

УДК 519.62  
МРНТИ 27.35.21; 27.35.47; 27.41.19

<https://doi.org/10.55452/1998-6688-2026-23-2-83-107>

<sup>1\*</sup>**Исахов А.А.,**

профессор, ORCID ID: 0000-0002-1937-8615,

\*e-mail: alibek.issakhov@gmail.com

<sup>2</sup>**Нигметова А.,**

докторант, ORCID ID: 0009-0004-5395-4569,

e-mail: a.nygmetova@kbtu.kz

<sup>1</sup>**Абылкасымова А.Б.,**

ассоциированный профессор, ORCID ID: 0000-0002-5967-6959,

e-mail: abylkassymova.aizhan@gmail.com

<sup>1</sup>**Аханова Н.Е.**

профессор, ORCID ID: 0000-0003-2767-8711,

e-mail: n.akhanova@kbtu.kz

<sup>1</sup>Казахстанско-Британский технический университет, г. Алматы, Казахстан

<sup>2</sup>Международный университет информационных технологий, г. Алматы, Казахстан

## ПРОГНОЗИРОВАНИЕ ЗАГРЯЗНЕНИЯ ВОЗДУШНОЙ СРЕДЫ ОТ ВЫБРОСОВ В ГОРОДСКИХ КАНЬОНАХ С ИСПОЛЬЗОВАНИЕМ ML-CFD МОДЕЛИРОВАНИЯ

### Аннотация

В работе представлен гибридный подход для прогнозирования распространения загрязняющих веществ в городских уличных каньонах с учетом шумозащитных барьеров. Методология объединяет детальное CFD-моделирование и суррогатную модель на основе нейросетевой архитектуры BiLSTM с механизмом внимания. Исследованы конфигурации с барьерами высотой 0.1H, 0.2H и 0.3H. CFD-расчеты выявили нелинейное влияние высоты барьера на аэродинамику и формирование зон накопления примеси, причем наиболее сложный нестационарный режим наблюдается при высоте 0.2H. Суррогатная модель успешно прогнозирует эволюцию концентрации для случаев без барьера и с барьером 0.1H, демонстрируя среднюю абсолютную процентную ошибку менее 15%. Для барьера 0.2H точность снижается в зонах интенсивной турбулентности, что связано с высокой нестационарностью процесса. Подход позволяет значительно сократить вычислительные затраты при сохранении физической достоверности, что перспективно для систем поддержки решений в области городской экологии. Модель обеспечивает ускоренное прогнозирования по сравнению с CFD-расчетом на 7–8 порядков, так потраченное время инференса показало 1–5 мс, тогда как одна CFD-симуляция занимает около 54 часов на CPU.

**Ключевые слова:** загрязнение воздуха, уличные каньоны, шумозащитные барьеры, вычислительная гидродинамика (CFD), машинное обучение, BiLSTM, суррогатное моделирование, прогнозирование концентраций

### Введение

Загрязнение атмосферного воздуха в городской среде остается одной из наиболее значимых угроз общественному здоровью и устойчивому развитию. Многочисленные эпидемиологические исследования указывают на устойчивую связь между повышенными концентрациями загрязняющих веществ в приземном слое атмосферы и ростом заболеваемости респираторными, сердечно-сосудистыми и онкологическими заболеваниями [1–4]. Наибольшую опасность представляют мелкодисперсные аэрозольные частицы PM<sub>2.5</sub> и PM<sub>10</sub>, способные глубоко проникать в дыхательные пути и накапливаться в тканях организма.

В условиях современных городов доминирующим источником загрязнения воздуха остается автотранспорт. Рост урбанизации, увеличение плотности застройки и транспортной нагрузки приводят к формированию локальных зон повышенной концентрации загрязняющих веществ, особенно в часы пик [5–8]. Расширение городских территорий и зависимость от личного транспорта усугубляют эту проблему, формируя устойчивые очаги загрязнения вблизи магистралей и жилых районов.

Ключевую роль в формировании пространственного распределения загрязняющих веществ играет архитектурная структура городской среды. Плотная и высотная застройка формирует так называемые уличные каньоны, в которых ограниченная вентиляция и сложная структура турбулентности способствуют накоплению загрязнений [9–11]. В таких условиях даже при умеренных скоростях ветра возникают застойные зоны и устойчивые вихревые структуры, приводящие к формированию длительных эпизодов смога, рассеивание которых возможно лишь при благоприятных метеорологических условиях [12].

Для анализа процессов переноса и дисперсии загрязняющих веществ в сложной городской геометрии широко применяются методы вычислительной гидродинамики. CFD-моделирование позволяет детально учитывать форму застройки, турбулентные эффекты и нестационарные граничные условия, обеспечивая высокую физическую достоверность результатов по сравнению с параметрическими моделями [13–16]. Численные исследования уличных каньонов показали, что CFD является эффективным инструментом для оценки распределения концентраций и анализа влияния архитектурных и инженерных элементов городской инфраструктуры.

