

«УТВЕРЖДАЮ»

Директор ИОА СО РАН

д.ф.-м.н., член-корр. РАН

И.В. Пташник

16 апреля 2025 года

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного бюджетного учреждения науки Института оптики атмосферы им. В.Е. Зуева Сибирского отделения Российской академии наук на диссертационную работу Борисова Данила Владимировича **«ПОСТПРОЦЕССИНГ ЧИСЛЕННЫХ ПРОГНОЗОВ КОНЦЕНТРАЦИЙ ВЗВЕШЕННЫХ ЧАСТИЦ (PM₁₀) И ПРИЗЕМНОГО ОЗОНА (O₃) С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ»**, представленную на соискание учёной степени кандидата географических наук по специальности 1.6.18 – Науки об атмосфере и климате.

Диссертационная работа Борисова Данила Владимировича посвящена актуальной задаче уточнения результатов численных расчетов химико-транспортных моделей на этапе постпроцессинга расчетов с использованием методов машинного обучения.

Актуальность темы. В настоящее время для исследований процессов формирования и переноса загрязнений в атмосфере активно используются химико-транспортные модели разной степени сложности. Они позволяют прогнозировать распространение загрязняющих веществ, оценивать влияние антропогенных и природных факторов, а также разрабатывать стратегии снижения вредных выбросов. Однако результаты численного моделирования неизбежно содержат погрешности, связанные с упрощением описания физико-химических процессов; ограниченным пространственно-временным разрешением (по времени и пространству); ошибками, содержащимися во входных данных и вычислительными артефактами (численная диффузия, неустойчивость схем). Актуальность данной работы состоит в том, что в ней впервые представлен подход, позволяющий значительно улучшить результаты расчетов численной модели CHIMERE на этапе постпроцессинга с использованием данных измерений автоматических станциях контроля загрязнений атмосферы в Московском регионе. Подход основан на использовании методов машинного обучения, которые позволяют преодолеть ограничения традиционных статистических методов используемых для коррекции результатов.

Целью исследования является разработка на основе методов машинного обучения инструмента постпроцессинга прогнозов химической транспортной модели приземных концентраций озона (O₃) и твердых взвешенных частиц (PM₁₀). Для достижения поставленной цели исследования решены следующие задачи: Проведен анализ данных измерений приземных концентраций озона и PM₁₀ на территории московского региона за последние 5 лет. Изучена сезонно-суточная изменчивость концентраций озона и PM₁₀, влияющих на них факторов, а также условий повышения концентраций озона и PM₁₀ до опасных значений; 2) проведен анализ современных систем прогнозирования качества

воздуха, методов повышения качества прогнозов, зарубежных и отечественных практик применения методов машинного обучения; 3) произведено изучение и освоение методов машинного обучения; 4) определены отклонения модельных прогнозов с горизонтальным шагом 2 км от измеренных данных на станциях концентраций озона и PM₁₀; 5) сформирована исследовательская база данных для разработки моделей машинного обучения; 6) разработана модель машинного обучения для постпроцессинга численных прогнозов приземных концентраций озона (ММО-О₃) и взвешенных частиц (ММО-PM₁₀) на территории московского региона; 7) выполнена верификация разработанных моделей постпроцессинга (ММО-О₃ и ММО-PM₁₀) на независимых выборках с использованием данных измерений концентраций в том же регионе; 8) сформулирована методика построения моделей машинного обучения для постпроцессинга численных прогнозов концентраций для применения в задачах прогнозирования качества воздуха.

Структура и основные результаты работы. Диссертационная работа состоит из введения, пяти глав, заключения, списка используемых источников наименований. Общий объем диссертации составляет 152 страницы; список использованных источников состоит из 129 наименований.

Во **введении** обосновывается актуальность исследования, сформулированы цели и задачи, приведены основные положения, выносимые на защиту, научная новизна и практическая значимость результатов работы.

В **первой главе** проведен аналитический обзор методов численного прогнозирования уровней атмосферного загрязнения и методов повышения точности полученных прогнозов. Обсуждены аспекты применения методов машинного обучения для повышения точности прогнозов качества воздуха на основе стационарных данных. Рассмотрены вопросы, связанные с использованием архивных прогнозов химико-транспортных моделей и данных измерений для методов машинного обучения.

Во **второй главе** описаны методы и данные, использованные в исследовании. В качестве основного инструмента машинного обучения выбраны искусственные нейронные сети. Рассмотрены их архитектура, принципы обучения, а также методы оптимизации. Проработаны вопросы, связанные с формированием массивов данных необходимых для обучения искусственной нейронной сети. Проведен анализ результатов измерений с целью оценки сезонных и суточных вариации концентраций O₃ и PM₁₀. Проанализированы основные факторы, влияющие на динамику концентрации исследуемых веществ. Получены оценки качества прогноза озона и PM₁₀ с использованием метеорологической модели COSMO-Ru и XTM CHIMERE.

