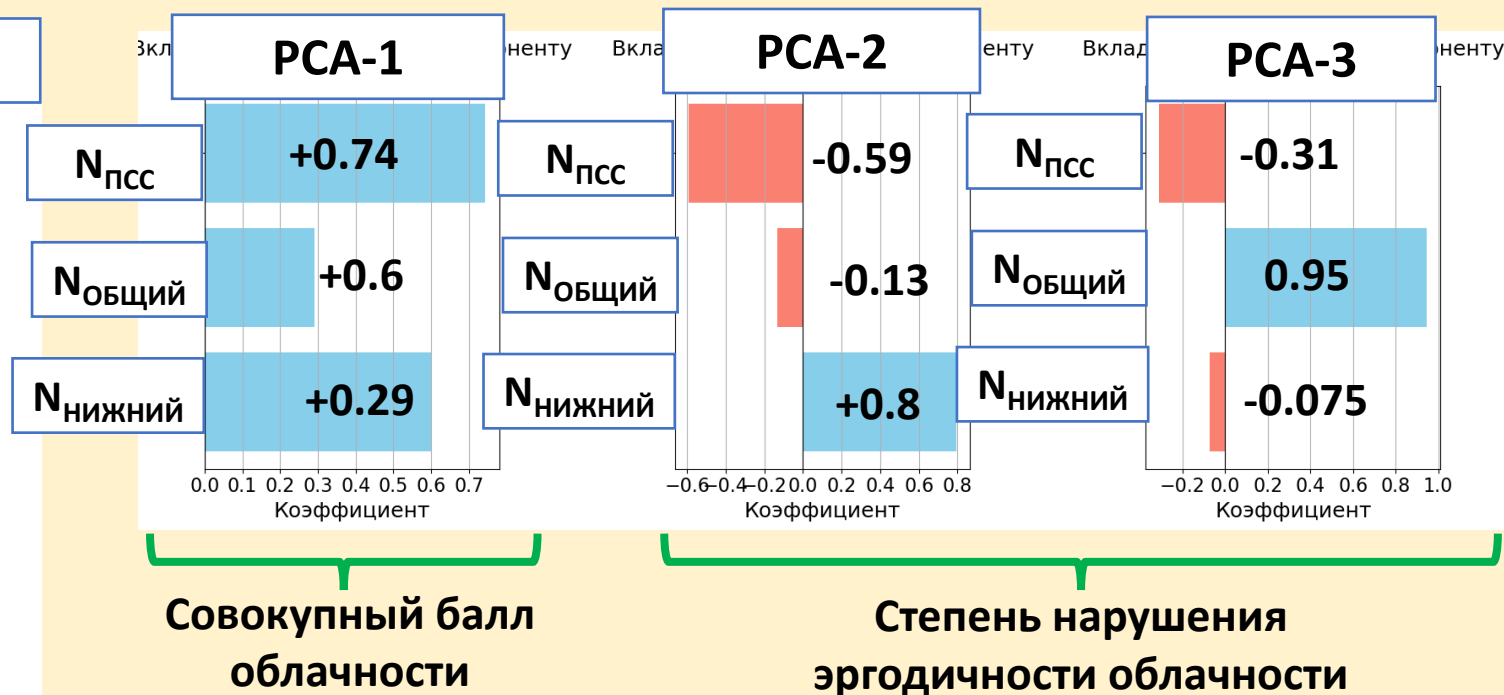
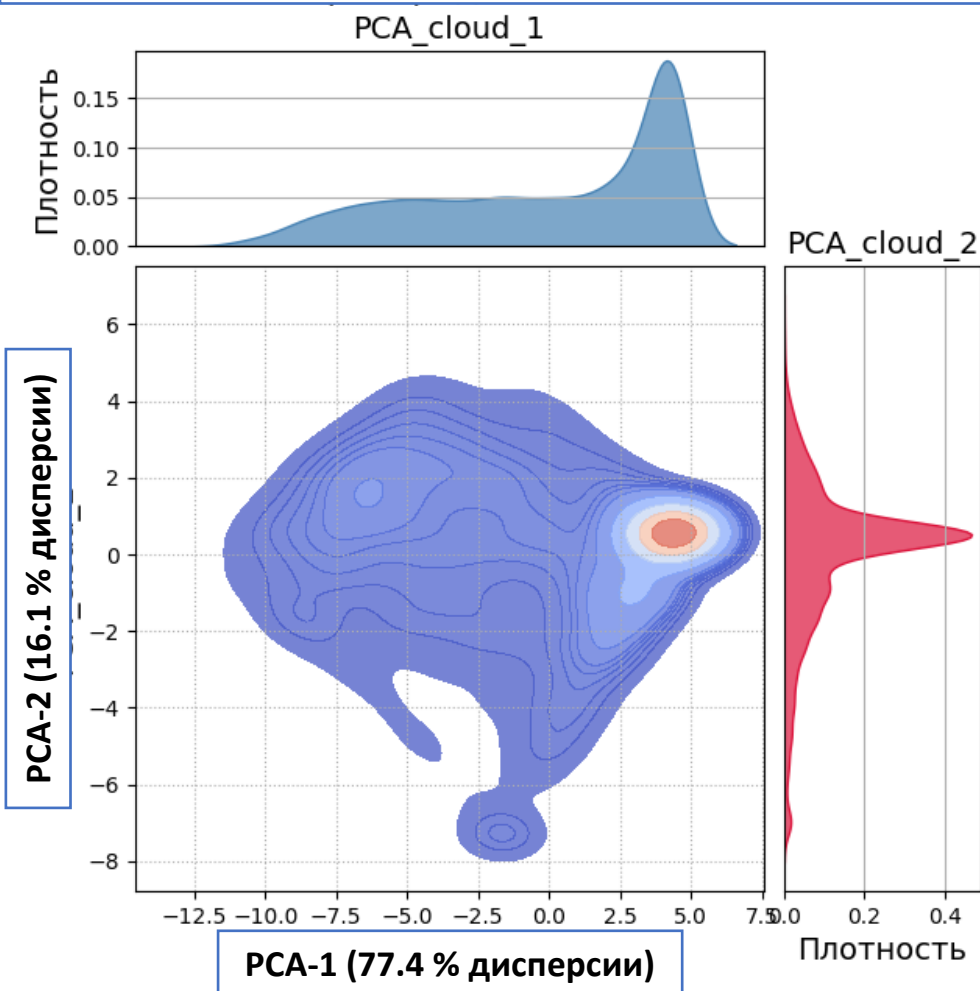
Характеристики балла облачности: N_{нижний}, N_{общий}, ПСС сформировать по методу PCA в три компоненты

Применение метода PCA и трансформация признаков

Балл по ПСС ($B_{\text{ПСС}}$) – характеристика времени *открытости* солнечного диска – по эргодичности соответствует нижнему баллу плотных облаков:

$$N_{\text{ПСС}} = 10 \left(1 - \frac{\text{ПСС}(\text{мин})}{60} \right)$$

PCA – метод главных компонент



Использованные классические модели машинного обучения

Линейная
регрессия

К-ближайших
соседей (KNN)