Однако практическое применение CFD в задачах оперативного прогнозирования и сценарного анализа существенно ограничено высокой вычислительной стоимостью. Детализированное моделирование турбулентных потоков в городской среде требует значительных вычислительных ресурсов, что затрудняет проведение многовариантных расчетов и использование таких моделей в системах поддержки принятия решений [15,16]. Эти ограничения стимулировали развитие альтернативных и комплементарных подходов, направленных на ускорение расчетов при сохранении приемлемой точности.

В последние годы все более активно развиваются методы машинного обучения, применяемые для прогнозирования качества воздуха и анализа результатов численного моделирования. ML-подходы используются как для обработки данных наблюдений, так и для построения суррогатных моделей, аппроксимирующих результаты CFD-расчетов [17, 18]. Такие модели позволяют существенно сократить вычислительные затраты и обеспечить быстрый отклик при анализе различных сценариев. Особый интерес представляют гибридные CFD–ML подходы, в которых машинное обучение используется в качестве дополнения к физическим моделям, а не их замены, что повышает устойчивость и интерпретируемость прогнозов.

Отдельного внимания заслуживает влияние инженерных решений и элементов городской инфраструктуры на процессы дисперсии загрязняющих веществ. Шумозащитные барьеры, элементы застройки и планировочные решения способны существенно изменять структуру потока, как снижая концентрации в отдельных зонах, так и формируя локальные области повышенного загрязнения [19–22]. Анализ таких эффектов требует сочетания физически обоснованных численных моделей и инструментов быстрого анализа, что делает гибридные подходы особенно актуальными.

Таким образом, разработка и применение методов, объединяющих CFD-моделирование и машинное обучение, представляется перспективным направлением для исследования загрязнения воздуха в уличных каньонах. Такой подход позволяет совместить физическую достоверность численных моделей с вычислительной эффективностью ML-алгоритмов и создать инструменты, пригодные для анализа сценариев, оптимизации городской инфраструктуры и поддержки экологически ориентированных решений.

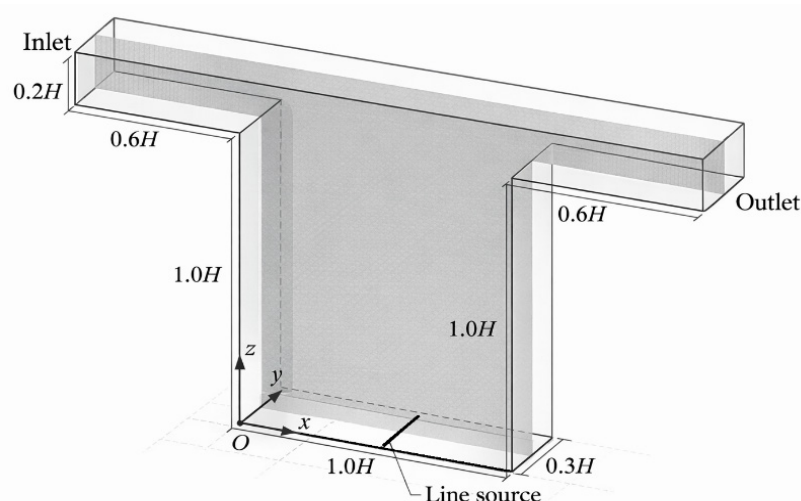
В рамках настоящего исследования рассматривается задача прогнозирования пространственного распределения загрязняющих веществ в идеализированных городских уличных каньонах с учетом влияния инженерных элементов городской инфраструктуры. Основное внимание уделяется разработке и валидации гибридного подхода, сочетающего численное моделирование на основе методов вычислительной гидродинамики и методы машинного обучения, ориентированные на построение вычислительно эффективных суррогатных моделей. CFD-расчеты используются для получения физически обоснованных данных о структуре потока и распределении концентраций при различных геометрических и аэродинамических конфигурациях, тогда как алгоритмы машинного обучения применяются для выявления скрытых закономерностей и аппроксимации результатов численного моделирования.

## Материалы и методы

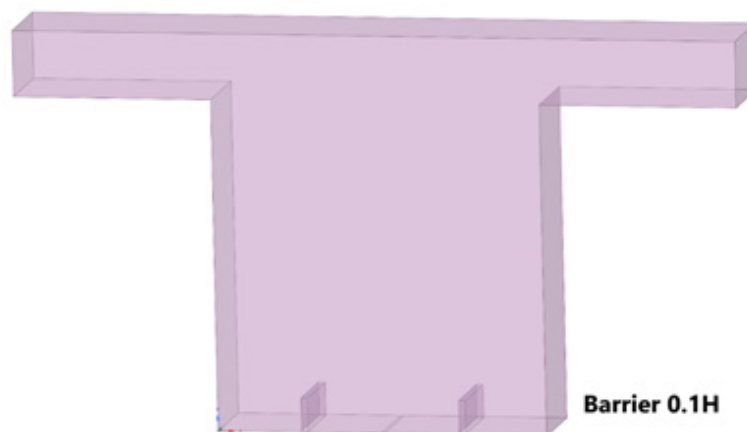
### Постановка задачи и геометрия вычислительной области

Численное моделирование процессов переноса и дисперсии загрязняющих веществ в условиях городской застройки проводилось на основе методов вычислительной гидродинамики (CFD). В качестве объекта исследования рассматривался идеализированный уличный каньон, образованный двумя параллельными зданиями одинаковой высоты  $H$  (рис. 1а). Длина каньона значительно превышала его ширину и высоту, что позволяло пренебречь влиянием торцевых границ на установившийся режим течения в центральной части.

