

DOI: <https://doi.org/10.37162/2618-9631-2023-4-86-104>

УДК 504.3.054+004.85+504.064

Постпроцессинг численных прогнозов концентраций приземного озона с использованием машинного обучения

Д.В. Борисов, И.Н. Кузнецова

*Гидрометеорологический научно-исследовательский центр
Российской Федерации, г. Москва, Россия;
lbj23.98@mail.ru*

Мониторинг качества прогнозирования концентраций загрязняющих веществ на основе химических транспортных моделей (ХТМ) в действующей в Гидрометцентре России технологии указывает на целесообразность применения постпроцессинга. Впервые для коррекции модельных расчетов концентрации приземного озона использованы искусственные нейронные сети (ИНС). Обучающая выборка сформирована из ретроспективных часовых прогнозов ХТМ CHIMERE в период 2019–2023 гг. Проведены эксперименты по подбору оптимальной конфигурации ИНС. Представляются результаты экспериментального тестирования лучшей ИНС на недельном периоде летом с эпизодом высоких концентраций озона и в весеннем эпизоде повышенного содержания озона вследствие активного тропосферного перемешивания. Показана эффективность применения ИНС для улучшения модельного прогноза приземного озона и его суточной динамики.

Ключевые слова: искусственные нейронные сети, приземный озон, численный прогноз загрязнения, химическая транспортная модель, CHIMERE

Post-processing of ground-level ozone numerical forecasts using machine learning

D. V. Borisov, I. N. Kuznetsova

*Hydrometeorological Research Center of Russian Federation, Moscow, Russia
lbj23.98@mail.ru*

Monitoring the quality of air pollutant concentration forecasts based on chemical transport models (CTMs) currently used in the technology of the Hydrometcentre of Russia indicates the feasibility of the post-processing procedure application. For the first time, artificial neural networks (ANNs) were used to correct ground-level ozone model calculations. Retrospective hourly CTM CHIMERE forecasts for 2019-2023 formed the training dataset. Experiments were carried out to select the optimal ANN settings. Results of the experimental testing of the best ANN on a week-long summer period with an episode of high ozone concentrations and a spring period with an episode of high ozone concentrations due to active tropospheric mixing are presented. The effectiveness of using ANNs to improve model forecasts of ground-level ozone and its daily dynamics is shown.

Keywords: artificial neural networks, ground-level ozone, numerical forecast of pollution, chemical transport model, CHIMERE

Введение

По данным Всемирной организации здравоохранения, озон (O_3) является одним из основных загрязняющих воздух веществ, которые представляют угрозу для здоровья населения (<https://www.who.int/publications/i/item/9789240034228>); содержание озона в приземном воздухе – приоритетный показатель при оценке качества воздуха [3].

В ФГБУ «Гидрометцентр России» разработана и реализована технология численного прогнозирования качества атмосферного воздуха на основе химической транспортной модели (ХТМ) CHIMERE. Расчеты концентраций загрязнителей воздуха CHIMERE проводятся ежедневно на 48 часов вперед с временным шагом 1 час в узлах регулярной сетки разрешением 2×2 км, покрывающей территорию московского региона. Для расчетов в технологии используется оперативный прогноз модели COSMO-Ru численного краткосрочного прогноза погоды (конфигурация COSMO-Ru2ETR) [8]. Для задания антропогенных эмиссий CHIMERE использует адаптированные региональные данные выбросов загрязняющих веществ в атмосферу кадастра ЕМЕП (European Monitoring and Evaluation Programme, <https://www.emep.int/>).

Регулярный мониторинг качества модельных расчетов CHIMERE проводится по данным автоматизированных измерений концентраций загрязнителей на более чем 50 автоматических станциях контроля загрязнения атмосферы (АСКЗА) ГПБУ «Мосэкомониторинг» в Москве (<https://mosecom.mos.ru/>). Оценки успешности модельного прогноза концентраций приземного озона показывают в среднем удовлетворительное качество прогнозов максимальной и средней суточной концентрации озона. Вместе с тем выявлена сезонная особенность модельных отклонений – завышение концентрации озона летом, в т. ч. в «озоновых» эпизодах, когда наблюдаются самые высокие уровни озона [4, 7].

Успешность модельного расчета концентрации приземного озона во многом зависит от точности данных об эмиссиях предшественников. Как показал выполненный автором анализ, в кадастре ЕМЕП в последние годы отмечаются значительные изменения отраслевой структуры и объемов выбросов отдельных ингредиентов, что, безусловно, вписывается в наблюдаемую реальную динамику выбросов загрязняющих веществ в атмосферу [5]. И для приближения модельных расчетов к наблюдаемым концентрациям требуется регулярная актуализация начальных данных о выбросах на этапе препроцессинга. Заметим, что данные ЕМЕП обновляются с задержкой 2 года; в настоящее время доступны данные 2020 г. (<https://www.ceip.at/the-emep-grid/gridded-emissions>).

Другим способом повышения успешности прогнозов приземного озона является статистическая коррекция расчетов CHIMERE на этапе постпроцессинга.

В данном исследовании предлагается метод коррекции прогнозов концентраций приземного озона ХТМ SHIMERE путем обучения и применения искусственной нейронной сети (ИНС). Этот метод в настоящее время активно применяется в различных областях науки и практической деятельности. Аппарат нейронных сетей применительно к прогнозированию приземного озона одними из первых в нашей стране был апробирован сотрудниками Института оптики атмосферы и океана [2]. Авторы использовали многослойный персептрон с тремя скрытыми слоями. Установлено, что нейронная сеть способна описать 70 % дисперсии среднего и 49 % дисперсии среднеквадратического отклонения среднесуточной приземной концентрации озона.

В целом ряде зарубежных научных публикаций последних лет приводятся результаты применения методов машинного обучения для прогнозирования и коррекции прогнозов качества воздуха; наибольшую эффективность показывают искусственные нейронные сети, в особенности глубокие нейронные сети [10–12].

Методы и данные

Искусственная нейронная сеть (ИНС) – математическая модель, возникающая из попыток смоделировать обработку информации человеческим мозгом. Основным преимуществом ИНС является возможность и способность накапливать, обобщать большие объемы информации, т. е. обучаться. Базовыми составляющими ИНС являются искусственные нейроны, которые образуют слои.

На вход ИНС подается набор предикторов. Нейроны входного слоя ИНС обрабатывают набор предикторов так, что значение на выходе каждого нейрона представляет собой взвешенную линейную комбинацию входных предикторов; нейроны скрытых слоев дают на выходе взвешенные линейные комбинации выходных значений нейронов предыдущего слоя. Функция активации масштабирует выходные сигналы нейронов, обычно в интервалах $[0, 1]$ или $[-1, 1]$. Прогнозом ИНС является выходное значение нейрона последнего слоя. В случае задачи регрессии на выходном слое всегда задан один нейрон. На рис. 1 схематично представлен процесс обработки информации искусственной нейронной сетью.