Случайный лес –
Random forest

Градиентный
бустинг – Catboost

Различные
конфигурации расчётов
по предикторам

Без реанализов

С PCA для облачности

Без PCA для облачности

С реанализами

С PCA для облачности

Без PCA для облачности

Метод кросс-валидации для оценки неопределённости качества

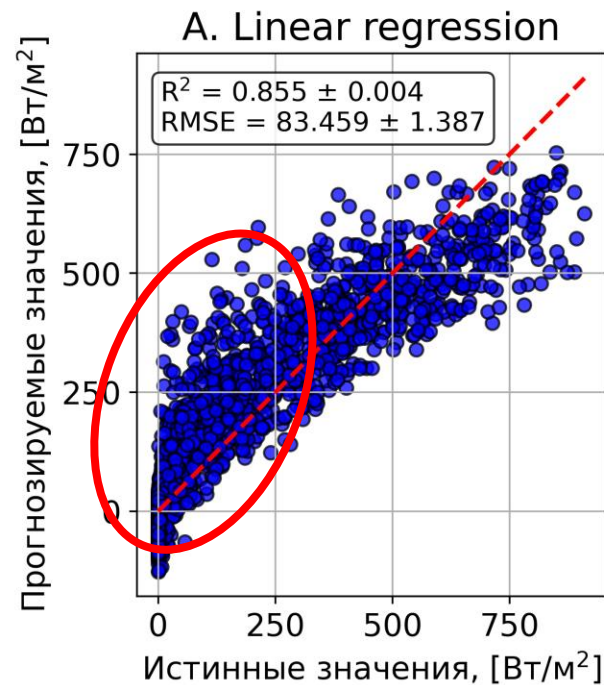
- ✓ Был применён метод Bootstrap для формирования выборок для оценки неопределённости моделей МО методом кросс-валидации
- ✓ Чаще всего выбирались параметры: ~7000 случаев в выборке (размер) и 50 bootstrap-выборок

Bootstrap



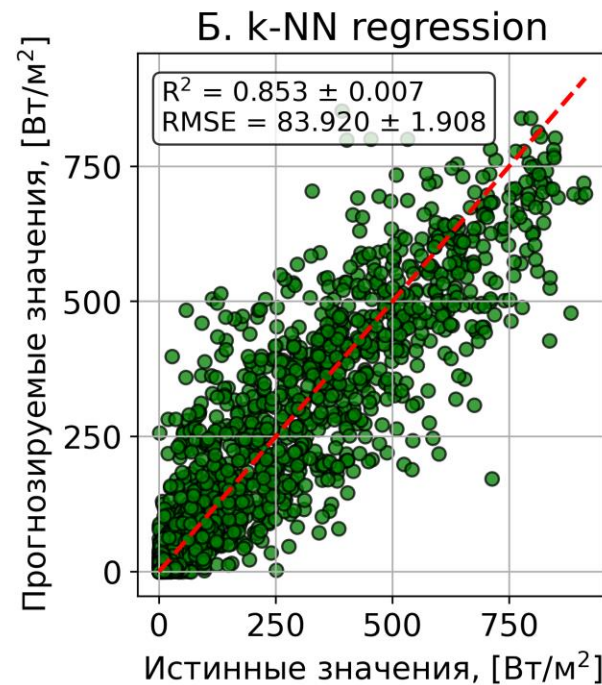
Результаты расчётов с «классическими» моделями МО

- Сравнение моделей МО различной архитектуры на тестовой выборке в подходе кросс-валидации методом Bootstrap
- Конфигурация предикторов: с учётом реанализа и метода PCA для облачности



