УДК 520.16, 520.18
DOI: 10.62955/0135-3748-2026-432
ORCID: 0009-0006-8094-0116

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ГРАДИЕНТНОГО БУСТИНГА И СЛУЧАЙНОГО ЛЕСА ДЛЯ ВОССТАНОВЛЕНИЯ ВЫСОТНОГО ПРОФИЛЯ ОПТИЧЕСКОЙ ТУРБУЛЕНТНОСТИ НАД КИСЛОВОДСКОЙ ОБСЕРВАТОРИЕЙ

М.Б. Дрига[1], А.Ю. Шиховцев[1], П.Г. Ковадло[1], Е.А. Копылов[2], С.А. Потанин[2, 3, 4]

[1]Институт солнечно-земной физики Сибирского отделения Российской академии наук, Иркутск, Россия, driga@iszf.irk.ru
[2]Институт Астрономии Российской Академии Наук, Москва, Россия
[3]Московский Государственный Университет им. М.В. Ломоносова, Физический Факультет, Москва, Россия
[4]Государственный Астрономический Институт им. П.К. Штернберга, Москва, Россия

COMPARATIVE ANALYSIS OF GRADIENT BOOSTING AND RANDOM FOREST FOR RECONSTRUCTING THE VERTICAL PROFILE OF OPTICAL TURBULENCE ABOVE THE KISLOVODSK OBSERVATORY

M.B. Driga[1], A.Yu. Shikhovtsev[1], P.G. Kovadlo[1], E.A. Kopylov[2], S.A. Potanin[2, 3, 4]

[1]Institute of Solar-Terrestrial Physics SB RAS, Irkutsk, Russia, driga@iszf.irk.ru
[2]Institute of Astronomy of the Russian Academy of Sciences, Moscow, Russia
[3]Lomonosov Moscow State University, Faculty of Physics, Moscow, Russia
[4]Sternberg Astronomical Institute, Moscow, Russia

Аннотация. Угловое разрешение наземных телескопов в оптическом и ИК-диапазонах ограничено атмосферной оптической турбулентностью (АОТ). Для прогноза качества изображений и проектирования систем адаптивной оптики необходимо знать вертикальное распределение структурной характеристики C_n^2. В работе на основе многолетних измерений на Кисловодской астрономической обсерватории (плато Шаджатмаз) построены модели, связывающие профили $C_n^2 \Delta z$ с метеорологическими данными реанализа ERA-5 и приземными наблюдениями. Исследованы два ансамблевых метода: случайный лес (Random Forest) и градиентный бустинг (XGBoost). Показано, что градиентный бустинг демонстрирует несколько более высокие коэффициенты корреляции и меньшие ошибки на большинстве высотных уровней, однако оба метода дают сопоставимые результаты (средняя корреляция выше 0,73). Анализ важности признаков выявил, что для нижней тропосферы ключевую роль играют приземные C_n^2 и завихренность, а для высот около 11 км — температура и ветер на уровне 200 гПа. Предложенная методика может быть распространена на другие обсерватории Северного Кавказа.

Ключевые слова: атмосферная турбулентность, градиентный бустинг, случайный лес, реанализ ERA-5.

Abstract. Angular resolution of ground-based telescopes in the optical and IR ranges is limited by atmospheric optical turbulence (AOT). To predict image quality and design adaptive optics systems, knowledge of the vertical distribution of the structural characteristic C_n^2 is required. In this paper, models linking $C_n^2 \Delta z$ profiles with meteorological data from the ERA-5 reanalysis and ground-based observations are constructed, using long-term measurements at the Kislovodsk Astronomical Observatory (Shatzhatmaz Plateau). Two ensemble methods are investigated: Random Forest and Gradient Boosting (XGBoost). It is shown that Gradient Boosting demonstrates slightly higher correlation coefficients and lower errors at majority of altitude levels; however, both methods yield comparable results (average correlation above 0.73). Feature importance analysis has revealed that the ground-level C_n^2 and vorticity play a key role for the lower troposphere; for altitudes around 11 km, temperature and wind at the 200 hPa level are the most important. The proposed methodology can be used in other observatories of the North Caucasus.

Keywords: atmospheric turbulence, gradient boosting, random forest, ERA-5 reanalysis

ВВЕДЕНИЕ

Качество астрономических наблюдений с земли существенно зависит от турбулентных флуктуаций показателя преломления воздуха [Shikhovtsev et al., 2023; Shikhovtsev et al., 2025]. Типичное значение seeing (угловое разрешение) на хороших астроплощадках составляет 0,5–1,5 угловых секунд, что на порядок хуже дифракционного предела крупных телескопов [Baki, Basu, 2024]. Особенно остро проблема стоит для дневных наблюдений, когда приземная турбулентность усиливается [Cai et al., 2020; Cuadros et al., 2025].