Основными параметрами ИНС, от которых зависит точность прогноза, являются синаптические связи между нейронами, или веса ИНС (w на рис. 1). Весь процесс вывода прогноза ИНС с любым количеством слоев и нейронов из набора предикторов можно формализовать как

$$Y = f(X, W, b),$$

где Y – целевая переменная, прогноз ИНС; X – набор входных данных – предикторов; W – набор весов синаптических связей в слоях ИНС; b – смещения в слоях ИНС.

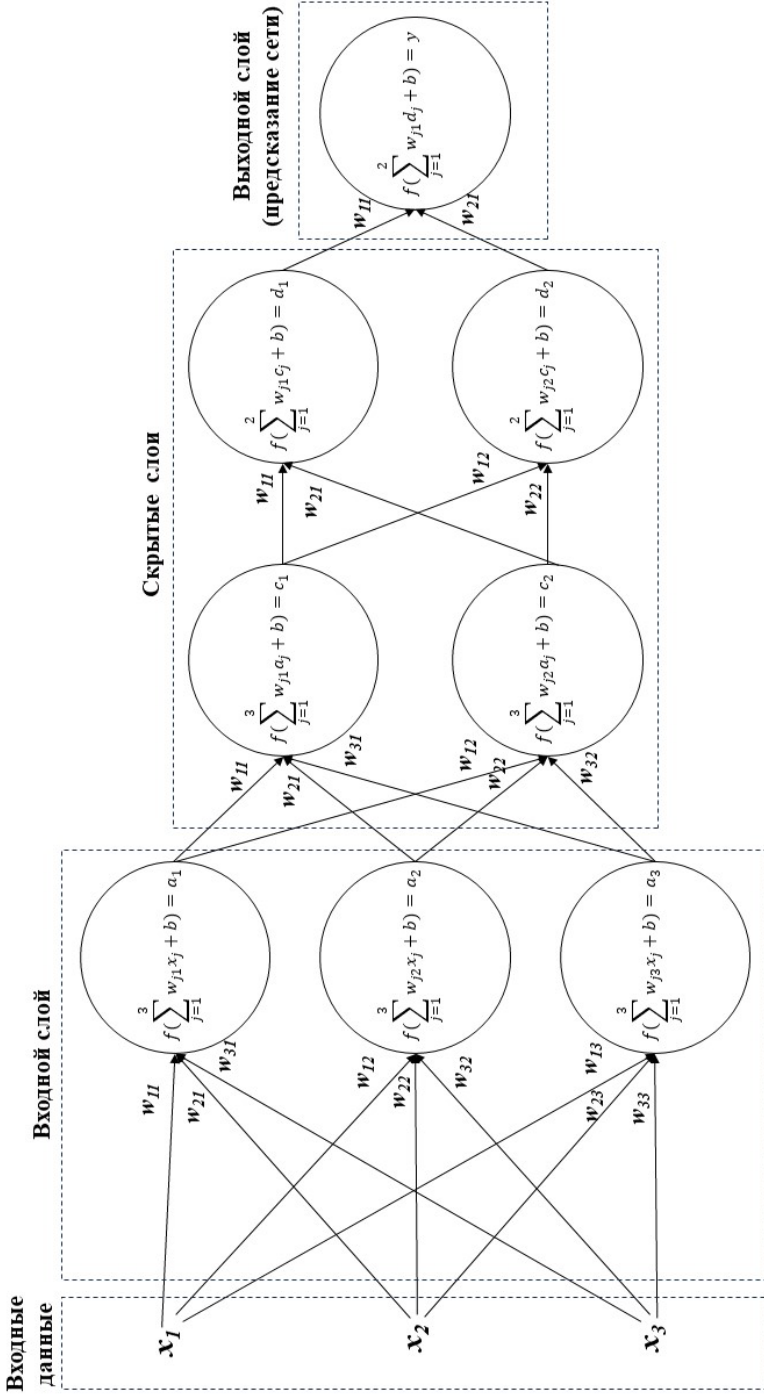


Рис. 1. Схематичное изображение процесса обработки данных искусственной нейронной сетью. Кружками обозначены нейроны, стрелками – синтаптические связи между нейронами, или веса, w .
 Fig. 1. Schematic representation of the data processing carried out by an artificial neural network. Circles indicate neurons, arrows indicate synaptic connections between neurons, or weights, w .

Обучение ИНС заключается в оптимизации весов W с целью минимизации функции потерь. Для обучения необходима обучающая выборка формы (n записей \times m предикторов) и массив истинных значений целевой переменной формы (n записей).

Каждый шаг обучения состоит из нескольких этапов:

1) на вход ИНС подается *пакет данных* – подвыборка из обучающей выборки. Перед обучением задается размер пакета, обучающая выборка разбивается на $n = \frac{\text{размер обучающей выборки}}{\text{размер пакета}}$ пакетов данных, соответственно, обучение проходит в n шагов;

2) на выходе ИНС вычисляется функция потерь L , например, среднее абсолютное отклонение прогнозов ИНС от истинных значений целевой переменной;

3) вычисляется градиент функции потерь по весам ИНС $\frac{\partial L}{\partial w}$;

4) веса ИНС обновляются в сторону снижения функции потерь по правилу:

$$W_t = W_{t-1} - \eta \cdot \frac{\partial L}{\partial w},$$

где W_t – обновленные веса; W_{t-1} – веса на предыдущем шаге обучения; η – шаг обучения, или скорость обучения.

Проход всего обучающего набора данных через ИНС называется *эпохой обучения*.

Формирование обучающей выборки для коррекции прогнозов озона ХТМ CHIMERE производилось из доступных авторам ежедневных почасовых прогнозов ХТМ CHIMERE для территории московского региона на двое суток (48 часов) вперед за период 02 февраля 2019 года – 25 сентября 2023 года, что составляет ~ 5 лет прогнозов.

Вместе с прогнозами концентраций загрязнителей в обучающую выборку на той же регулярной сетке включены данные препроцессинга (прогнозы метеопараметров, данные об эмиссиях и др.), которые использует CHIMERE при построении прогнозов.

В обучающую выборку вошли 34848 часовых прогнозов CHIMERE на сетке разрешением 2×2 км за период 02.07.2019 – 25.09.2023; по техническим причинам отсутствуют прогнозы CHIMERE за период с 22.05 по 30.09.2021 – за теплый период, когда зачастую наблюдаются повышенные и высокие концентрации приземного озона (т.н. «озоновые эпизоды»). Также имеются длительные пропуски прогнозов во второй половине декабря 2019 г. и с середины сентября по середину октября 2022 г. Таким образом, в обучающей выборке содержатся модельные часовые концентрации приземного озона во все сезоны. Для формирования массива целевой переменной в выборку включены часовые данные измерений концентраций приземного озона на АСКЗА в г. Москве. В рассматриваемый период

измерения концентрации приземного озона проводились на 17 АСКЗА, расположенных в жилых и примагистральных районах Москвы (<https://mosecom.mos.ru/karta/>). В базу данных за ~5 лет внесены 516009 почасовых измерений концентраций приземного озона на АСКЗА различной функциональной принадлежности.