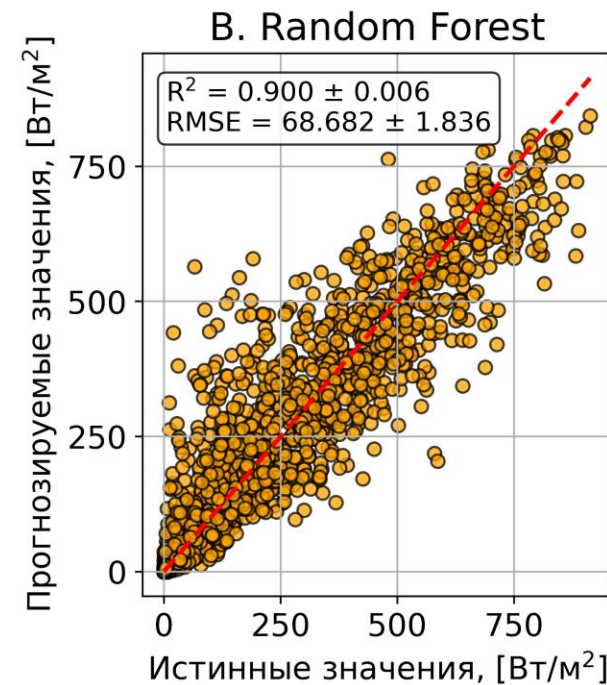
$$RMSE = 83.5 \pm 1.4 \text{ Вт/м}^2$$

$$R^2 = 0.855 \pm 0.004$$



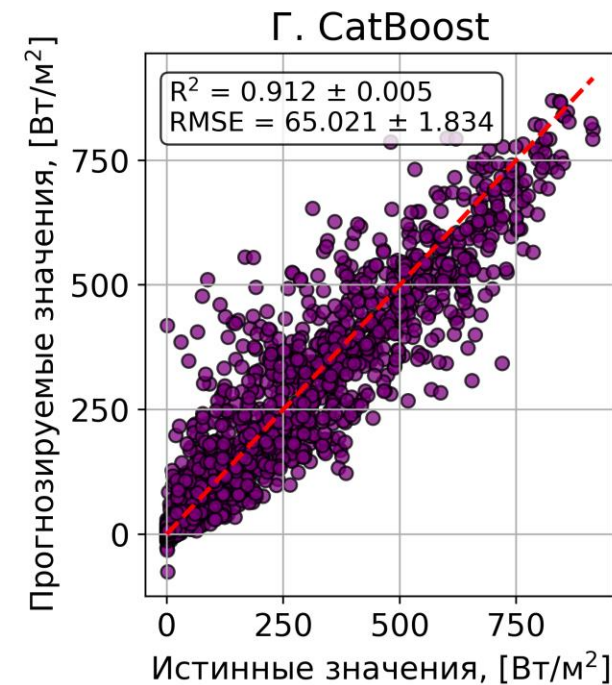
$$RMSE = 83.9 \pm 1.9 \text{ Вт/м}^2$$

$$R^2 = 0.853 \pm 0.007$$



$$RMSE = 68.7 \pm 1.8 \text{ Вт/м}^2$$

$$R^2 = 0.900 \pm 0.006$$



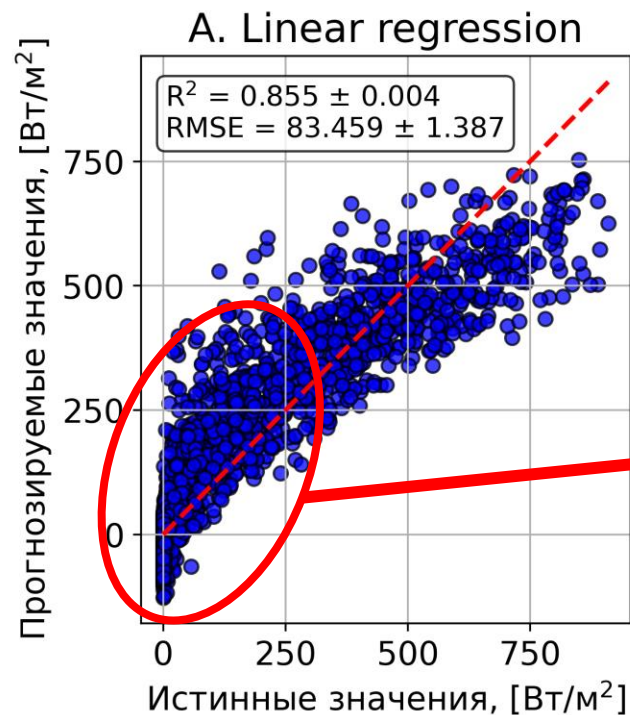
$$RMSE = 65.0 \pm 1.8 \text{ Вт/м}^2$$

$$R^2 = 0.912 \pm 0.005$$

Эффект нелинейности прямой радиации от балла облачности

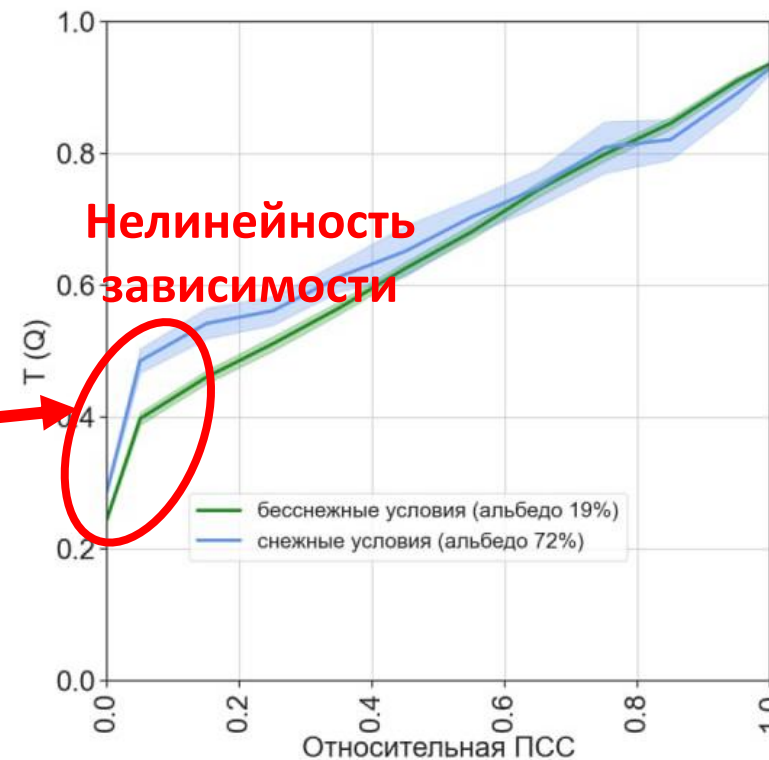
- Сравнение моделей МО различной архитектуры на тестовой выборке в подходе кросс-валидации методом Bootstrap
- Конфигурация предикторов: с учётом реанализа и метода PCA для облачности

Облачное пропускание



$$RMSE = 83.3 \pm 1.4 \text{ Вт/м}^2$$

$$R^2 = 0.855 \pm 0.004$$



Зависимость коэффициента пропускания облаков $T(Q)$ от относительной часовой продолжительности солнечного сияния (ПСС) в бесснежных и снежных условиях. Август 2021 – апрель 2025. Заливкой показаны доверительные интервалы (95%).

Коэффициент пропускания:

$$T(Q) = \frac{Q}{Q_0}$$

Q – фактическая суммарная радиация, приведенная к среднему расстоянию от Земли до Солнца
 Q_0 – суммарная радиация в безоблачном небе

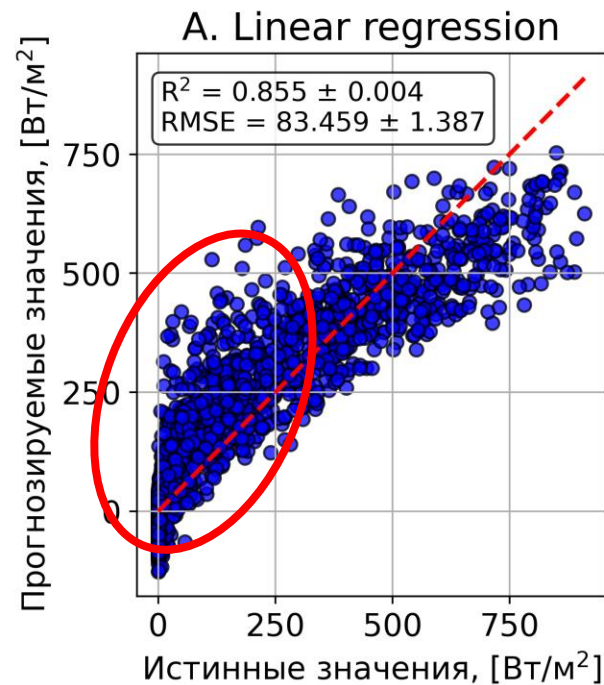
Относительная ПСС:

$$\text{Отн. ПСС} = \frac{\text{ПСС}}{60}$$

[Piskunova et al., 2024]

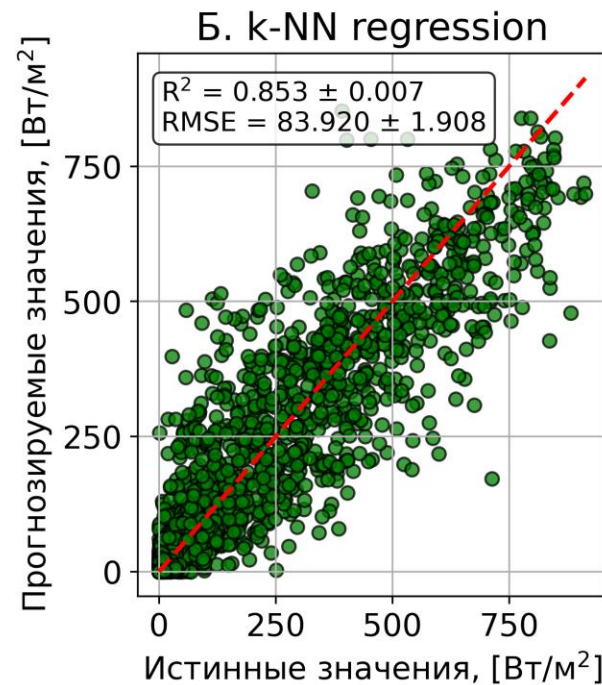
Результаты расчётов с «классическими» моделями МО

- Сравнение моделей МО различной архитектуры на тестовой выборке в подходе кросс-валидации методом Bootstrap.
- Конфигурация предикторов: с учётом реанализа и метода PCA для облачности