Границы вычислительной области, включая поверхности зданий, дорожное полотно и верхнюю границу, задавались в соответствии с условиями, имитирующими нейтральную стратификацию атмосферного пограничного слоя. Для анализа влияния инфраструктурных элементов вдоль линии, соответствующей расположению источника выбросов, были смоделированы шумозащитные барьеры различной высоты (рис. 1б). Высота барьера варьировалась в долях от высоты зданий и принимала значения  $0.1H$ ,  $0.2H$  и  $0.3H$ , что позволило исследовать зависимость структуры потока и распределения примеси от этого параметра. Толщина барьера принята  $0.01H$ .



a



b

Рисунок 1 – Геометрия вычислительной области:  
а) без барьера, б) барьер с высотой 0.1H

Уравнения, описывающие процесс, и физические предположения

Моделирование источника загрязняющих веществ было реализовано в виде линейного источника, расположенного на уровне дорожного полотна и ориентированного вдоль оси каньона (рисунок 1). Такая постановка соответствует выбросам от непрерывного потока автотранспорта. В качестве модельного загрязнителя использовался этилен ( $C_2H_4$ ). Вещество рассматривалось как пассивная примесь, не влияющая на плотность и динамику воздушного потока и не участвующая в химических реакциях.

Движение воздушной среды в уличном каньоне описывается осредненными по Рейнольдсу уравнениями Навье–Стокса для несжимаемого турбулентного течения. Такой подход является общепринятым при моделировании атмосферных потоков в городской среде и позволяет адекватно воспроизводить основные характеристики средних полей скорости и давления при умеренных вычислительных затратах. Плотность воздуха считалась постоянной, а влияние температурной стратификации и плавучести не учитывалось, что соответствует приближению нейтральной атмосферной устойчивости.

Система уравнений включает уравнение неразрывности, уравнения сохранения импульса и уравнение конвективно-диффузионного переноса пассивной примеси:

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0 \quad (1)$$

$$\frac{\partial \rho \bar{u}_i}{\partial t} + \frac{\partial \rho \bar{u}_i \bar{u}_j}{\partial x_j} = - \frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j} \left[ (\mu + \mu_t) \left( \frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) \right] - \frac{2}{3} \delta_{ij} \rho k \quad (2)$$

$$\frac{\partial \rho \bar{c}}{\partial t} + \frac{\partial \rho \bar{c} \bar{u}_i}{\partial x_i} = \frac{\partial}{\partial x_i} \left[ \rho \left( D + \frac{\nu_t}{Sc_t} \right) \frac{\partial \bar{c}}{\partial x_i} \right] \quad (3)$$

Здесь  $\bar{u}_i$  – компоненты осредненной скорости,  $\bar{p}$  – осредненное давление,  $\bar{c}$  – осредненная концентрация примеси,  $\mu$  – динамическая вязкость (для воздуха –  $1.79 \times 10^{-5}$  кг/(м·с), для этилена –  $1.03 \times 10^{-5}$  кг/(м·с)),  $\mu_t$  – турбулентная вязкость,  $k$  – турбулентная кинетическая энергия,  $D$  – молекулярный коэффициент диффузии (для этилена –  $2.88 \times 10^{-5}$  м<sup>2</sup>/с),  $\nu_t$  – турбулентная кинематическая вязкость ( $\nu_t = \mu_t / \rho$ ),  $\delta_{ij}$  – символ Кронекера, а  $Sc_t$  – турбулентное число Шмидта, принятое равным 0.7 [14].

Для замыкания системы использовалась модель турбулентности SST  $k-\omega$ . Эта двухпараметрическая модель сочетает точность стандартной  $k-\omega$  модели в пристенной области с устойчивостью  $k-\epsilon$  модели в удаленных от стен зонах, что делает ее эффективной для задач обтекания зданий, где существенны отрывные течения, сдвиговые слои и взаимодействие вихревых структур. Модель включает два транспортных уравнения – для турбулентной кинетической энергии  $k$  и удельной скорости диссипации  $\omega$ , которые решались совместно с уравнениями (1)–(3).