Список предикторов (или признаков) обучающей выборки составлялся из имеющихся представлений о физических связях озона с предшественниками в урбанизированной среде, зависимости изменчивости концентрации приземного озона от метеорологических факторов, включая тропосферный обмен [1, 3, 4, 6, 7, 9].

Список предикторов включает в себя следующие величины:

- Прогностические часовые концентрации O_3 и предшественников NO , NO_2 , по расчетам CHIMERE.

- Из данных препроцессинга – часовые прогностические значения метеорологических характеристик системы COSMO-Ru на сетке CHIMERE: температура на высоте 2 м, высота пограничного слоя, относительная влажность воздуха, осадки, скорость ветра на высоте 10 м, коротковолновая радиация, температурный градиент в слоях 2–200 м и 2–700 м, скорость ветра на высоте 700 м. Температурные градиенты и высотный ветер вычислялись из значений метеопараметров на соответствующих высотных уровнях сетки CHIMERE.

- Среднее многолетнее значение общего содержания озона (ОСО) в атмосферном столбе за каждый календарный день (предоставлены Центральной аэрологической обсерваторией), тип АСКЗА, месяц, день недели, час прогноза.

Обучающая выборка включает в себя 18 описанных выше предикторов. Часовые данные измерений озона на АСКЗА синхронизированы по времени с предикторами в ближайшем к координатам каждой АСКЗА узле модельной сетки, выбранными из данных препроцессинга (метеорологические параметры) и прогнозов CHIMERE.

Массив целевой переменной (Y) формировался как отклонение прогноза концентрации озона CHIMERE от соответствующих измерений на АСКЗА. После синхронизации прогнозов CHIMERE и данных измерений на АСКЗА объем обучающей выборки после удаления пропусков составил 428733 записи, которые охватывают все сезоны в ~5-летний период 02.07.2019 – 25.09.2023.

Следует отметить, что разработанный алгоритм формирования обучающих выборок, включая сбор, обработку и синхронизацию прогнозов CHIMERE и данных измерений концентраций на АСКЗА, добавление данных из внешних файлов, конструирование признаков и удаление некорректных данных (многократное превышение средних уровней, отрицательные значения концентраций озона), может использоваться для формирования любых обучающих выборок, а также для пополнения обучающих выборок

новыми данными. Алгоритм разработан на языке программирования Python с использованием библиотек Pandas, Numpy, netCDF4.

Также разработан отдельный алгоритм для проведения экспериментов по обучению и тестированию ИНС. Алгоритм разрабатывался с целью подбора такой конфигурации ИНС, которая показала бы лучшую точность и устойчивость при решении основной задачи исследования – уточнения численных прогнозов приземного озона CHIMERE.

Разработка алгоритма велась по принципам универсальности и автоматизации; он базируется на применении Python-функций, которые создают и обучают ИНС по заданным конфигурациям отдельно или в цикле. Этот алгоритм позволяет создавать ИНС с любым количеством полносвязных слоев и нейронов, добавлять в слои исключения (dropout), задавать любые схемы затухания скорости обучения, задавать импульс обновления весов, свободно задавать настройки процесса обучения – размер пакета, количество эпох обучения и др. При разработке алгоритма на языке программирования Python использовались библиотеки машинного обучения Keras и Tensorflow.

Эксперименты по обучению ИНС

В табл. 1 представлен перечень конфигураций ИНС, обученных в ходе проведения экспериментов по установлению лучших ИНС. Отметим, что эксперименты 1–3 были направлены на обучение различных архитектур ИНС; в экспериментах 4–7 опробованы дополнительные методы повышения эффективности процесса обучения – затухание скорости обучения, задание импульса обновления весов. В экспериментах 8–9 опробован метод исключения нейронов (англ. dropout) в слоях ИНС; в экспериментах 10–12 задана комбинация затухания скорости обучения и импульса обновления весов при различных архитектурах ИНС.

Алгоритм проводит обучение всех ИНС по принципу перекрестной проверки, с установкой количества блоков k . Перекрестная проверка подразумевает разбиение обучающей выборки на k блоков и обучение одной и той же вариации ИНС k раз. Каждый раз ИНС обучается на $k-1$ блоках данных, на оставшемся контрольном блоке, т. е. на независимой выборке, вычисляется оценка точности ИНС, например, среднее абсолютное отклонение прогнозов ИНС от истинных значений целевой переменной. Оценки точности вычисляются после каждой эпохи обучения.

В завершении оценки точности по эпохам осредняются по k блокам, таким образом сглаживаются флуктуации оценок точности ИНС на разных данных и повышается репрезентативность оценок. По осредненным оценкам строятся т. н. кривые обучения – графики точности ИНС на каждой эпохе обучения (см. рис. 2), по которым можно оценить эффективность обучения ИНС.

Таблица 1. Протокол экспериментов по обучению ИНС
Table 1. Protocol of ANN training experiments

№ эксперимента	Архитектура в формате (кол-во нейронов в слое 1, кол-во нейронов в слое 2 ...)	Исключение (доля исключенных нейронов)	Начальная скорость обучения	Затухание скорости обучения	Коэффициент импульса обновления весов	Кол-во весов ИНС
1	(10,1)	-	0.001	-	-	621
2	(100, 1)	-	0.001	-	-	6201
3	(100, 50, 1)	-	0.001	-	-	11201
4	(100, 50, 1)	-	0.1	Ступенчатое	-	11201
5	(100, 50, 1)	-	0.1	Линейное	-	11201
6	(100, 50, 1)	-	0.1	Полиномиальное	-	11201
7	(100, 50, 1)	-	0.001	-	0,9	11201
8	(200, 100, 50, 1)	Вход – 0.2 Слой 1 – 0.5 Слой 2 – 0.5 Слой 3 – 0.5	0.001	-	0,9	37401
9	(400, 200, 100,50, 20, 1)	Вход – 0.2 Слой 1 – 0.5 Слой 2 – 0.5 Слой 3 – 0.5 Слой 4 – 0.5 Слой 5 – 0.5	0.1	Линейное	0,99	130791
10	(100, 50, 1)	-	0.1	Линейное	0,9	11201
11	(100, 50, 25, 1)	-	0.1	Линейное	0,9	12451
12	(100, 50, 1)	-	0.1	Линейное*	0,9	11201

Тестирование обученных ИНС на независимых выборках

Ключевым этапом проведения экспериментов является тестирование обученных ИНС на независимых выборках. Независимые выборки представляют собой временные ряды данных в определенном пункте – на АСКЗА, что позволяет проводить сравнение качества первичных и скорректированных ИНС прогнозов CHIMERE с измеренными на АСКЗА концентрациями приземного озона.