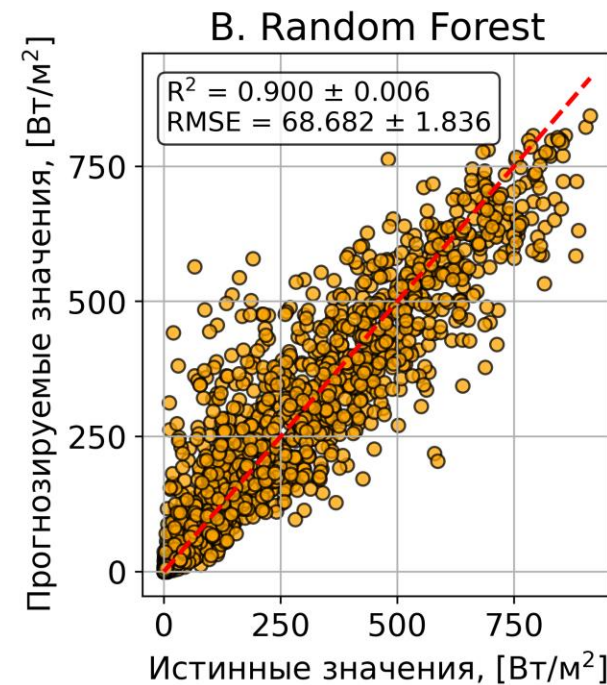
$$RMSE = 83.5 \pm 1.4 \text{ Вт/м}^2$$

$$R^2 = 0.855 \pm 0.004$$



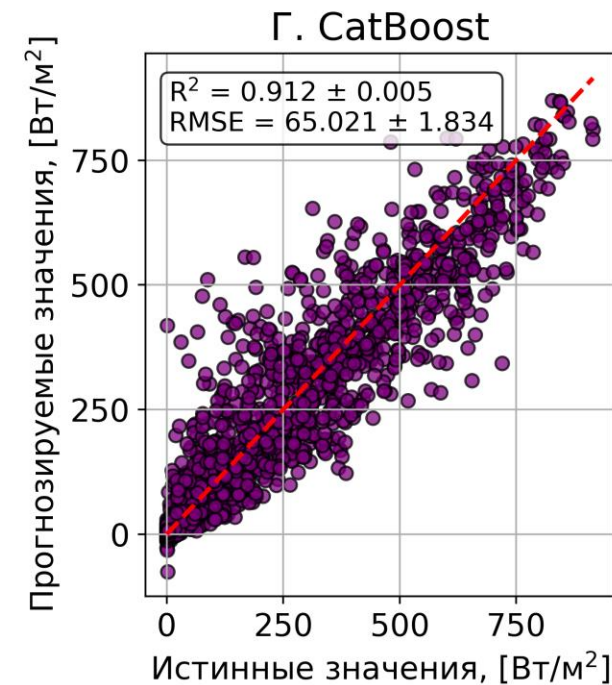
$$RMSE = 83.9 \pm 1.9 \text{ Вт/м}^2$$

$$R^2 = 0.853 \pm 0.007$$



$$RMSE = 68.7 \pm 1.8 \text{ Вт/м}^2$$

$$R^2 = 0.900 \pm 0.006$$



$$RMSE = 65.0 \pm 1.8 \text{ Вт/м}^2$$

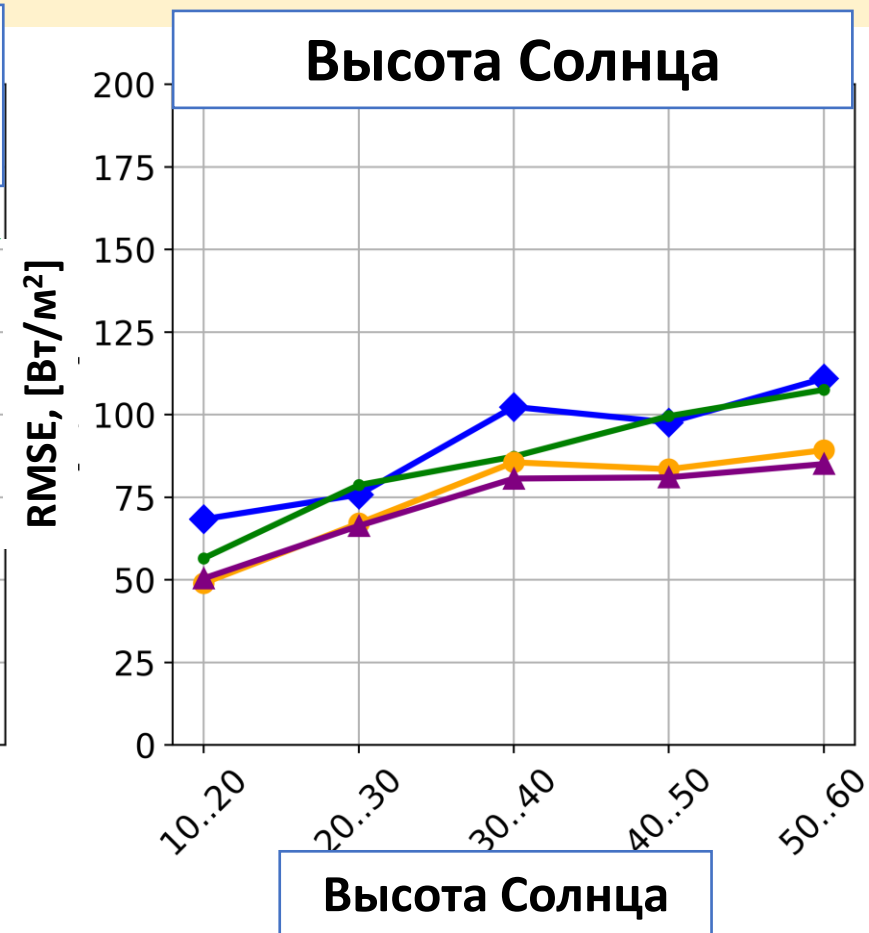
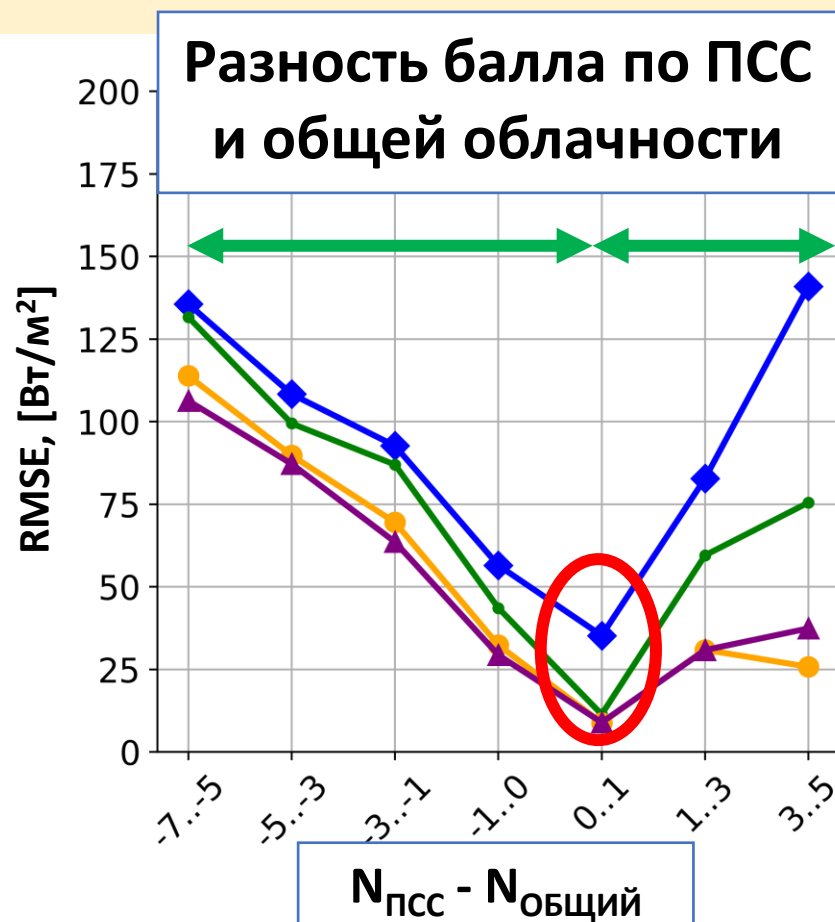
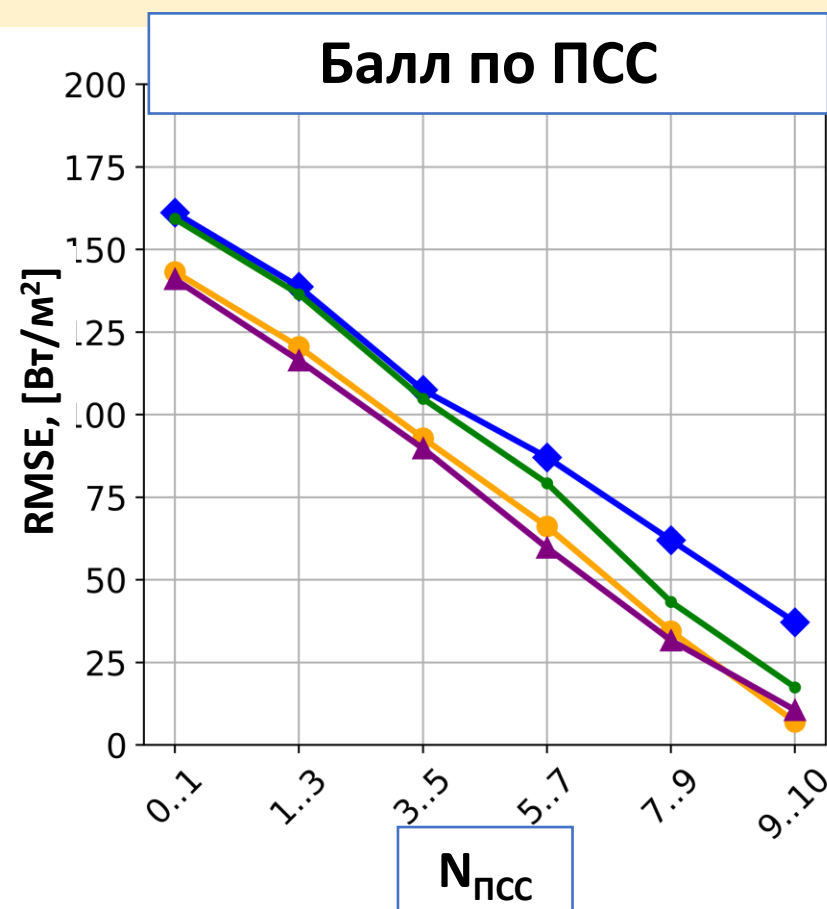
$$R^2 = 0.912 \pm 0.005$$

