

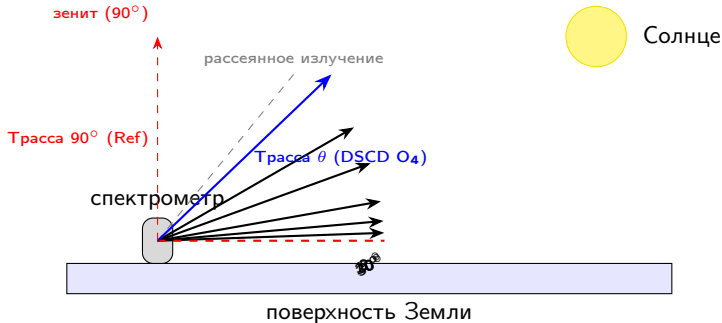
Определение вертикального профиля коэффициента экстинкции аэрозоля в нижней тропосфере по данным спектральных измерений метода MAX-DOAS с применением трансформер-архитектуры глубокого обучения.

Шамсутдинов Д.Р.⁽¹⁾ Постыляков О.В.⁽¹⁾ Боровский А.Н.⁽¹⁾

⁽¹⁾ Институт физики атмосферы имени А.М. Обухова РАН, Москва, Россия

Введение

- Контроль вертикального распределения аэрозоля – важная задача экологического мониторинга
- MAX-DOAS – эффективный метод дистанционного зондирования для измерения содержания малых газовых примесей и характеристик аэрозоля в нижней тропосфере
- Основная сложность – решение задачи восстановления профиля экстинкции аэрозоля и двуокиси азота NO_2
- Рассмотрим один из методов её решения



SCD (Slant Column Density) — наклонное содержание поглощающего газа (O_4 используется как индикатор аэрозольного ослабления). DSCD — разность между наклонным и зенитным направлениями.

Принцип метода MAX-DOAS

- **MAX-DOAS (Multi-Axis Differential Optical Absorption Spectroscopy)** — метод измерения содержания газов в нижней тропосфере.
- Измеряются спектры рассеянного солнечного света при нескольких углах визирования (от горизонта до зенита).
- Для каждого спектра выполняется **DOAS-анализ**:

- 1 Нормализация спектров:

$$I_{\text{norm}}(\lambda) = \frac{I_{\text{meas}}(\lambda)}{I_{\text{ref}}(\lambda)}$$

- 2 Оценивание дифференциальных наклонных толщ (dSCD) по спектрам поглощения газов:

$$\ln \left(\frac{I_{\text{meas}}(\lambda)}{I_{\text{ref}}(\lambda)} \right) = \sum_k \sigma'_k(\lambda) \cdot \text{dSCD}_k + \sum_{m=0}^M a_m \lambda^m$$

где σ'_k — дифференциальные сечения поглощения, dSCD_k — оцениваемая дифференциальная наклонная толщина, $\sum a_m \lambda^m$ — полином для учета широкополосной экстинкции

- 3 Получение дифференциальных наклонных толщ (DSCD):

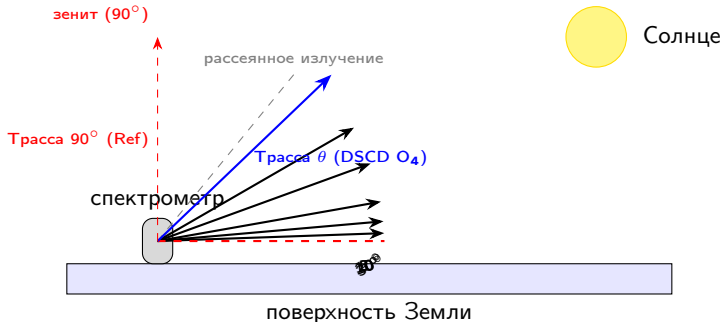
$$\text{DSCD}(\theta) = \text{SCD}(\theta) - \text{SCD}(\theta_{\text{ref}})$$