Блок тестирования включает в себя:

1. Задание набора тестовых подвыборок вида «временной период – АСКЗА».

2. Вырезка подвыборок из обучающей выборки, т.к. тестовые данные никак не должны участвовать в обучении ИНС.

3. Прогон тестовых данных через обученные ИНС и получение рядов скорректированных прогнозов.

4. Расчет статистических показателей качества первичных модельных и скорректированных ИНС прогнозов озона. Показатели успешности применения ИНС определяются как по отдельным подвыборкам вида «временной период – АСКЗА», так и осредняются по всей выборке, по заданным временным периодам, по типовым АСКЗА.

5. Визуализация результатов тестирования.

Наибольший интерес представляет установление эффективности коррекции ИНС в периоды высоких концентраций озона, которые весной связаны чаще всего со значительным притоком озона из верхних слоев атмосферы, а летом образуются из-за активной фотохимической генерации озона в жаркую погоду в загрязненном городском воздухе [3, 4, 9]. Также отдельный интерес представляют изменения конфигураций ИНС в условиях различной антропогенной нагрузки, что определило задачу оценки результатов отдельно по АСКЗА транспортного, жилого, смешанного и условно фонового типа.

Общий объем тестовой (независимой) выборки составил 3282 записи данных; из них сформировано 20 периодов семисуточной продолжительности. Для тестирования экспериментальных ИНС отобраны два летних и три весенних семисуточных периодов в разные годы. Для каждого из периодов выбраны данные четырех типов АСКЗА (табл. 2).

Обсуждение результатов

По итогам обучения экспериментальных ИНС (табл. 1) по анализу кривых обучения лучшие результаты показали ИНС 4–7 и 10–12 – двух- и трехслойные ИНС с включением затухания скорости обучения и импульса обновления весов, среднее абсолютное отклонение (САО) которых на 100 эпохе достигает значений в пределах 8–9. Менее успешные результаты показали ИНС 1–3 с вариациями архитектур полносвязных слоев – САО

на 100 эпохе достигает значений 9–11; худшие результаты показали ИНС 8 и 9 с исключениями нейронов в слоях – САО на 100 эпохе 12.36 и 18.76 в эксперименте 8 и 9 соответственно. На рис. 2 для сравнения приведены кривые обучения ИНС-7 и ИНС-9.

Таблица 2. Временные ряды данных на АСКЗА, составляющие независимую выборку для тестирования экспериментальных конфигураций ИНС

Table 2. ASKZA data time series constituting an independent sample for experimental ANN configurations testing

Временные периоды		Типовая принадлежность АСКЗА			
Начало периода	Конец периода	Транспортная	Смешанная	Жилая	Фоновая
5.06.2019	11.06.2019	Хамовники	Марьино	Спиридоновка	МГУ
10.05.2022	16.05.2022	МАДИ	Марьино	Толбухина	МГУ
13.04.2021	19.04.2021	Площадь Гагарина	Кожухово	Спиридоновка	МГУ
16.03.2020	22.03.2020	Хамовники	Марьино	Туристская	МГУ
28.06.2022	04.07.2022	Площадь Гагарина	Марьино	Спиридоновка	МГУ

В табл. 3 представлены общие – осредненные по всем тестовым подвыборкам (табл. 2) – статистические показатели качества первичных модельных и скорректированных ИНС прогнозов озона для каждой обученной в ходе экспериментов ИНС.

Используемые в табл. 3 обозначения:

- средняя за период концентрация озона: OBS – измерения на АСКЗА, Chim – прогноз CHIMERE, ANN – скорректированная ИНС.

- Bias, ABS, RMS, R – среднее отклонение, среднее абсолютное отклонение, стандартное отклонение ошибки, коэффициент корреляции прогнозов CHIMERE (Chim) / скорректированных ИНС прогнозов (ANN) и измерений на АСКЗА.

Обращают на себя внимание знаки средних отклонений: при том что CHIMERE в среднем завьшает концентрации озона по сравнению с измерениями, расчеты всех 12 тестируемых ИНС приводят к занижению концентраций, но оно по величинам меньше, чем завьшения ХТМ CHIMERE (за исключением экспериментов 1, 2, 8, 9). Наименьшие и близкие к 0 средние отклонения «расчет-наблюдение» оказались свойственными для ИНС 5 и 7, 11 и 12.

Отметим, во всех экспериментах (кроме 9) наблюдается снижение средней абсолютной ошибки прогноза CHIMERE после коррекции ИНС. Наибольшее уменьшение среднего абсолютного отклонения – практически в 2 раза – показывали ИНС 4–7 и 10–12, то есть двух- и трехслойные ИНС

с применением затухания скорости обучения и импульса обновления весов. Эти же ИНС показывают наибольшее снижение стандартного отклонения ошибки прогноза.

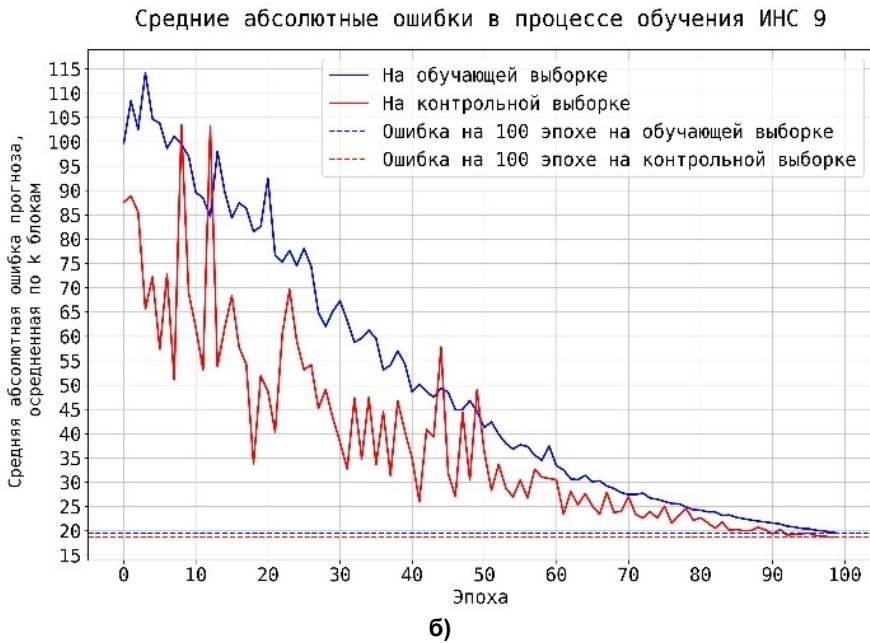
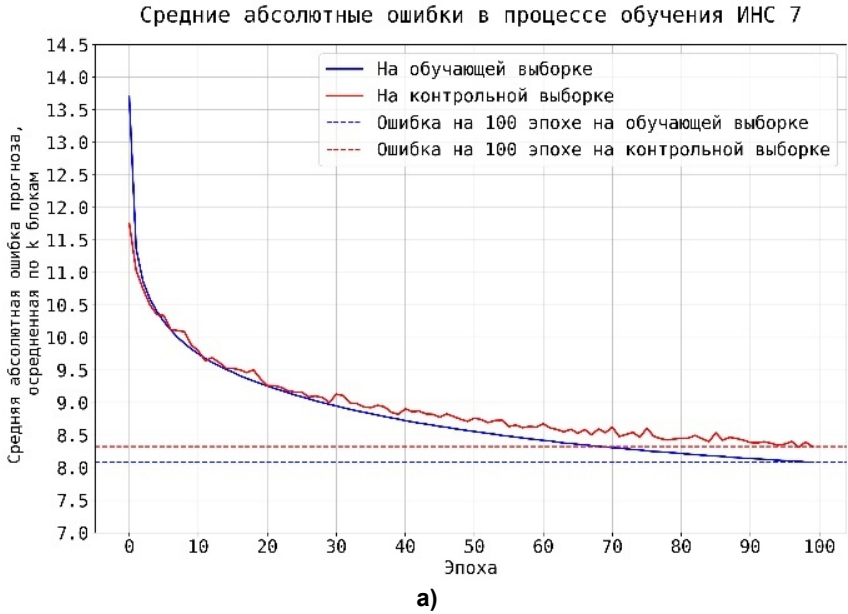