Граничные условия были заданы следующим образом. На входной границе моделировался набегающий поток воздуха с постоянной скоростью  $V=1\text{ м/с}$ , направленной перпендикулярно оси каньона и линии источника. Такой сценарий соответствует наихудшим условиям вентиляции и способствует накоплению примеси, что важно для оценки защитной эффективности барьеров. Концентрация загрязняющего вещества на входе принималась равной нулю. Источник этилена задавался через эквивалентную скорость подачи, равную  $0.01923\text{ м/с}$ . На выходной границе использовались условия свободного истечения для давления и нулевых градиентов для скорости и концентрации. Все твердые поверхности (здания, дорожное полотно, барьеры) моделировались как стенки с условием прилипания для скорости и нулевым нормальным диффузионным потоком для концентрации. Плотность воздуха принималась постоянной ( $\rho = 1.225\text{ кг/м}^3$ ).

#### Численная дискретизация и расчетная сетка

Решение системы управляющих уравнений выполнялось методом конечных объемов. Конвективные члены в уравнениях переноса импульса и примеси аппроксимировались схемой второго порядка  $crwind$ , что позволило минимизировать численную диффузию и корректно описать перенос в зонах с высокой турбулентностью. Для давления применялась стандартная схема второго порядка. Связь полей скорости и давления осуществлялась с помощью алгоритма SIMPLE, обеспечивающего устойчивую итерационную процедуру для несжимаемых течений. Для проведения гидродинамических расчетов был использован ANSYS Fluent.

Дискретизация расчетной области выполнялась с использованием неструктурированной вычислительной сетки, сформированной из элементов додекаэдрической формы (рисунок 2). Повышенная плотность сетки задавалась в зонах, определяющих аэродинамику и перенос примеси: вблизи дорожного покрытия, линейного источника выбросов, поверхностей зданий и шумозащитного барьера. Общее количество ячеек в наиболее детализированной конфигурации составило порядка  $1.2 \times 10^6$ , что обеспечило баланс между точностью решения и вычислительными затратами.

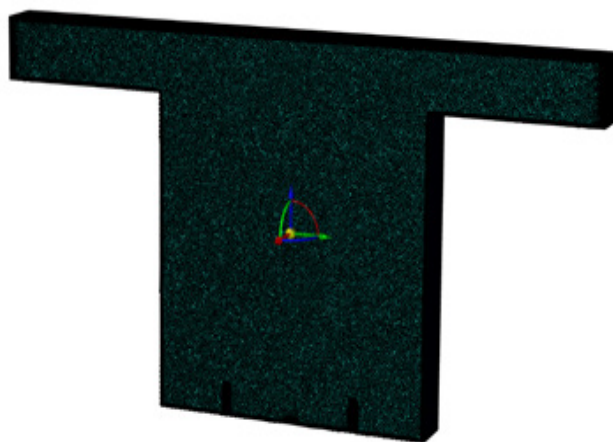


Рисунок 2 – Вычислительная сетка

Независимость численного решения от размера вычислительной сетки оценивалась путем сравнения результатов расчетов, выполненных на нескольких уровнях дискретизации (таблица 1) и экспериментальными данными Kikumoto & Ooka (2017) [14]. Анализ проводился на основе распределений продольной скорости и концентрации загрязняющего вещества в характерных сечениях уличного каньона. Различия между результатами для наиболее детализированных сеток находились в пределах численной погрешности (рисунок 3), что позволяет считать выбранную сетку достаточной для дальнейших расчетов.