Зависимости RMSE локальных моделей МО от важнейших предикторов

- RMSE линейно уменьшается с баллом облачности по ПСС
- Эффекты нарушения эргодичности облачного покрова
- Увеличение RMSE с ростом высоты Солнца

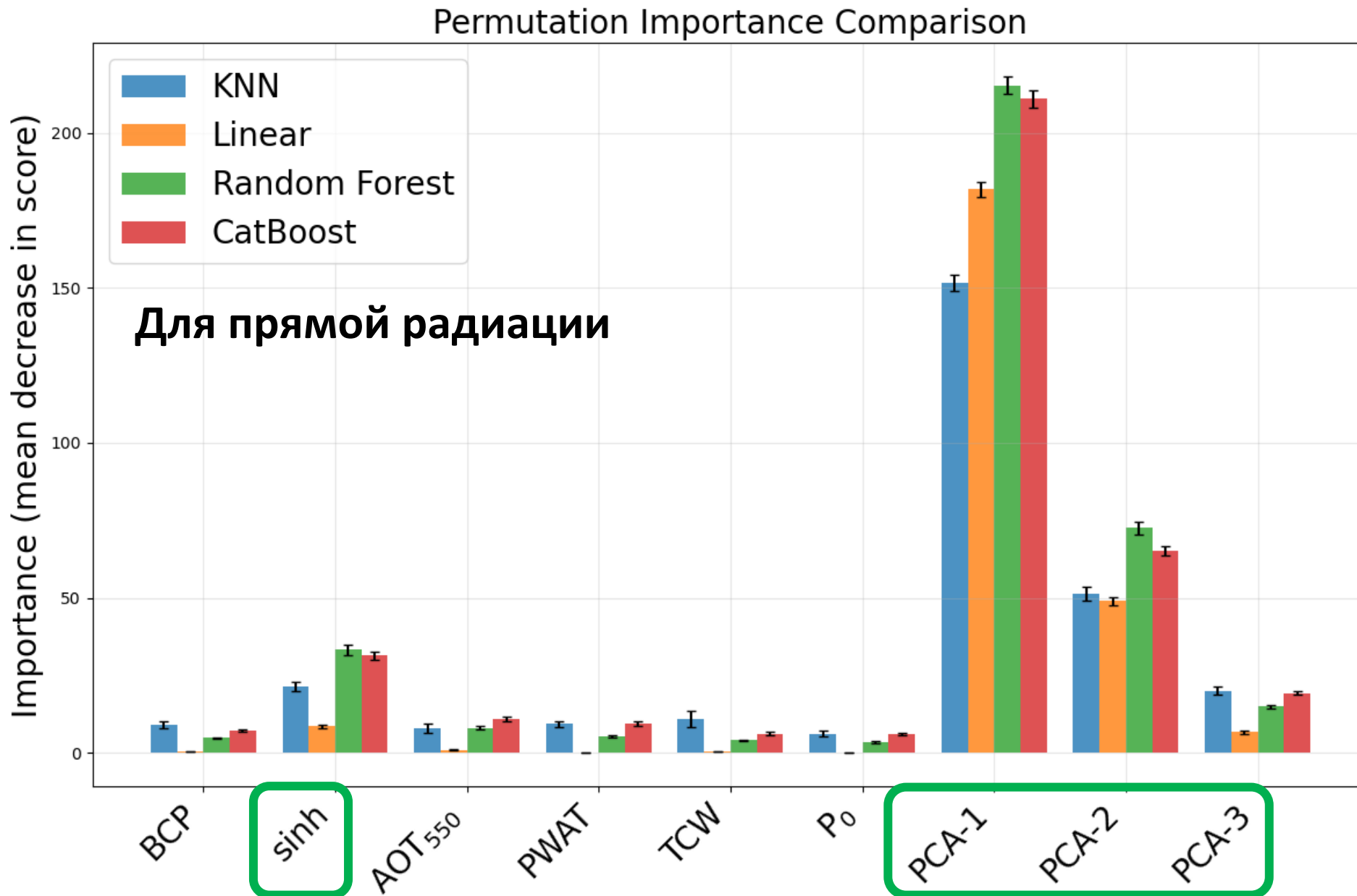
Балл облачности по ПСС:

$$N_{\text{ПСС}} = 10 \left(1 - \frac{\text{ПСС}(\text{мин})}{60} \right)$$



—◆— Linear —●— k-NN regression —●— Random Forest —▲— CatBoost

Перестановочная значимость признаков для моделей МО



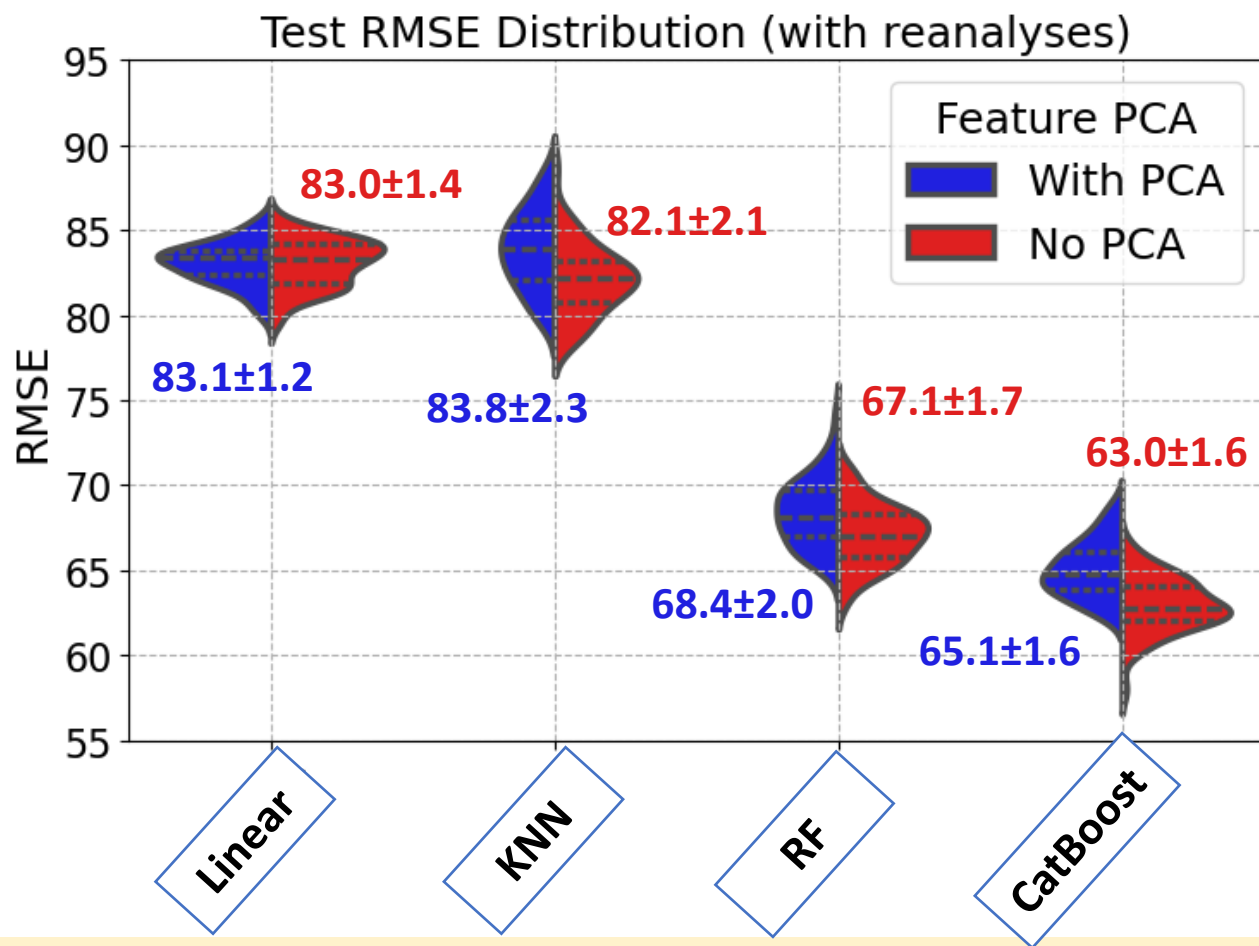
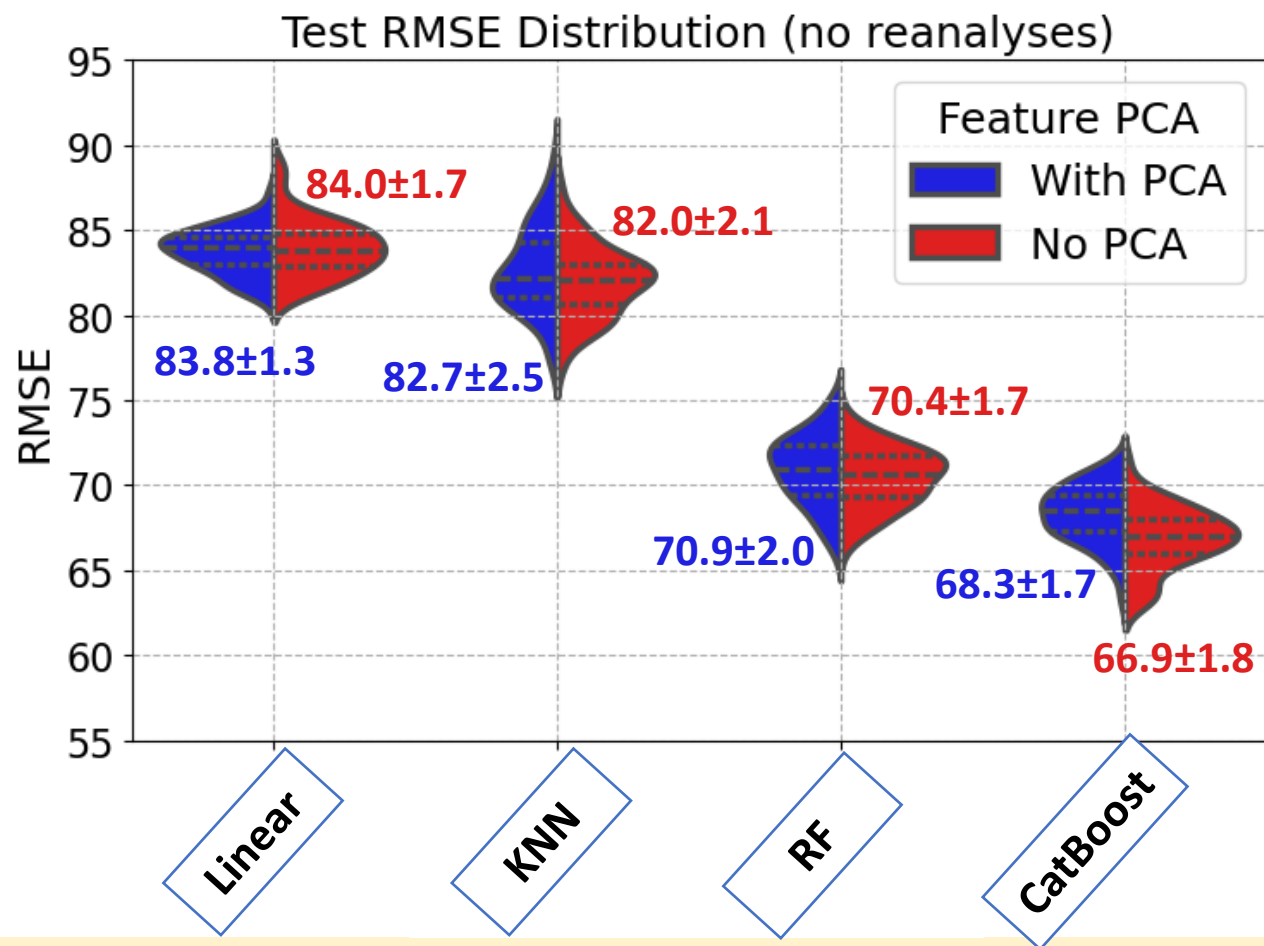
Наиболее значимыми параметрами для всех моделей по принципу перестановочной важности (feature importances) являются:

1. Компоненты балла облачности (PCA-1, PCA-2, PCA-3)
2. Синус высоты Солнца (sinh)