Рис. 2. Кривые обучения ИНС конфигураций 7 (а) и 9 (б).
Fig. 2. Learning curves of ANN configurations 7 (a) and 9 (б).

Таблица 3. Статистические показатели качества первичных модельных и скорректированных экспериментальными ИНС прогнозов озона**Table 3.** Statistical indicators of the quality of primary model and experimentally ANN corrected ozone forecasts

Эксперимент	1	2	3	4	5	6	7	8	9	10	11	12
Средняя за период концентрация озона												
OBS	56,67	56,67	56,67	56,67	56,67	56,67	56,67	56,67	56,67	56,67	56,67	56,67
Chim	62,21	62,21	62,21	62,21	62,21	62,21	62,21	62,21	62,21	62,21	62,21	62,21
ANN	49,02	50,78	53,47	53	54,23	54,06	55,72	48,49	42,14	53,39	54,26	53,91
Bias												
Chim	5,54	5,54	5,54	5,54	5,54	5,54	5,54	5,54	5,54	5,54	5,54	5,54
ANN	-7,65	-5,88	-3,2	-3,66	-2,44	-2,61	-0,94	-8,17	-14,52	-3,27	-2,41	-2,75
ABS												
Chim	23,61	23,61	23,61	23,61	23,61	23,61	23,61	23,61	23,61	23,61	23,61	23,61
ANN	17,14	16,21	15,05	13,45	13,4	13,36	13,18	17,5	24,02	13,82	13,49	13,96
RMS												
Chim	28,87	28,87	28,87	28,87	28,87	28,87	28,87	28,87	28,87	28,87	28,87	28,87
ANN	21,56	20,5	19,23	17,19	17,41	17,28	17,14	21,72	27,97	17,75	17,44	18
R												
Chim	0,57	0,57	0,57	0,57	0,57	0,57	0,57	0,57	0,57	0,57	0,57	0,57
ANN	0,77	0,8	0,82	0,86	0,86	0,86	0,87	0,77	0,56	0,85	0,86	0,85

Для коррекции всех ИНС, кроме ИНС-9, характерно увеличение коэффициента корреляции (R) «прогноз – измерение» по сравнению с первичными прогнозами CHIMERE, наибольшее увеличение с $R=0.57$ до $R=0.85-0.87$ показали те же ИНС 4–7 и 10–12.

В среднем по тестовой выборке наилучшие результаты по всем полученным оценкам – снижение Bias, ABS, RMS; увеличение R – показала ИНС-7 – двуслойная ИНС 100–50 нейронов со значением импульса обновления весов 0.9. Худшие результаты с увеличением средних отклонений (по модулю) и снижением коэффициента корреляции после коррекции показали ИНС 8 и 9 – с исключениями нейронов в слоях.

Далее обсуждаются результаты коррекции прогнозов CHIMERE с помощью лучшей ИНС-7 в летнем и весеннем эпизодах на отдельных подвыборках вида *временной период-АСКЗА*.

Как видно в табл. 4, по всем осредненным за летний эпизод 5–11 июня 2019 г. показателям и критериям расчеты ИНС-7 имеют преимущества перед первичными прогнозами CHIMERE.

Таблица 4. Средняя за период концентрация приземного озона ($\mu\text{г}/\text{м}^3$) – измеренная (OBS), рассчитанная по ХТМ CHIMERE (Chim) и ИНС-7 (ANN), статистические характеристики качества прогнозов CHIMERE и ИНС-7. Эпизод 5-11 июня 2019 г.
Table 4. Period-average ground-level ozone concentration ($\mu\text{g}/\text{m}^3$) – measured (OBS), calculated using CHIMERE (Chim) and ANN-7 (ANN), statistical characteristics of the quality of CHIMERE and ANN-7 forecasts. June 5-11, 2019 episode

Эксперимент	АСКЗА			
	Хамовники Транспортная	Марьино Смешанная	Спиридоновка Жилая	МГУ Фоновая
OBS	59,11	60,85	70,37	76,77
Chim	47	45,16	42,76	45,55
ANN	51,27	65,31	67,8	60,98
Bias Chim	-12,11	-15,69	-27,61	-31,22
Bias ANN	-7,84	4,46	-2,57	-15,79
ABS Chim	17,59	20,55	28,23	35,2
ABS ANN	13,77	12,13	14,36	20,46
RMS Chim	19,28	21,42	20,47	27,86
RMS ANN	16,03	14,56	17,05	20,34
R Chim	0,89	0,9	0,92	0,86
R ANN	0,93	0,96	0,94	0,92

На фоне хороших средних за период показателей качества коррекции первичных прогнозов ХТМ с помощью ИНС-7, при анализе временного хода модельных расчетов и измеренных концентраций можно отметить некоторые особенности. На рис. 3 иллюстрируются расчеты ХТМ, ИНС-7 для станций четырех типов (см. табл. 4).

Обсуждаемый летний эпизод характеризуется высокими уровнями озона, превысившими 6 и 7 июня ПДК_{м.р.}, (равная $160 \mu\text{г}/\text{м}^3$, отмечено черной пунктирной линией на рис. 3). Повышенные концентрации наблюдались в послеполуденные часы, что объясняется фотохимическим производством озона в приземном слое в присутствии предшественников – оксидов азота, при высокой УФ-облученности. В темное время суток концентрация приземного озона уменьшалась по наблюдениям и модельным расчетам до $1\text{--}10 \mu\text{г}/\text{м}^3$.