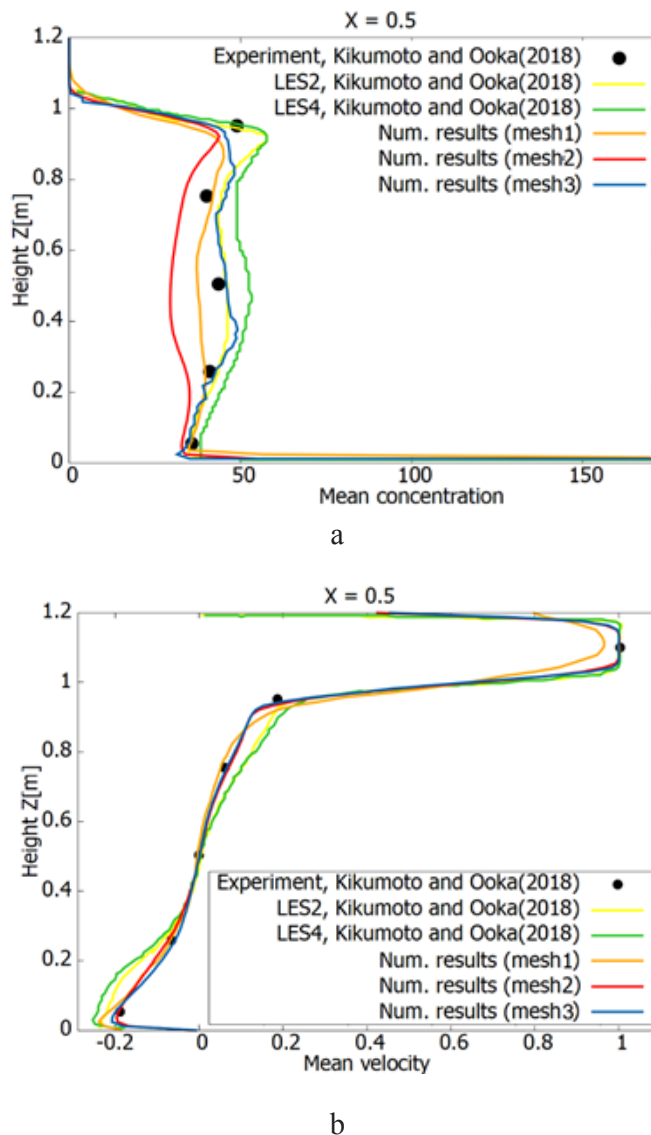


Рисунок 3 – Сравнение численных результатов с экспериментальными данными [14] для:  
а) усредненная скорость; б) усредненная концентрация

Таблица 1 – Параметры вычислительной сетки

Сетка №	Минимальный размер элемента [м]	Максимальный размер элемента [м]	Количество элементов
1	0.0125	0.0625	8763
2	0.003125	0.015625	227030
3	0.001	0.01	1195547

### Суррогатная модель на основе BiLSTM: данные, архитектура и обучение

Для построения суррогатной модели, предназначенной для прогнозирования временной эволюции концентрации загрязняющих веществ на основе данных CFD-моделирования, использовалась нейросетевая архитектура двунаправленной долгой краткосрочной памяти (BiLSTM), дополненная механизмом внимания. Выбор данной архитектуры обусловлен необходимостью учитывать временные зависимости различной протяженности, характерные для нестационарного переноса примеси в уличных каньонах, а также способностью механизма attention адаптивно выделять наиболее значимые участки временных последовательностей при формировании прогноза.

#### Формирование обучающего набора данных

Обучающая выборка формировалась на основе результатов нестационарных CFD-расчетов, выполненных для четырех геометрических конфигураций уличного каньона: без шумозащитного барьера и с барьерами относительной высоты  $0.1H$ ,  $0.2H$  и  $0.3H$ . Для каждой конфигурации регистрировались временные ряды концентрации этилена в пяти контрольных точках, расположенных в характерных зонах внутри каньона, включая области рециркуляции и основной поток (рисунок 4).

Продолжительность моделирования составляла 200 с. Шаг по времени принимался равным  $0.25$  с для всех конфигураций, что обеспечивало достаточное временное разрешение для анализа динамики концентрации при различных режимах обтекания. Сохранение численных результатов выполнялось с интервалом  $0.5$  с.

На основе исходных временных рядов формировались обучающие примеры в виде последовательностей типа «прошлое–будущее». Входные данные представляли собой фрагменты временного ряда фиксированной длины, соответствующие предыдущим значениям концентрации, тогда как выходными данными являлись значения концентрации в последующие моменты времени. Такая организация данных позволила сформулировать задачу как прогнозирование временной эволюции концентрации на основе ее предыстории.



Рисунок 4 – Расположение контрольных вычислительных узлов в области вычисления

Сформированный набор данных (16 000 временных точек) был разделен на обучающую и тестовую выборки в соотношении 90/10 без перемешивания временных последовательностей. Формирование набора данных основывалась в таком виде ( $800$  точек  $\times$   $5$  (контрольные точки)  $\times$   $4$  (варианта вычисления) =  $16\ 000$ ). Сохранение временного порядка обеспечивало корректную оценку прогностических свойств модели и исключало утечку информации между выборками.

В рамках данного исследования для каждой конфигурации (без барьера,  $0.1H$ ,  $0.2H$  и  $0.3H$ ) обучалась отдельная модель машинного обучения. Такой подход обусловлен существенными

различиями в структуре потока и характере распределения загрязняющих веществ при изменении геометрии области. Обучение отдельных моделей позволяет избежать смещения различных аэродинамических режимов в одном наборе данных и обеспечивает более корректное воспроизведение зависимостей между входными параметрами и концентрацией загрязняющего вещества. Кроме того, раздельное обучение моделей повышает точность предсказаний, поскольку каждая модель адаптируется к специфике конкретной конфигурации потока. Сравнение качества предсказаний выполнялось независимо для каждой модели с использованием метрик MAE, MAPE и коэффициента детерминации ( $R^2$ ).

Архитектура нейронной сети и обоснование ее проектирования

Суррогатная модель построена на основе двунаправленной рекуррентной архитектуры LSTM, предназначенной для аппроксимации нестационарной временной эволюции концентрации загрязняющего вещества. Выбор архитектуры LSTM обусловлен рядом особенностей рассматриваемой задачи. Нестационарный перенос загрязняющих веществ в уличном каньоне характеризуется сложными временными зависимостями различной протяженности: краткосрочными флуктуациями концентрации, вызванными турбулентными пульсациями, и долгосрочными тенденциями, связанными с эволюцией вихревых структур. Рекуррентные ячейки LSTM, в отличие от стандартных RNN, устойчиво улавливают такие зависимости благодаря механизму вентиля, предотвращающему затухание градиентов при обучении на длинных последовательностях [23].