Для моделирования функции размытия точки и разработки адаптивной оптики необходима информация о вертикальном распределении структурной постоянной C_n^2 [Jellen et al., 2023]. Мировые обсерватории (Параналь, Майданак) ведут рутинный мониторинг АОТ [Kornilov et al., 2007], однако на территории России такие данные доступны лишь для нескольких площадок [Natekin et al., 2013]. Особый интерес представляет Кисловодская обсерватория (гора Шаджатмаз, 2127 м), где с 2007 по 2013 год проводились регулярные измерения профилей C_n^2 методом MASS-DIMM.

Современное развитие методов машинного обучения открывает возможность восстанавливать характеристики турбулентности по косвенным метеорологическим данным, например, реанализу ERA-5 [Pierzyna et al., 2025]. Ранее были предложены нейросетевые подходы для прогноза seeing [Pierzyna et al., 2025]. В настоящей работе мы ставим целью сравнить два популярных ансамблевых алгоритма — градиентный бустинг

(XGBoost) и случайный лес (Random Forest) — при решении задачи регрессии вертикального профиля $C_n^2 \Delta z$ в диапазоне высот 0,5–22,6 км.

ИСХОДНЫЕ ДАННЫЕ И МЕТОДЫ

Измерения проводились на базе 12-дюймового телескопа Meade RCX400 с использованием методики MASS-DIMM [Roddier et al., 1981]. Получены профили $C_n^2 \Delta z$ для 12 высот: 0,5; 0,71; 1,0; 1,41; 2,0; 2,82; 4,0; 5,66; 8,0; 11,3; 16,0; 22,6 км. Входные признаки включали: приземные метеоданные (температура, ветер, влажность), данные реанализа ERA-5 для уровней 750, 500 и 200 гПа (компоненты ветра *u*,*v*,*w*, температура *T*, завихренность *vo*, высота пограничного слоя *blh*, турбулентные и гравитационные напряжения), а также измеренное значение C_n^2 на высоте 0 км. Все данные приведены к часовому разрешению.

Обучены две модели: градиентный бустинг (XGBoost) и случайный лес (RF). Гиперпараметры подбирались с помощью GridSearchCV. Обучающая выборка – 70%, валидационная – 15%, тестовая – 15%. Качество оценивалось по коэффициенту корреляции Пирсона (C), средней абсолютной ошибке (MAE) и среднеквадратичной ошибке (MSE) для десятичных логарифмов $lg(C_n^2 \Delta z)$.

РЕЗУЛЬТАТЫ

В таблице 1 приведены метрики качества для обоих методов на тестовой выборке для каждого высотного слоя. Видно, что градиентный бустинг в большинстве случаев даёт более высокие значения коэффициента корреляции (например, 0,81 против 0,78 на высоте 4 км) и несколько меньшие ошибки MAE и MSE. Исключение составляет высота 1,0 км, где корреляция случайного леса ниже (0,71 против 0,74 у XGBoost). На верхних слоях (8–22 км) преимущество градиентного бустинга также заметно: для 22,6 км $C_{GB} = 0{,}84$ против $C_{RF} = 0{,}79$.

Таблица 1. Метрики качества моделей градиентного бустинга (GB) и случайного леса (RF)

Высота, км	C_{GB}	C_{RF}	MAE_{GB}	MAE_{RF}	MSE_{GB}	MSE_{RF}
0,5	0,77	0,78	0,33	0,33	0,19	0,19
0,71	0,73	0,73	0,43	0,43	0,31	0,31
1,0	0,74	0,71	0,41	0,43	0,28	0,31
1,41	0,79	0,76	0,41	0,45	0,29	0,33
2,0	0,78	0,76	0,34	0,36	0,20	0,22
2,82	0,78	0,76	0,27	0,29	0,13	0,14
4,0	0,81	0,78	0,25	0,27	0,11	0,12
5,66	0,80	0,76	0,23	0,26	0,09	0,11
8,0	0,71	0,68	0,21	0,24	0,08	0,09
11,3	0,73	0,68	0,16	0,19	0,05	0,06
16,0	0,71	0,67	0,16	0,18	0,04	0,05
22,6	0,84	0,79	0,13	0,15	0,03	0,04

Для каждого слоя вычислена важность признаков по методу встроенной важности RF (среднее уменьшение нечистоты). Выявлены следующие закономерности:

— Высота 0,5 км: доминируют $lg(C_n^2(0))$ и завихренность vo на уровне 750 гПа;

— Высота 5,66 км: ключевые факторы — меридиональная скорость ветра v(500) и температура T(500);

— Высота 11,3 км (область струйного течения): наибольший вклад вносят T(200), u(200) и v(200). Для XGBoost распределение важностей оказалось похожим, но с более равномерными весами для верхних слоёв.