Как видно на рис. 3, в данном эпизоде CHIMERE занижает концентрации озона по сравнению с измерениями на всех АСКЗА, больше всего – на АСКЗА жилого и фоновых типов. В то же время CHIMERE довольно точно воспроизводила суточную изменчивость озона, время дневных максимумов и ночных минимумов, что подтверждается высокими коэффициентами корреляции между измерениями и прогнозами на всех АСКЗА ($R_{\text{Chim}} = 0.86\text{--}0.92$).

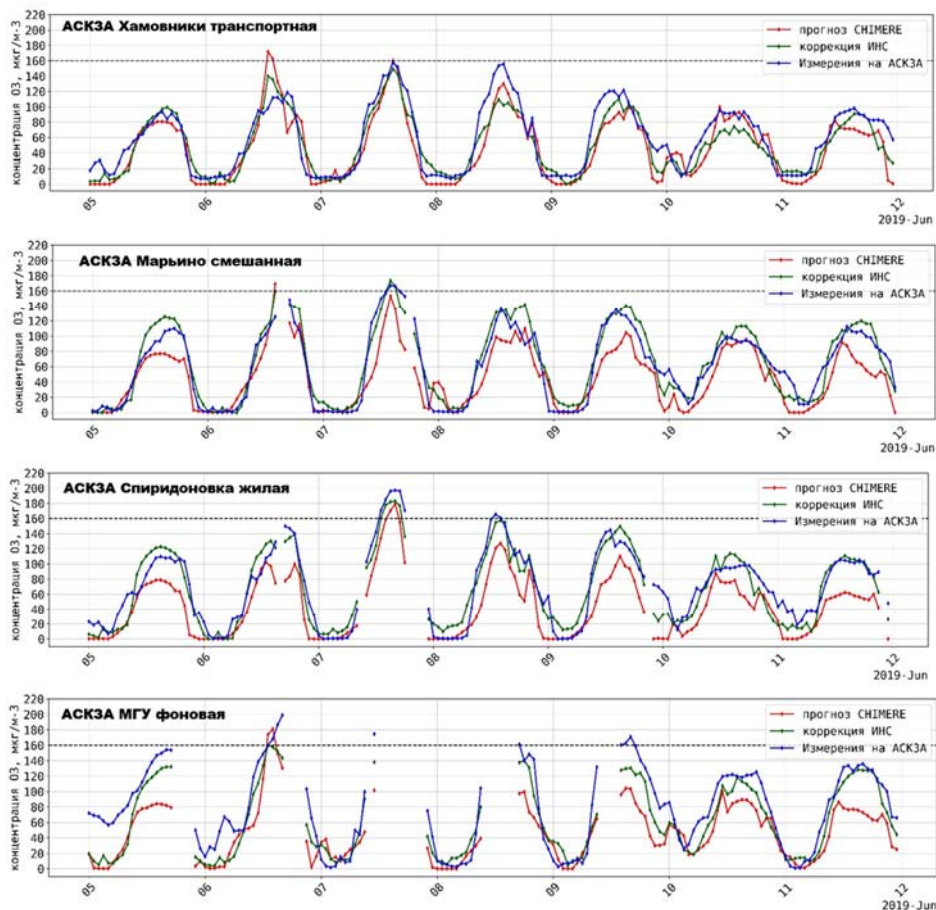


Рис. 3. Временной ход концентрации приземного озона по измерениям на АСКЗА, прогнозам CHIMERE и скорректированным ИНС-7 прогнозам в период 5-11 июня 2019 г.

Fig. 3. Dynamics of ground-level ozone concentration according to measurements at ASKZA, CHIMERE forecasts and adjusted with ANN-7 forecasts for the June 5-11, 2019 period.

ИНС заметно корректирует прогноз концентраций озона по ХТМ CHIMERE – на всех АСКЗА уменьшила модельные занижения, приблизив значения прогноза к измерениям. Наиболее выраженная эффективность коррекции ИНС установлена для АСКЗА смешанного и жилого типа (Марьино и Спиридоновка). На АСКЗА транспортного типа Хамовники обнаружены ошибочные занижения ИНС-7 8 и 10 июня. На фоновой станции МГУ ИНС-7 почти вдвое уменьшила отрицательное отклонение прогноза CHIMERE. В целом следует отметить, что в данном эпизоде ИНС-7 повысила

точность прогноза максимумов и суточных циклов концентрации приземного озона, также на всех АСКЗА увеличился коэффициент корреляции «прогноз – измерение» до 0,92–0,96.

Выше рассматривался летний период с зафиксированными «озоновыми» эпизодами фотохимического происхождения. Отдельный интерес представляет анализ возможностей моделирования и коррекции с помощью ИНС в эпизоде весеннего значительного повышения содержания озона в приземном воздухе вследствие усиления притока тропосферного озона. Один из таких эпизодов наблюдался в середине апреля 2021 г., что хорошо видно на рис. 4.