где $\text{SCD}(\theta) = \int_0^\infty n(z) \cdot m(z, \theta) dz$ — наклонная толщина (k — индекс газа, $m(z, \theta)$ — весовая функция/воздушная масса)

Ключевые особенности:

- Содержание O_4 связано с вертикальным распределением **объемного коэффициента ослабления** аэрозоля
- Вертикальные профили **объемного коэффициента ослабления** аэрозоля восстанавливаются через решение обратной задачи

Схема измерений MAX-DOAS



На различных углах визирования измеряются спектры рассеянного солнечного света. Из них определяются наклонные содержания O_4 ($DSCD(\theta)$) и восстанавливаются профили коэффициента экстинкции аэрозоля $\alpha_{aer}(z)$

Ключевые соотношения

- $DSCD(\theta) = SCD(\theta) - SCD(90^\circ)$
- $AOD(\text{аэрозольная оптическая толщина}) = \int \alpha_{aer}(h) dh$

$DSCD$ – дифференциальная наклонная толщина газа (O_4),
 α_{aer} – коэффициент экстинкции аэрозоля,
 AOD – оптическая толщина аэрозоля на длине волны 440нм.

Уравнение для наклонных содержаний O_4

$$\xi_{O_4}(\theta_i) = \int_0^H m(h, \theta_i, \alpha_{aer}(\cdot)) n_{O_4}(h) dh + \nu_i^{O_4}$$

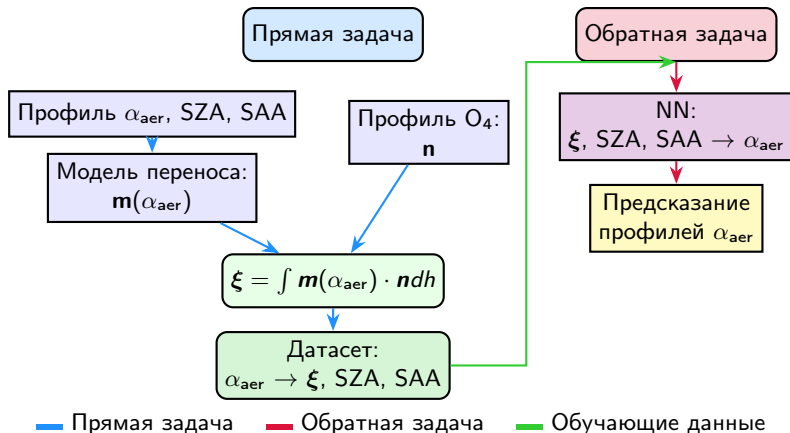
($\xi_{O_4}(\theta_i)$ здесь соответствует измеренному DSCD O_4 при угле θ_i)

- $\xi_{O_4}(\theta_i)$ — измеренные наклонные содержания O_4 при углах визирования θ_i
- m — послойные воздушные массы (весовые функции), которые зависят от геометрии и **профиля аэрозольного ослабления** и рассчитываются через модель переноса излучения
- $n_{O_4}(h)$ — известный вертикальный профиль O_4 (пропорционален квадрату концентрации O_2)
- $\alpha_{aer}(h)$ — **искомый профиль** коэффициента экстинкции аэрозоля (нелинейный параметр)

Подход: По измеренным $\xi_{O_4}(\theta_i)$ восстанавливаем вертикальный профиль экстинкции $\alpha_{aer}(h)$

O_4 — димер кислорода, используется как реперный газ;

Схема решения обратной задачи



Обучение на синтетических данных, сгенерированных моделью переноса на основе реалистичных профилей аэрозоля.

Создание обучающего датасета

Источник профилей PM_{10} :

CAMS European air quality reanalyses,

Модель: SILAM

Область: Звенигород (55.7°N, 36.8°E).

Ограничения: 2018–2023г, тёплый сезон

(май–сент), 06:00–20:00 LT.