Использование двунаправленной конфигурации дополнительно обосновано спецификой задачи суррогатного моделирования: в отличие от задач онлайн-прогнозирование, где будущие значения недоступны, суррогатная модель обучается на полных временных последовательностях CFD-расчетов, что делает двунаправленную обработку физически оправданной и позволяет модели формировать более полное внутреннее представление динамики процесса. Дополнительный механизм внимания обеспечивает адаптивное взвешивание временных шагов, что особенно важно в условиях нестационарного режима течения с резкими изменениями концентрации. Размерность скрытого состояния составляет 256 нейронов в каждом направлении, что обеспечивает достаточную выразительную способность модели без избыточного увеличения числа параметров.

Последовательность скрытых состояний, формируемая двунаправленным LSTM-слоем (рисунок 5), далее обрабатывается механизмом внимания, предназначенным для адаптивного перераспределения весов между временными шагами. Механизм внимания реализован как параметрическое отображение скрытых состояний во временные коэффициенты значимости с последующей нормализацией. Полученные коэффициенты используются для вычисления контекстного вектора в виде взвешенной суммы скрытых состояний, который аккумулирует наиболее информативные элементы временной последовательности. Контекстный вектор поступает на выходной полносвязный слой, формирующий прогноз концентрации загрязняющего вещества на заданном временном горизонте.

Данная архитектура обеспечивает баланс между аппроксимационной способностью и интерпретируемостью модели. Использование внимания позволяет не только повысить точность прогноза, но и проанализировать, какие участки временной эволюции оказывают наибольшее влияние на формирование результата, что представляет интерес с точки зрения физической интерпретации нестационарных процессов переноса примеси.

В качестве основной архитектуры использовалась двунаправленная рекуррентная нейронная сеть BiLSTM, дополненная механизмом внимания. В данной работе применен механизм additive attention (Bahdanau attention), который позволяет учитывать вклад скрытых состояний на каждом временном шаге при формировании итогового прогноза. Такой подход обеспечивает более гибкое взвешивание информации по временной последовательности по сравнению с простыми схемами усреднения.

Входные данные формировались в виде временных окон фиксированной длины. Размер входного окна (look-back) составлял ( $L = 20$ ) временных шагов, что обеспечивало достаточный контекст для учета динамики изменения концентрации. Горизонт прогноза (forecast horizon) был выбран равным ( $H = 5$ ) временным шагам, что соответствует краткосрочному прогнозированию концентрации загрязняющего вещества.

В качестве входного вектора использовались многомерные временные ряды, включающие значения концентрации во всех 9 контрольных точках одновременно. Таким образом, размерность входного вектора на каждом временном шаге составляла 9. Это позволило модели учитывать пространственные корреляции между различными точками наблюдения, что особенно важно в условиях сложной аэродинамики городского каньона.

Подбор гиперпараметров модели выполнялся с использованием метода grid search с последующей локальной донастройкой. В процессе оптимизации варьировались следующие параметры: количество слоев BiLSTM (1–3), размер скрытого слоя (32–128 нейронов), размер батча (16–64), скорость обучения ( $10^{-4} - 10^{-2}$ ), а также параметры механизма внимания. Критерием выбора оптимальной конфигурации служило минимальное значение функции потерь на валидационной выборке (MSE). Дополнительно применялась ранняя остановка (early stopping) для предотвращения переобучения. Оптимальная конфигурация для расчетов была выбрана как 2 слоя BiLSTM, размер скрытого слоя 128 нейронов, размер батча 32, скорость обучения  $10^{-3}$ . Финальная модель обучалась с указанными параметрами; ранняя остановка применялась при отсутствии улучшения MSE на валидации в течение 30 эпох.