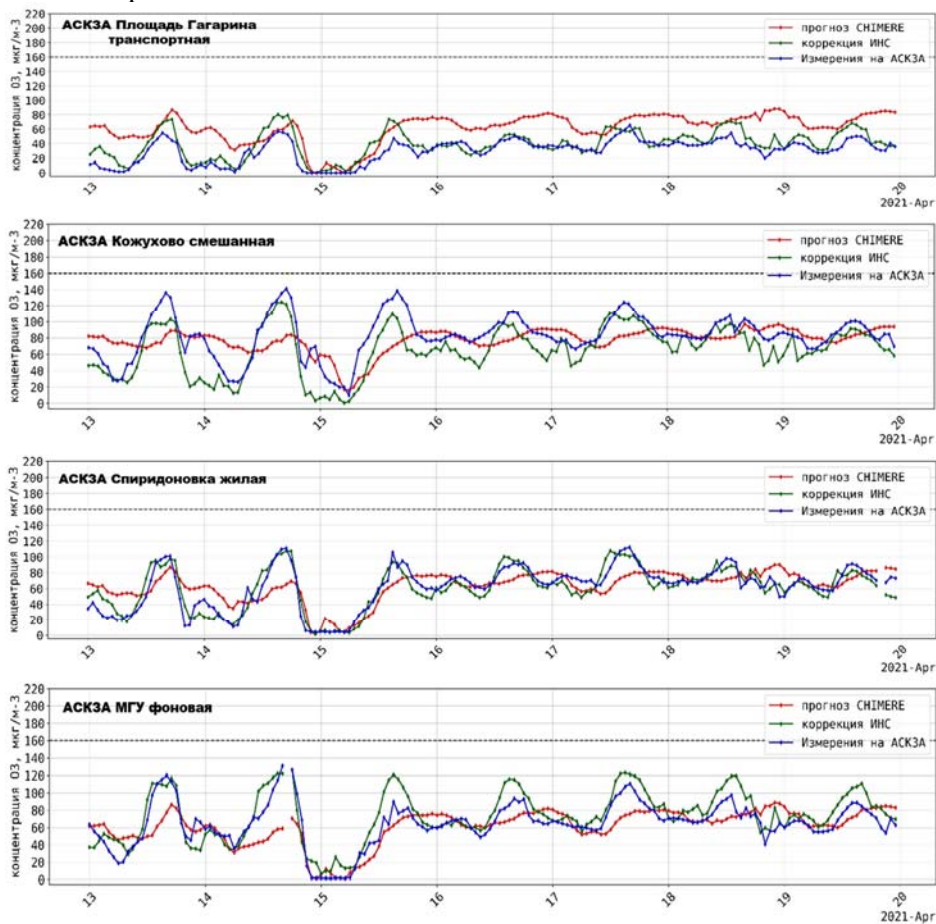


Рис. 4. Временной ход концентрации приземного озона по измерениям на АСКЗА, прогнозам CHIMERE и скорректированным ИНС-7 прогнозам в период 13-19 апреля 2021 г.

Fig. 4. Dynamics of ground-level ozone concentration according to measurements at ASKZA, CHIMERE forecasts and adjusted with ANN-7 forecasts for the period April 13-19, 2021 period

Рассматриваемый весенний эпизод содержит два режима приземного озона: 13–15 апреля наблюдался близкий к типовому суточный ход с хорошо выраженной амплитудой озона до 80–100 мкг/м³; вторая половина эпизода 16–19 апреля отражала характерный для ветреной неустойчивой погоды режим приземного озона с небольшими внутрисуточными колебаниями концентрации на повышенном фоне.

Как видно на рис. 4, расчеты ИНС-7 имеют явное преимущество над прогнозами концентрации приземного озона CHIMERE на всех типовых АСКЗА. CHIMERE так же, как в предыдущем эпизоде, чаще всего искажает суточную динамику концентраций озона. ИНС-7 на АСКЗА транспортного типа (пл. Гагарина) уменьшает завышенные прогнозы CHIMERE и корректирует суточную изменчивость озона. На АСКЗА смешанного типа (Кожухово) ИНС ошибочно занижает прогнозы CHIMERE, которые оказались в среднем близки к измерениям, но улучшает воспроизведение суточной изменчивости концентраций озона ($R_{\text{Chim}} = 0,38$, $R_{\text{ANN}} = 0,87$). При этом следует отметить, что на этой АСКЗА, расположенной за пределами МКАД, измеренные концентрации озона почти на 20 мкг/м³ превышали уровень озона на городских станциях (табл. 5).

Таблица 5. Средняя за период концентрация приземного озона (мкг/м³) – измеренная (OBS), рассчитанная по ХТМ CHIMERE (Chim) и ИНС-7 (ANN), статистические характеристики качества прогнозов CHIMERE и ИНС-7. Эпизод 13–19 апреля 2021 г.

Table 5. Period-average ground-level ozone concentration ($\mu\text{g}/\text{m}^3$) – measured (OBS), calculated using CHIMERE (Chim) and ANN-7 (ANN), statistical characteristics of the quality of CHIMERE and ANN-7 forecasts. April 13-19, 2021 episode

Эксперимент	АСКЗА			
	Площадь Гагарина Транспортная	Кожухово Смешанная	Спиридоновка Жилая	МГУ Фоновая
OBS	30,4	82,34	62,16	64,8
Chim	61,59	77,39	63,31	61,27
ANN	39,37	65,62	61	73,72
Bias Chim	31,19	-4,94	1,15	-3,53
Bias ANN	8,97	-16,71	-1,16	8,92
ABS Chim	31,19	19,64	15,35	14,13
ABS ANN	10,34	17,37	8,96	12,64
RMS Chim	14,85	24,82	18,88	18,72
RMS ANN	9,77	14,33	10,96	13,17
R Chim	0,68	0,38	0,71	0,68
R ANN	0,85	0,87	0,92	0,89

На АСКЗА жилого типа (Спиридоновка) CHIMERE довольно точно спрогнозировала средний уровень озона, но ИНС успешнее отразила суточную изменчивость прогнозов CHIMERE ($R_{\text{Chim}} = 0,71$, $R_{\text{ANN}} = 0,92$). На АСКЗА фоновом типа МГУ ИНС-7 ошибочно завысила прогнозы CHIMERE, в то же время значительно улучшила воспроизведение суточной динамики прогностических концентраций ($R_{\text{Chim}} = 0,68$, $R_{\text{ANN}} = 0,89$).

По результатам тестовых сравнений первичных модельных прогнозов концентрации приземного озона по ХТМ CHIMERE с расчетами ИНС в двух недельных эпизодах выявлена и подтверждена количественными показателями эффективность применения ИНС для уточнения суточной динамики и уровня озона на большинстве АСКЗА.

Заключение

Выполнен анализ возможностей применения ИНС в области прогнозирования качества воздуха. Изучены методы построения ИНС, разработаны авторские алгоритмы формирования обучающих выборок; создания, обучения и тестирования конфигураций ИНС для выбора оптимальных моделей коррекции прогнозов атмосферных загрязнений.

Впервые для статистической коррекции модельных прогнозов концентрации приземного озона, рассчитываемых на основе ХТМ CHIMERE, применены ИНС. Для разработки ИНС сформирована база данных часовых модельных прогнозов озона, его предшественников и наблюдений в период 2019–2023 гг. Используются непрерывные измерения концентраций приземного озона на АСКЗА «Мосэкомониторнг». Проведены эксперименты по построению, обучению и тестированию ИНС с вариациями конфигураций ИНС.

Экспериментальное тестирование всех созданных ИНС показало целесообразность применения метода для повышения качества прогноза приземного озона.

Авторы выражают признательность и благодарность руководству ГПБУ «Мосэкомониторинг» за предоставление данных, позволяющее выполнять научно-практические разработки на современном уровне.

Список использованных источников

1. Антохин П.Н., Белан Б.Д. Регулирование динамики тропосферного озона через стратосферу // Оптика атмосферы и океана. 2012. Т. 25, № 10. С. 890-895.
2. Антохин П.Н., Белан Б.Д., Савкин Д.Е., Толмачев Г.Н. Сравнение различных методов статистического прогнозирования суточной динамики приземной концентрации озона // Оптика атмосферы и океана. 2013. Т. 26, № 12. С. 1082-1089.
3. Белан Б.Д. Озон в тропосфере. Томск: Изд-во ИОА СО РАН, 2010. 488 с.
4. Борисов Д.В., Шалыгина И.Ю., Лезина Е.А. Исследование сезонной и суточной изменчивости концентраций приземного озона // Гидрометеорологические исследования и прогнозы. 2020. № 3 (377). С. 122-135.