При этом для всех моделей сохраняется важность одних и тех же параметров

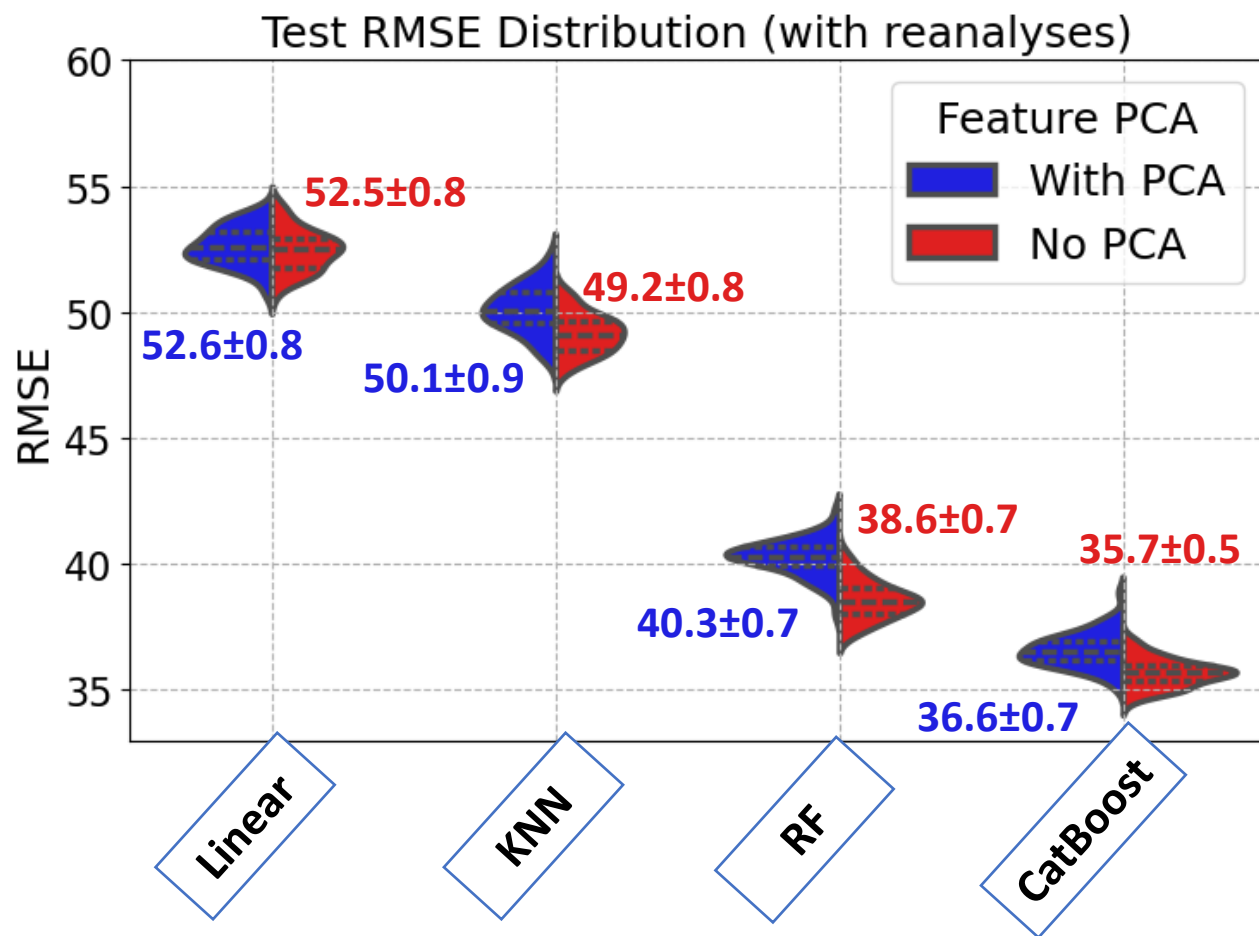
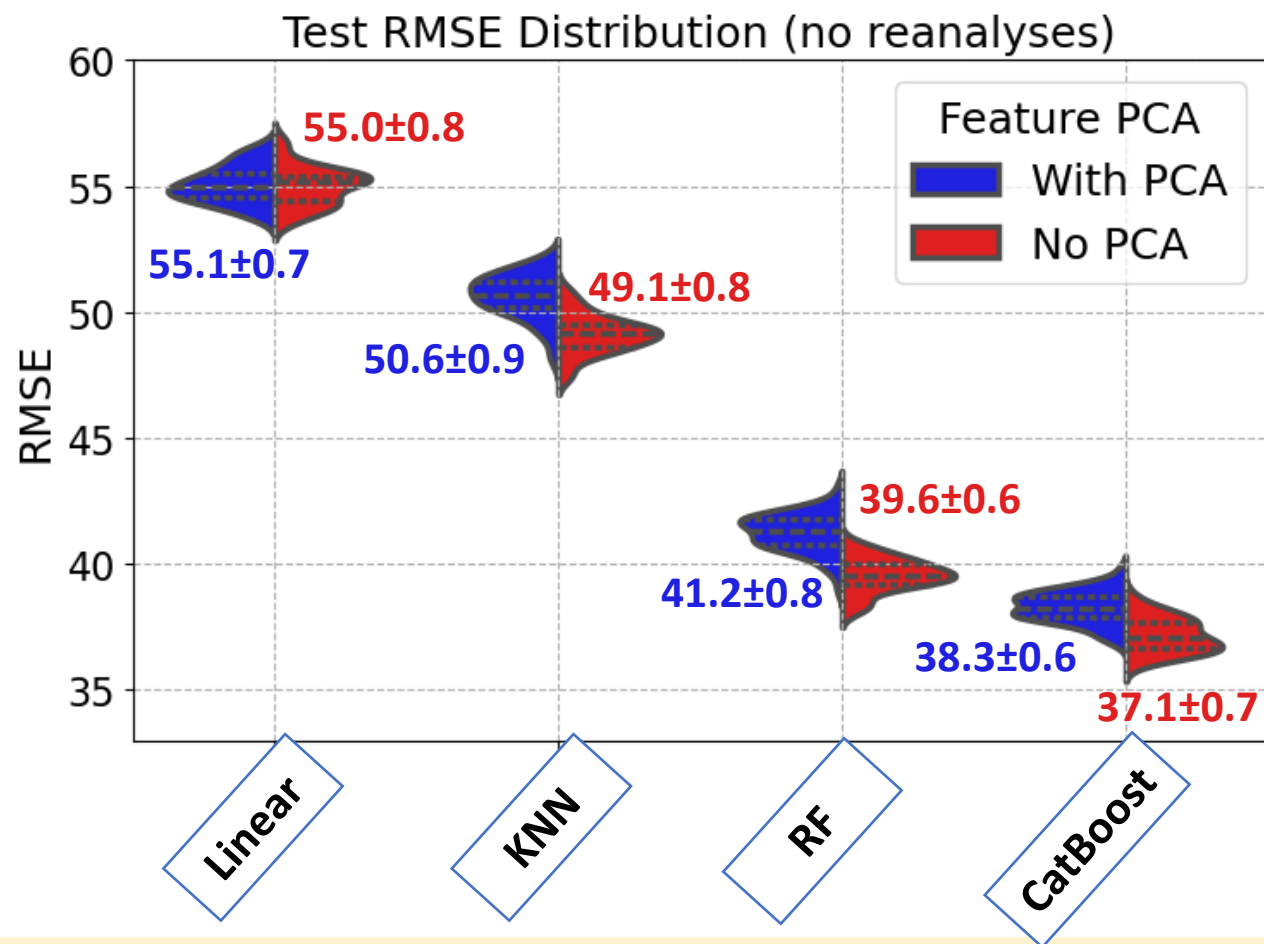
Сравнение различных конфигураций моделей машинного обучения для **прямой** радиации