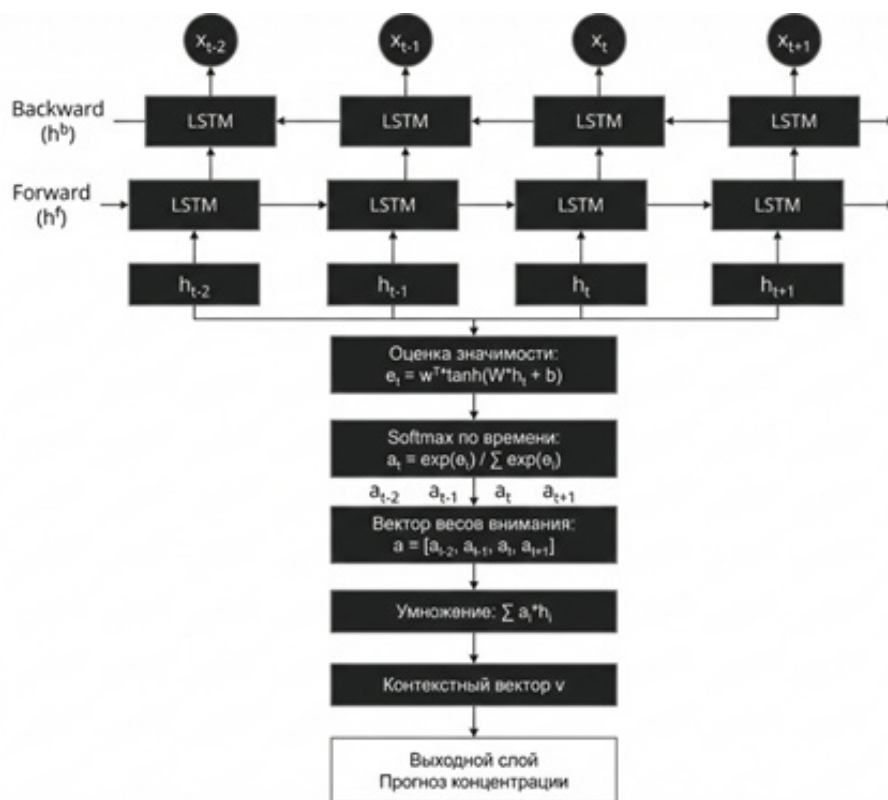


Рисунок 5 – Bi-LSTM модель с механизмом внимания

#### Обучение и регуляризация

Обучение нейросетевой модели выполнялось с использованием оптимизатора RAdam. Его применение позволило избежать неустойчивого поведения градиентов на начальной стадии обучения и обеспечить более предсказуемую сходимость без дополнительной подстройки

гиперпараметров. Для сглаживания траектории оптимизации и повышения устойчивости получаемого решения использовался алгоритм Lookahead, в рамках которого обновления параметров осуществлялись по схеме чередования быстрых и медленных весов.

В качестве функции потерь было выбрано среднеквадратичное отклонение с временными весами, смещающими акцент ошибки в сторону начальных шагов прогнозной последовательности. Такое взвешивание позволило улучшить качество краткосрочного прогноза, не влияя на структуру модели и не вводя дополнительных ограничений в процесс обучения.

Для снижения риска переобучения после рекуррентного слоя применялся dropout с коэффициентом 0.2. Обучение контролировалось с использованием механизма ранней остановки: процесс оптимизации прекращался при отсутствии улучшения значения функции потерь на валидационной выборке в течение 30 последовательных эпох. Этот критерий оказался достаточным для стабилизации параметров модели без преждевременного прекращения обучения. Все расчеты выполнялись с использованием фреймворка PyTorch. Обучение модели осуществлялось с аппаратным ускорением на графическом процессоре с поддержкой CUDA.

## Результаты и обсуждение

Влияние высоты барьера на аэродинамику и распределение концентрации

Нестационарные CFD-расчеты позволили проследить временную эволюцию концентрации загрязняющего вещества в уличном каньоне при различных конфигурациях шумозащитного барьера. Во всех рассмотренных случаях формирование концентрационного поля носило выражено нестационарный характер и определялось взаимодействием основного поперечного потока с вихревыми структурами внутри каньона и в приповерхностной зоне.

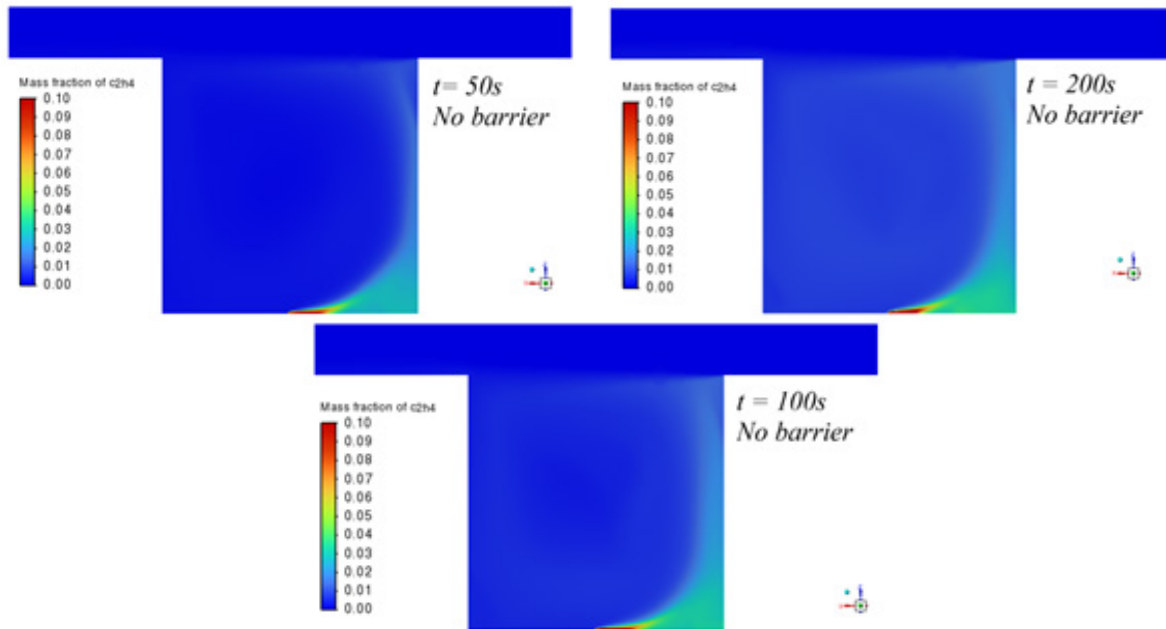
Введение барьера вдоль линии источника кардинально изменяет структуру течения внутри каньона. При отсутствии барьера (рисунок 6а) основной вихрь занимает почти весь объем каньона, обеспечивая относительно однородное перемешивание. Линейный источник создает шлейф примеси, который захватывается этим вихрем и распределяется по всей высоте подветренной зоны.