5. Борисов Д.В., Кузнецова И.Н., Нахаев М.И. Изменения кадастровых данных о выбросах в атмосферу загрязняющих веществ в московском регионе // Гидрометеорологические исследования и прогнозы. 2023. № 2 (388). С. 156-173. DOI: <https://doi.org/10.37162/2618-9631-2023-2-156-173>.

6. Еланский Н.Ф., Шилкин А.В., Пономарев Н.А., Захарова П.В., Качко М.Д., Поляков Т.И. Пространственно-временные вариации содержания загрязняющих примесей в воздушном бассейне Москвы и их эмиссии // Известия РАН. ФАО. 2022. Т. 58, № 1. С. 92-108.

7. Кузнецова И.Н., Нахаев М.И., Курсанов А.А., Борисов Д.В., Ткачева Ю.В., Ривин Г.С., Лезина Е.А. Тестирование и перспективы технологии прогнозирования загрязнения воздуха с применением химических транспортных моделей CHIMERE и COSMO-Ru2ART // Гидрометеорологические исследования и прогнозы. 2022. № 4 (386). С. 147-170. DOI: <https://doi.org/10.37162/2618-9631-2022-4-147-170>.

8. Ривин Г.С., Розинкина И.А., Астахова Е.Д., Блинов Д.В., Бундель А.Ю., Курсанов А.А., Шатунова М.В., Чубарова Н.Е., Алферов Д.Ю., Варенцов М.И., Захарченко Д.И., Копейкин В.В., Никитин М.А., Полюхов А.А., Ревокатова А.П., Татарникович Е.В., Чурюлин Е.В. Система краткосрочного численного прогноза высокой детализации COSMO-Ru, ее развитие и приложения // Гидрометеорологические исследования и прогнозы. 2019. № 4 (374). С. 27-53.

9. Borisov D.V., Kuznetsova I.N., Shalygina I.Y., Nahaev M.I., Lezina E.A., Lapchenko V.A. Monitoring of surface ozone in the Moscow region // Тезисы конференции QOS (Online Meeting Quadrennial Ozone Symposium). 3–9 October 2021.

10. Freeman B.S., Taylor G., Gharabaghi B., Th'e J. Forecasting air quality time series using deep learning // J. Air Waste Manag. Assoc. 2018. Vol. 68, no. 8. P. 866-886. <https://doi.org/10.1080/10962247.2018.1459956>.

11. Masood A., Kafeel A. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance // Journal of Cleaner Production. 2021. Vol. 322. P. 129072. DOI: 10.1016/j.jclepro.2021.129072.

12. Min Xu et al. Machine learning based bias correction for numerical chemical transport models // Atmospheric Environment. 2021. Vol. 248. <https://doi.org/10.1016/j.atmosenv.2020.118022>.

References

1. Antokhin P.N., Belan B.D. Control of the Dynamics of Tropospheric Ozone through the Stratosphere. *Atmospheric and Oceanic Optics*, 2013, vol. 26, no. 3, pp. 207-213.

2. Antokhin P.N., Belan B.D., Savkin D.E., Tolmachev G.N. The comparison of different methods of statistical prediction of diurnal dynamics in the ground ozone concentration. *Atmospheric and Oceanic Optics*, 2013, vol. 26, no. 12.

3. Belan B.D. Ozon v troposphere [Ozone in the troposphere]. Tomsk, IOA SO RAN publ., 2010, 488 p. [in Russ.].

4. Borisov D.V., Shalygina I.Yu., Lezina E.A. Studying seasonal and daily variability of surface ozone. *Gidrometeorologicheskie issledovaniya i prognozy [Hydrometeorological Research and Forecasting]*, 2020, vol. 377, no. 3, pp. 122-135 [in Russ.].

5. Borisov D.V., Kuznetsova I.N., Nakhaev M.I. Changes in inventory data on pollutant emissions into the atmosphere over the Moscow region. *Gidrometeorologicheskie issledovaniya i prognozy [Hydrometeorological Research and Forecasting]*, 2023,

vol. 388, no. 2, pp. 156-173. DOI: <https://doi.org/10.37162/2618-9631-2023-2-156-173> [in Russ.].

6. Elansky N.F., Shilkin A.V., Ponomarev N.A., Zakharova P.V., Kachko M.D., Poliakov T.I. Spatiotemporal variations in the content of pollutants in the Moscow air basin and their emissions. *Izvestiya, Atmospheric and Oceanic Physics*, 2022, vol. 58, no. 1, pp. 80-94.

7. Kuznetsova I.N., Nakhaev M.I., Kirsanov A.A., Borisov D.V., Tkacheva Yu.V., Rivin G.S., Lezina E.A. Testing and prospects of air pollution prediction technology based on CHIMERE and COSMO-Ru2ART chemical transport models. *Gidrometeorologicheskie issledovaniya i prognozy [Hydrometeorological Research and Forecasting]*, 2022, vol. 386, pp. 147-170. DOI: [10.37162/2618-9631-2022-4-147-170](https://doi.org/10.37162/2618-9631-2022-4-147-170) [in Russ.].

8. Rivin G.S., Rozinkina I.A., Astakhova E.D., Blinov D.V., Bundel' A.Yu., Kirsanov A.A., Shatunova M.V., Chubarova N.Ye., Alferov D.Yu., Varentsov M.I., Zakharchenko D.I., Kopeykin V.V., Nikitin M.A., Poliukhov A.A., Revokatova A.P., Tatarinovich E.V., Churiulin E.V. COSMO-Ru high-resolution short-range numerical weather prediction system: its development and applications. *Gidrometeorologicheskie issledovaniya i prognozy [Hydrometeorological Research and Forecasting]*, 2019, vol. 374, no.4, pp. 27-53 [in Russ.].

9. Borisov D.V., Kuznetsova I.N., Shalygina I.Y., Nahaev M.I., Lezina E.A., Lapchenko V.A. Monitoring of surface ozone in the Moscow region. Online Meeting Quadrennial Ozone Symposium. 3–9 October 2021.

10. Freeman B.S., Taylor G., Gharabaghi B., Th'e J. Forecasting air quality time series using deep learning. *J. Air Waste Manag. Assoc.*, 2018, vol. 68, no. 8, pp. 866-886. DOI: [10.1080/10962247.2018.1459956](https://doi.org/10.1080/10962247.2018.1459956).

11. Masood A., Kafeel A. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production*, 2021, Vol. 322, pp. 129072. DOI: [10.1016/j.jclepro.2021.129072](https://doi.org/10.1016/j.jclepro.2021.129072).

12. Min Xu et al. Machine learning based bias correction for numerical chemical transport models. *Atmospheric Environment*. 2021, vol. 248. DOI: [10.1016/j.atmosenv.2020.118022](https://doi.org/10.1016/j.atmosenv.2020.118022).

Поступила 17.11.2023; одобрена после рецензирования 23.11.2023;
принята в печать 12.12.2023.

Submitted 17.11.2023; approved after reviewing 23.11.2023;
accepted for publication 12.12.2023.