ОБСУЖДЕНИЕ

Полученные результаты согласуются с известной физической картиной: в приземном слое турбулентность сильно коррелирует с локальными метеоусловиями, тогда как в свободной атмосфере её генерация связана со сдвигом ветра и температурной стратификацией. Преимущество градиентного бустинга в ряде метрик объясняется его способностью последовательно минимизировать ошибку и лучше адаптироваться к сложным нелинейным зависимостям. Случайный лес, однако, более устойчив к шуму и требует меньше вычислительных ресурсов. Ограничения работы связаны с дискретностью реанализа (30 км) и использованием только трёх изобарических уровней.

ЗАКЛЮЧЕНИЕ

Проведено сравнение двух ансамблевых методов машинного обучения (XGBoost и случайный лес) при решении задачи восстановления вертикального профиля оптической турбулентности над Кисловодской обсерваторией на основе реанализа ERA-5 и приземных измерений. Основные выводы:

— Оба метода показывают высокое качество (корреляция 0,7–0,8), что подтверждает принципиальную возможность оценки $C_n^2(z)$ по метеоданным.

— Градиентный бустинг даёт в целом более высокие коэффициенты корреляции (например, 0,84 на высоте 22,6 км) и меньшие ошибки MAE и MSE по сравнению со случайным лесом, особенно на верхних уровнях.

— Наиболее информативные признаки зависят от высоты: у поверхности — локальная турбулентность и завихренность, в средней тропосфере — ветер и температура, в области струйного течения — параметры уровня 200 гПа.

— Исключение приземных данных ухудшает метрики на 5–8%, но модель остаётся работоспособной.

Разработанная методика может быть применена для других обсерваторий Северного Кавказа, где отсутствуют прямые измерения профилей C_n^2, но доступен реанализ ERA-5.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке гранта Российского научного фонда № 24-72-10043.

СПИСОК ЛИТЕРАТУРЫ

Baki H., Basu S. Estimating high-resolution profiles of wind speeds from a global reanalysis dataset using TabNet. Environmental Data Science. 2024, vol. 3, p. e32. DOI: 10.1017/eds.2024.41.

Cai J., Xu K., Zhu Y., et al. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. Applied Energy. 2020, vol. 262, p. 114566. DOI: 10.1016/j.apenergy.2020.114566.

Cuadros E. T., Almeida A. P., Carlesso F., et al. Assessment of Daytime Sky Quality and Atmospheric Refractive Index (C_n^2) Analysis at Pico dos Dias Observatory. The Astrophysical Journal Supplement Series. 2025, vol. 281, iss. 1, p. 24. DOI: 10.3847/1538-4365/ae0a2a.

Jellen C., Nelson C., Burkhardt J., et al. Hybrid optical turbulence models using machine-learning and local measurements. Appl. Opt. 2023, vol. 62, iss. 18, pp. 4880–4890. DOI: 10.1364/AO.487280.

Kornilov V., Tokovinin A., Shatsky N., et al. B. Combined MASS-DIMM instrument for atmospheric turbulence studies. Monthly Notices of the Royal Astronomical Society. 2007, vol. 382, pp. 1268–1278. DOI: 10.1111/j.1365- 2966.2007.12467.x.

Natekin A., Knoll A. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics. 2013, vol. 7, article e21. DOI: 10.3389/fnbot.2013.00021.

Pierzyna M., Basu S., Saathof R. OTCliM: Generating a Near-Surface Climatology of Optical Turbulence Strength (C_n^2) Using Gradient Boosting. Artificial Intelligence for the Earth Systems. 2025, vol. 4, iss. 2, p. e240076. DOI: 10.1175/AIES-D-24-0076.1.

Roddier F. V The Effects of Atmospheric Turbulence in Optical Astronomy. Progress in Optics / Ed. E. Wolf. Amsterdam: Elsevier, 1981, vol. 19, pp. 281–376. DOI: 10.1016/S0079-6638(08)70204-X.

Shikhovtsev A. Yu., Kovadlo P. G., Kiselev A. V., et al. Application of Neural Networks to Estimation and Prediction of Seeing at the Large Solar Telescope Site. Publications of the Astronomical Society of the Pacific. 2023, vol. 135, iss. 1043, p. 014503. DOI: 10.1088/1538-3873/acb384.

Shikhovtsev A.Yu., Kiselev A.V., Kovadlo P.G. Physically informed neural networks for assessing atmospheric optical turbulence intensity. Izvestiya Vuzov. Fizika [Russ. Phys. J.]. 2025, vol. 68, iss. 5. DOI: 10.17223/00213411/68/5/8.