Перевод $PM_{10} \rightarrow \alpha_{aer}$ (сроки и осреднение):

Сопоставление **среднесуточных** значений PM_{10} и $AOD_{AERONET}$ за период 2018–2023 гг.

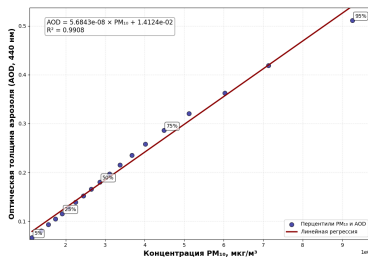
Калибровка по перцентилям AOD (440 нм):

$$\alpha_{aer}(440) = k \cdot PM_{10}(h), \quad k = \frac{AOD_{AERONET}}{\int PM_{10} dh}$$

(МНК по перцентилям, график справа)

Аугментация и расчёт DSCD O_4 :

- Сдвиг даты ± 10 дн., времени ± 2 ч $\rightarrow$ новые SZA/SAА (10 вар./объект)
- Добавлен гауссов шум ($\sigma = 7.24 \cdot 10^{41} \text{ см}^{-2}$) (30 вар./объект)



Калибровка PM_{10} по AOD (перцентилям)

Параметры модели переноса излучения (МПИ)

- Длина волны: 440 нм (согласовано с каналом AERONET)
- Геометрия: SZA, углы визирования, относительный азимут
- Газовый состав: Озон (ТОС = 360 DU, вертикальный профиль Крюгера)
- Аэрозоль: 2-слойная модель (тропосфера до 10 км + стратосфера)
- Поверхность: Ламбертово отражение, альbedo = 0.05

Архитектура Transformer для профилей аэрозоля

Угловая ветвь (обработка DSCD)

- 5 токенов DSCD $O_4 \rightarrow \text{AngleTokenizer}$ (Linear 1 $\rightarrow d_{\text{model}} = 64$)
- Позиционное кодирование (sin/cos)
- TransformerEncoder ($L = 4$, $h = 2$, $d_{\text{ff}} = 128$, GELU, dropout=0.012)
- Mean pooling $\Rightarrow e_{\text{angle}} \in \mathbb{R}^{64}$

Контекстная ветвь (геометрия)

- MLP($4 \rightarrow 64 \rightarrow 64$) с GELU для sin / cos(SZA/SAA)

Fusion

- Конкатенация $[e_{\text{angle}}; e_{\text{context}}]$
- FusionMLP($128 \rightarrow 64 \rightarrow 64$)
- LayerNorm $\rightarrow$ Linear($64 \rightarrow 46$)

Гиперпараметры и обучение

Параметр	Значение
d_{model}	64
n_{heads}	2
n_{layers}	4
d_{ff}	128
dropout	0.012
batch size	1024
lr (max)	$1.20 \cdot 10^{-3}$
weight decay	$6.89 \cdot 10^{-4}$

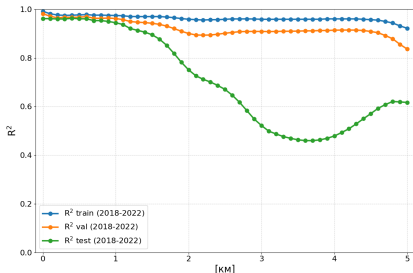
Препроцессинг: RobustScaler(DSCD)

Оптимизатор: AdamW

Планировщик: косинусный
с разогревом

Подбор: Optuna (TPE, 120 trials)

Результаты обучения модели

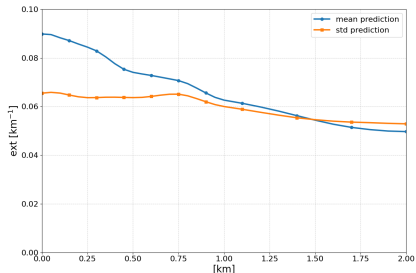


R^2 по высоте для train/val/test-выборок:

0–1 км: $\overline{R^2} = 0.936$

1–2 км: ≈ 0.85

выше 2 км – падение чувствительности



Средний профиль и СКО (до 2 км).