В **третьей главе** сформулированы основные этапы разработанной методики построения моделей машинного обучения для постпроцессинга численных прогнозов концентраций загрязняющих веществ. Описана реализация методики в виде разработки моделей машинного обучения для постпроцессинга численных прогнозов концентраций приземного озона и PM₁₀ на территории московского региона. Проведен анализ возможных реализаций и подходов. Выбран оптимальный набор входных параметров для обучения нейронной сети. Сформированы необходимые базы данных для каждого измерительного пункта и исследуемого вещества. Проведены численные эксперименты для выбора оптимальной конфигурации нейронной сети на основе блочного подхода. Получены оптимально настроенные нейронные сети и проведен анализ их чувствительности к используемым предикторам.

В четвертой главе проведена верификация постпроцессинга основанного на модели машинного обучения для приземной концентрации озона (ММО-ОЗ) на независимой выборке (март–сентябрь 2024 г.). На основе проведенного анализа для 8-часовых максимумов концентрации озона показано уменьшение среднеквадратической ошибки прогноза в среднем на 30% (с 27 до 19 мкг/м³), с максимальным эффектом в июне (–43%). В рассмотренных случаях повышенной концентрации приземного озона (май, август) ММО-ОЗ успешно компенсировала занижение прогнозов ХТМ, сократив отклонения прогноза от данных измерений примерно в 2 раза. Используемая в работе методология с исключением станций из обучения подтвердила возможность применения ММО-ОЗ для территорий, на которых не проводятся измерения озона, демонстрируя корректную подстройку полей концентраций с учетом городской застройки.

В пятой главе приведены результаты верификации постпроцессинга основанного на модели машинного обучения для прогнозов РМ10 (ММО-РМ10). Анализ, выполненный для среднесуточных значений концентрации аэрозольных частиц РМ10 на основе данных независимой выборки (март–сентябрь 2024 г.) показал снижение среднеквадратичной ошибки прогноза на 40% - с 25 до 15 мкг/м³. С максимальным эффектом в июнь–июль, (60–62%). Применение постпроцессинга ММО-РМ10 успешно корректировало систематическое завышение прогнозов ХТМ CHIMERE, особенно в марте снизив ошибку прогноза на 48%. На городских станциях СКО_{п-н} (прогноз-измерение) уменьшилось в 1.5–2 раза (до 10–15 мкг/м³), однако для станций вблизи автомагистралей (например, МКАД 52) и в зонах действия локальных источников (строительные работы в районе станции Долгопрудный) эффект коррекции был минимален. В рассмотренных эпизодах повышенных концентраций аэрозоля РМ10 связанных с неблагоприятными метеоусловиями, применение постпроцессинга ММО-РМ10 позволило значительно улучшить прогноз и приблизить его измеренным значениям.

В заключении приведены основные результаты и сформулированы выводы диссертационной работы.

К наиболее важным результаты работы относятся:

1) Разработана и применена методика построения моделей машинного обучения для постпроцессинга численных прогнозов приземных концентраций загрязняющих веществ. ММО обучаются для каждого отдельного загрязнителя.

2) Разработанная методика реализована в виде построения моделей машинного обучения для постпроцессинга численных прогнозов приземных концентраций озона (ММО-ОЗ) и твердых взвешенных частиц диаметром менее 10 мкм (ММО-РМ10) на территории московского региона. Обучающие выборки обеих ММО охватывают 5-летний период прогнозов-измерений февраль 2019 г. – февраль 2024 г.

3) Разработанные методики ММО-ОЗ и ММО-РМ10 применимы для постпроцессинга численных прогнозов концентраций в периоды повышенного озонowego и/или аэрозольного загрязнения при неблагоприятных для очищения воздуха метеоусловиях.

4) Показана применимость разработанных ММО-ОЗ и ММО-РМ10 для постпроцессинга численных прогнозов приземных концентраций по всему расчетному домену ХТМ, в том числе в ячейках модельной сетки, в которых отсутствуют измерения концентраций.

Научная новизна полученных результатов состоит в следующем: 1) Разработана методика построения моделей машинного обучения для постпроцессинга численных

прогнозов концентраций загрязняющих веществ; 2) Разработаны и протестированы модели машинного обучения для постпроцессинга численных прогнозов концентраций приземного озона и взвешенных частиц PM10 на территории московского региона 3) Разработанные модели машинного обучения применимы для постпроцессинга численных прогнозов ОЗ и PM10 на территориях, где не проводятся измерения концентраций.

Практическая значимость работы определяется тем, что реализованные по разработанной методике модели машинного обучения позволяют улучшать качество рассчитанных химической транспортной моделью прогнозов приземных концентраций ОЗ и PM10, в том числе на территориях, где отсутствуют измерения концентраций. Разработанные модели машинного обучения могут применяться для постпроцессинга численных часовых прогнозов приземных концентраций ОЗ и PM10 на модельной сетке с горизонтальным шагом $2 \text{ км} \times 2 \text{ км}$ для территории московского региона.