Split Violin Plot Comparison: With vs Without Feature PCA



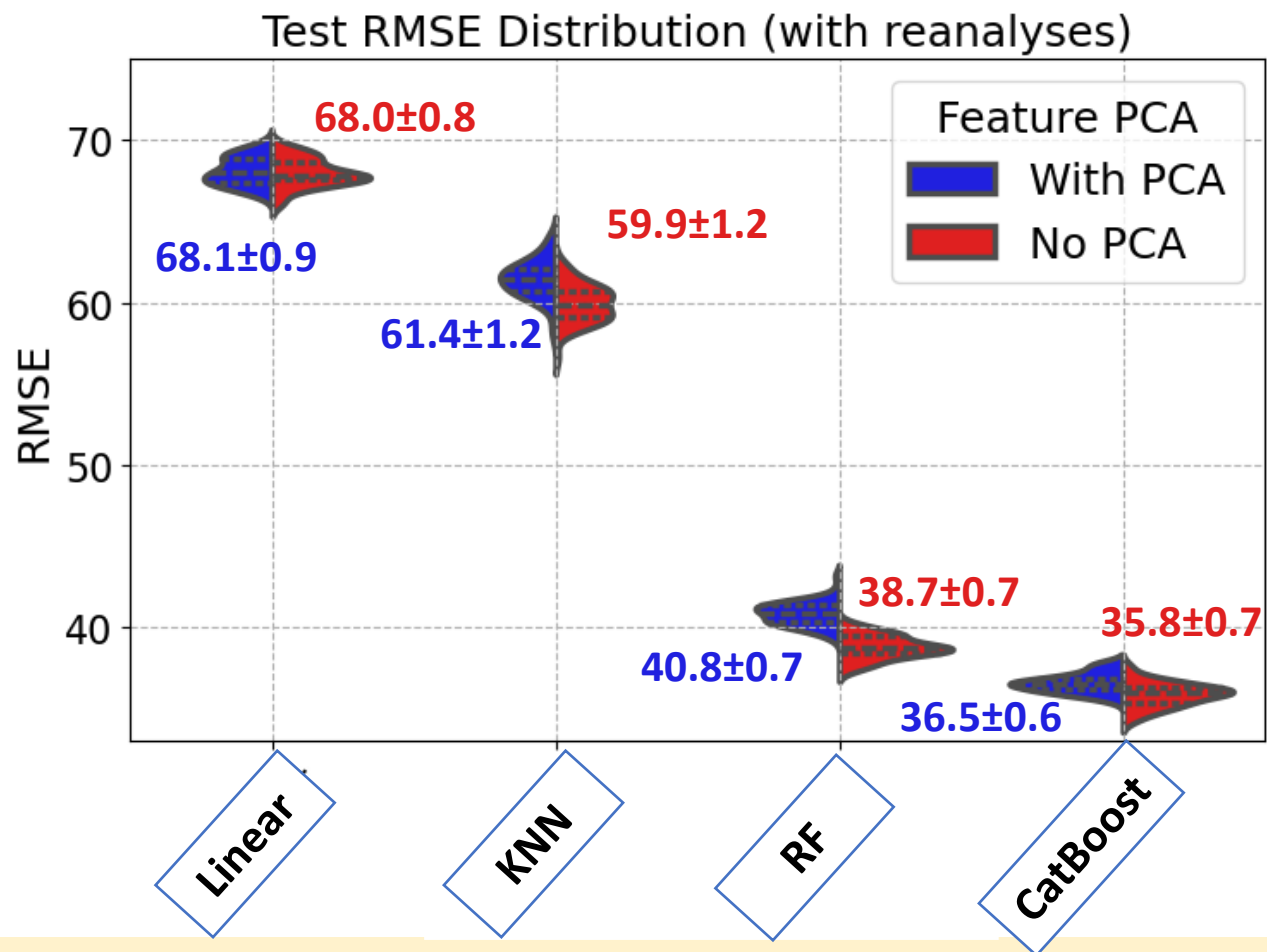
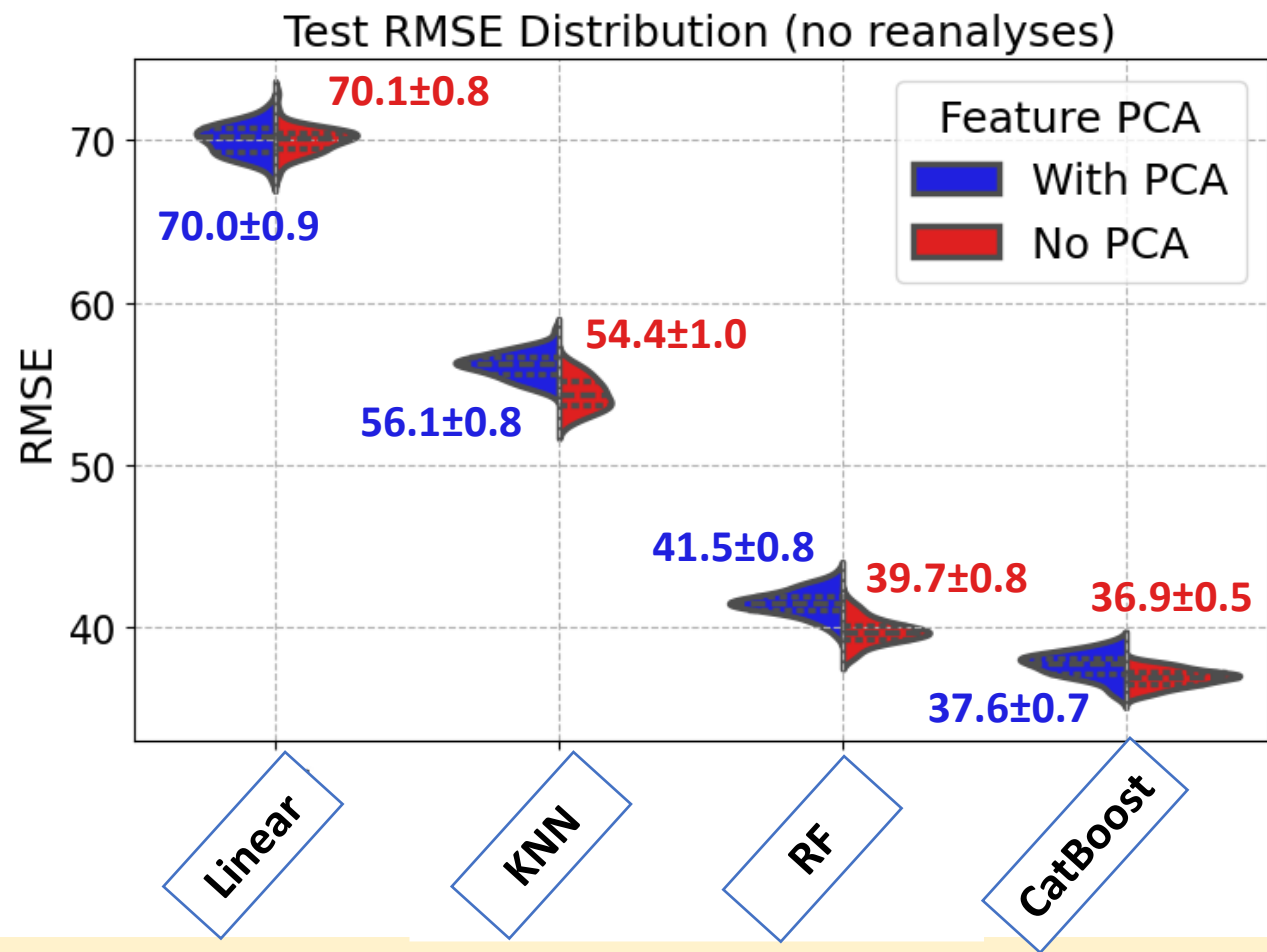
Сравнение различных конфигураций моделей машинного обучения для рассеянной радиации

Split Violin Plot Comparison: With vs Without Feature PCA



Сравнение различных конфигураций моделей машинного обучения для суммарной радиации

Split Violin Plot Comparison: With vs Without Feature PCA



Выводы работы

1. Экспертным способом отобраны физически значимые предикторы для восстановления прямой, рассеянной и суммарной солнечной радиации, подготовлены базы данных;
2. Реализованы 4 различных по архитектуре алгоритмов машинного обучения: *Линейная регрессия*, *K-ближайших соседей (KNN-Regression)*, *Случайный лес (Random Forest)* и *Градиентный бустинг (CatBoost)* с использованием 4-ых баз данных предикторов.
3. Во всех конфигурациях расчётов наилучший результат показан моделью CatBoost. Линейная модель не способна воспроизвести известную нелинейную зависимость прямой радиации от балла облачности при малых её значениях;
4. Ошибки баллов облачности имеют выраженную зависимость от балла облачности и степени нарушения эргодичности облачного покрова;
5. Наибольшую значимость из предикторов имеют комплексный балл облачности и степень нарушения эргодичности облачного покрова;

Обученные локальные модели Random Forest и CatBoost показали удовлетворительный результат качества, что позволяет их использовать в будущем в практических целях

Реализация локальных моделей машинного обучения для аппроксимации коротковолновой радиации у поверхности земли в облачных условиях по геофизическим предикторам для Московского региона

Петров Н.А. (nial.03@mail.ru), Н.Е. Чубарова, М.И. Варенцов

Благодарности: *А.А. Полюхову* и *Д.А. Пискуновой* за создание архивов данных RAD-MSU

Работа выполнена при поддержке Российского научного фонда (проект №25-27-00014)



Хибины, 2023. Автор: Д. Леонова



Атлантика, 2025. Автор: А. Нарижная