Ключевые метрики

- Высокая точность в нижних 500 м – благодаря сильному сигналу O_4 на малых углах
- Разница R^2 между train и val ≤ 0.05 – отсутствие переобучения
- Out-of-time тест (2023 г.) подтверждает обобщающую способность

Валидация на реальных измерениях

Место измерений: ЗНС

- Звенигородская научная станция (40 км к западу от Москвы).
- Измерения MAX-DOAS в период 2011–2019 гг.
- Входит в сеть AERONET, работает фотометр CIMEL.

Данные для валидации

- Спектрометр Ocean Optics USB4000.
- 124 дня с ясной погодой и углами 87° , 85° , 80° , 70° , 60° .
- Отобраны дни с **одновременными** измерениями MAX-DOAS и CIMEL.

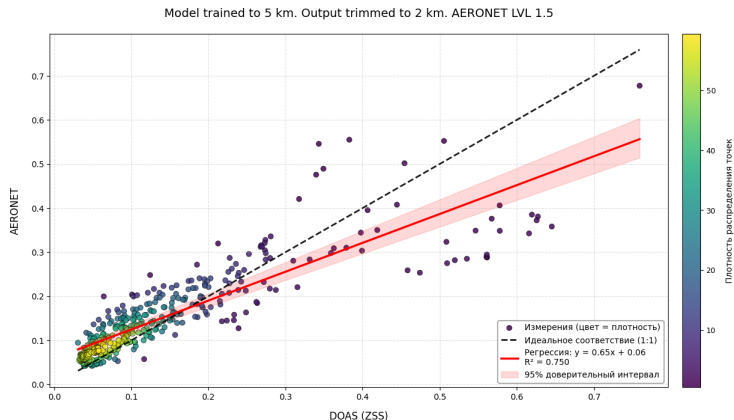


Фото спектрометра Ocean Optics USB4000 на ЗНС

Метод сравнения

- Восстановление профиля $\alpha_{\text{aer}}(h)$ по DSCD O_4 .
- Расчёт $\text{AOD}_{\text{модель}}$ интегрированием профиля.
- Сравнение с $\text{AOD}_{\text{AERONET}}$ (CIMEL).

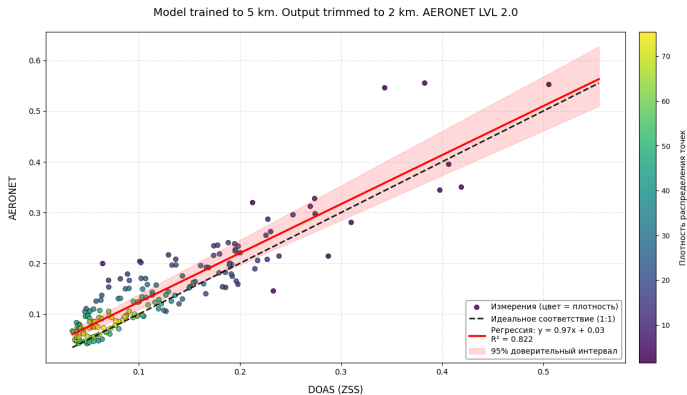
Валидация на реальных измерениях. AERONET LEVEL 1.5



Результат

Коэффициент детерминации $R^2 = 0.750$ между $AOD_{\text{MAX-DOAS}}$ и AOD_{AERONET} .
Level 1.5 — данные с автоматической очисткой от облачности, без пост-калибровки.

Валидация на реальных измерениях. AERONET LEVEL 2.0



Результат

Коэффициент детерминации $R^2 = 0.822$ между $AOD_{\text{MAX-DOAS}}$ и AOD_{AERONET} . Модель, обученная только на синтетических CAMS-данных, успешно работает с реальными измерениями.