Обоснованность и достоверность полученных результатов работы подтверждается использованием физических и математических обоснованных методов исследования. Представленные результаты находятся в соответствии с теоретическими и экспериментальными результатами аналогичных исследований других авторов.

Положения, выносимые на защиту, находят развернутое и аргументированное подтверждение в тексте диссертационной работы, а также в 28-и публикациях: 10 публикаций в журналах рекомендованных ВАК, 4 свидетельства Роспатента о государственной регистрации программы для ЭВМ. Результаты работы неоднократно докладывались на российских и международных конференциях по профилю диссертации.

Личный вклад автора состоит в том, что представленные в диссертационной работе методика, модели машинного обучения, программный комплекс полностью спроектированы, реализованы и протестированы лично автором. Все результаты исследования получены, проанализированы и описаны автором лично или в соавторстве с научным руководителем.

Автореферат полностью отражает содержание диссертации, её основные результаты и выводы.

К недостаткам работы следует отнести следующее:

1. В работе упущено описание процедуры предварительной обработки и фильтрации данных измерений, полученных на автоматизированных измерительных станциях.

2. Недостаточно обоснован выбор метрик оценки качества модели машинного обучения (средняя ошибка, среднеквадратичная ошибка и коэффициент корреляции). В чем заключается их преимущества перед альтернативными показателями и соответствия решаемой задаче. Особую озабоченность вызывает методика усреднения метрик при блочной кросс-валидации. В частности из текста не ясно, проводилось ли корректное объединение среднеквадратичных ошибок через усреднение квадратов ошибок перед извлечением квадратного корня, а также, не применялось ли и z-преобразование Фишера для коэффициентов корреляции, что может исказить итоговые оценки.

3. Наблюдается некоторое методическое несоответствие между данными использовавшимися для обучения и верификации ММО-постпроцессинга: искусственная нейронная сеть обучалась на часовых прогнозах с суточной заблаговременностью, тогда оценка точности проводилась для 8-часовых максимумов озона и среднесуточных концентраций PM10. Такой подход требует дополнительного обоснования, поскольку прямое обучение на агрегированных показателях могло бы повысить точность прогнозирования именно тех величин, которые использовались для верификации. Автору следует пояснить

причины выбора часовых данных для обучения, включая анализ возможных преимуществ (например, большой объем обучающей выборки, учет суточной динамики) и недостатков (необходимость последующей агрегации с потенциальной потерей точности).

4. В работе отсутствует описание языка программирования, на котором были написаны программы постпроцессинга и какие библиотеки были использованы.

Перечисленные недостатки не затрагивают основных результатов исследования и не снижают общую положительную оценку работы.

Заключение о соответствии диссертации критериям, установленным Положением о порядке присуждения ученых степеней

Диссертация Борисова Д. В. является законченной научно-квалификационной работой в актуальном направлении исследований. Автором работы проведено комплексное исследование методов улучшения качества прогнозов химико-транспортных моделей основанных на постпроцессинге. Диссертация Борисова Д. В. Соответствует паспорту специальности 1.6.18 - Науки об атмосфере и климате (пункт 2 - вычислительные методы и технологии систем анализа и усвоения данных наблюдений, пункт 3 - озон и другие малые примеси в атмосфере, загрязнение атмосферы, пункт - 14 - антропогенное загрязнение атмосферы, перенос примесей оценка концентраций и прогноз загрязнения).

Таким образом, диссертация Борисова Д. В. удовлетворяет требованиям пунктов 9-11, 13,14 Положения о присуждении учёных степеней, утверждённого Постановлением Российской Федерации от 24 сентября 2013 года №842 (в ред. от 25.01.2024), а Данил Владимирович Борисов заслуживает присуждения искомой степени кандидата географических наук по специальности 1.6.18 — Науки об атмосфере и климате.

Отзыв составлен доктором физико-математических наук, профессором М.В. Панченко.

Отзыв на диссертацию обсужден и одобрен на заседании секции Ученого совета отделения «Радиационные составляющие климата» ИОА СО РАН 15.04.2025 г., протокол №1.

Председатель СУС отделения РСК

ИОА СО РАН, д.ф.-м.н., профессор

Адрес: 634055, г.Томск, пл.Академика Зуева,1

ИОА СО РАН. Тел: (3822)492-050

E-mail: pmv@iao.ru

М.В.Панченко

Секретарь СУС отделения РСК

ИОА СО РАН, к.ф.-м.н., с.н.с.

Н.Н.Щелканов

Подпись М.В. Панченко и

Н.Н. Щелканова заверяю

Ученый секретарь ИОА СО РАН, к.ф.-м.н.

Т.Е.Кураева