Рост R^2 объясняется тем, что Level 2.0 — верифицированные данные: пост-калибровка фотометров и ручная очистка от тонкой облачности и инструментальных шумов.

Восстановление профилей NO₂: Прямая модель и проблема ЛСО

Прямая модель MAX-DOAS для NO₂

$$\xi_{\text{NO}_2}(\theta_i) = \int_0^H m(h, \theta_i, \alpha_{\text{aer}}(\cdot)) \eta_{\text{NO}_2}(h) dh + \nu_i$$

- $m(h, \theta_i, \alpha_{\text{aer}})$ — весовая функция, зависит от профиля аэрозоля
- $\eta_{\text{NO}_2}(h)$ — искомый профиль концентрации NO₂
- ξ_{NO_2} — измеренные DSCD NO₂ при углах θ_i

Специфика вертикального распределения NO₂

- Приземные источники: инверсии → накопление, вторичные максимумы
- Дневная конвекция: перемешивание, один максимум
- Профили не гауссовы, многоэкстремальные

Метод ЛСО (Linear Statistical Estimation)

Постановка задачи:

$$\min_{\mathbf{R}} \mathbb{E} \|\mathbf{R}(\xi) - \mathbf{n}\|_2^2$$

Аналитическое решение:

$$\hat{\mathbf{n}} = \mathbf{n}_0 + \mathbf{FM}^*(\mathbf{FM}^* + \Sigma)^{-1}(\xi - \mathbf{Mn}_0)$$

Ковариационная матрица:

$$F_{ij} \propto \exp\left(-\frac{|h_i - h_j|}{h_0}\right)$$

Проблемы ЛСО

- Субъективный выбор h_0 (25 м – 1 км)
- Гауссово предположение о профилях
- Сглаживание реальных экстремумов

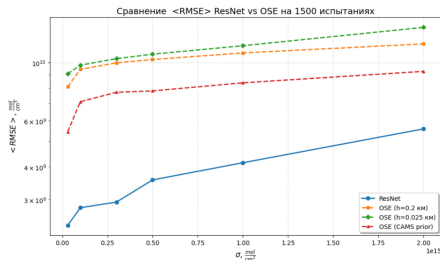
Нейросетевой подход и сравнение с ЛСО

Архитектура ResNet для NO₂

- **Вход:** 9 признаков (5 DSCD NO₂ + sin/cos SZA/SAA)
- **5 остаточных блоков:** LayerNorm → Linear(256) → LeakyReLU → Skip
- **Выход:** 46 высот (0–5 км)

Преимущества НС перед ЛСО

- Не требует априорной статистики n_0 , F , Σ
- Обучается на реалистичных не-гауссовых профилях (CAMS)
- Восстанавливает сложную многомодальную структуру
- Автоматически учитывает нелинейную зависимость $m(\alpha_{aer})$



Методология сравнения

- **Метрика:** $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\hat{n}_i - n_i\|^2}$
- **Шумы:** $0.03 - 2 \cdot 10^{15}$ мол/см²
- **Тестовая выборка:** 30 000 профилей

Результаты на профилях CAMS

- **НС превосходит ЛСО на 20–200%** по RMSE
- Устойчивость при шуме $> 1 \cdot 10^{15}$ мол/см²
- ЛСО — теоретический предел только для нормальных распределений

Основные результаты

- Разработана архитектура нейронной сети для восстановления профилей аэрозольной экстинкции по данным MAX-DOAS
- Точность в приземном слое: $R^2 > 0.93$
- Успешная валидация на независимых натурных данных (AERONET, $R^2 = 0.822$)

Перспективы развития

- Многозадачное обучение: совместное восстановление профилей экстинкции аэрозоля и NO_2
- Масштабирование на другие станции сети MAX-DOAS
- Интеграция с моделями переноса для оперативного мониторинга качества воздуха

Благодарим за внимание!