Барьер высотой 0.1H (рисунок 6б) не разрушает основной вихрь, но создает локальную зону застоя непосредственно перед препятствием. Поток разделяется, огибая барьер сверху, что приводит к формированию компактного вторичного вихря между барьером и подветренной стеной. Концентрация примеси резко возрастает в этой замкнутой зоне, создавая локальный максимум вблизи точки 5 (согласно нумерации на рис. 4). Напротив, за барьером (точка 3) концентрация остается низкой, что подтверждает его экранирующую функцию.

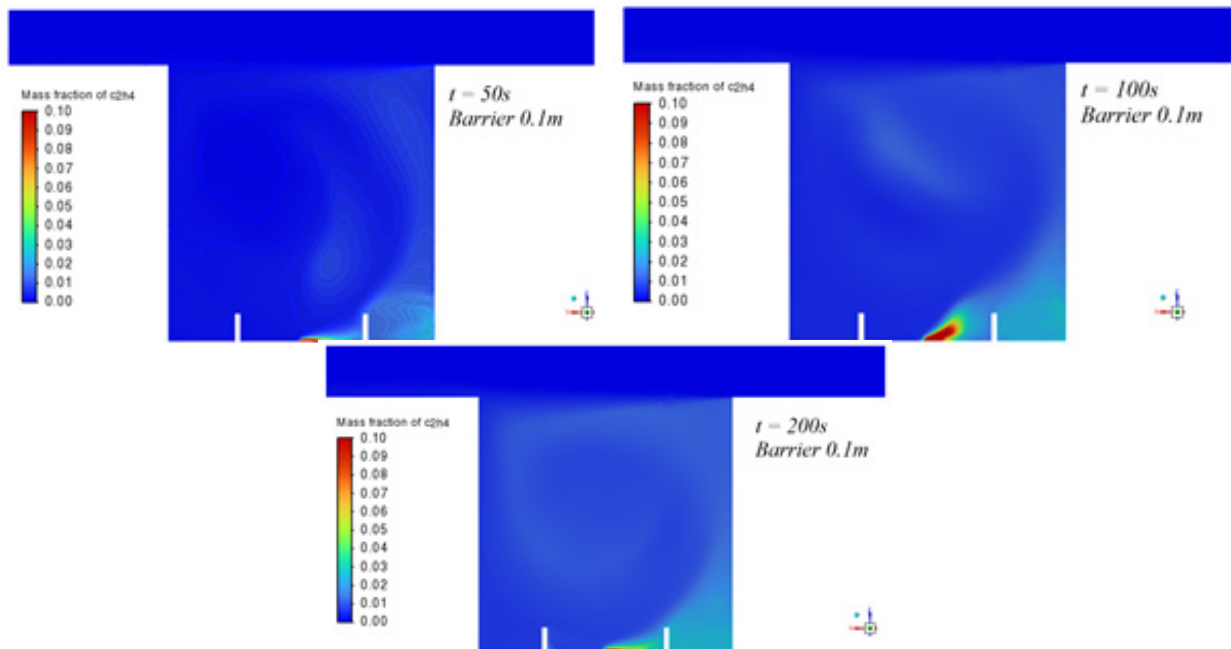
Увеличение высоты барьера до 0.2H (рис. 6в) усиливает этот эффект. Основной вихрь деформируется, его центр смещается вверх. Пространство между барьером и подветренной стеной становится областью сложной турбулентной структуры с несколькими мелкими вихрями. Динамика переноса примеси здесь становится существенно нестационарной, а распределение концентрации – менее предсказуемым. Пиковые значения концентрации в зоне между барьером и стеной возрастают.

При высоте барьера 0.3H (рисунок 6г) аэродинамическая картина снова упорядочивается. Барьер эффективно разделяет поток на два канала. Большая часть потока перенаправляется над барьером, создавая область пониженного давления за ним и устойчивую рециркуляцию. В этой более стабильной, хотя и интенсивной, вихревой зоне концентрация примеси остается высокой, но ее временная изменчивость снижается по сравнению со случаем 0.2H.

Полученные CFD-результаты сформировали основу для построения и валидации суррогатной модели машинного обучения, позволяя оценивать не только средние концентрации, но и временную динамику загрязняющего вещества в характерных зонах уличного каньона.



a



б

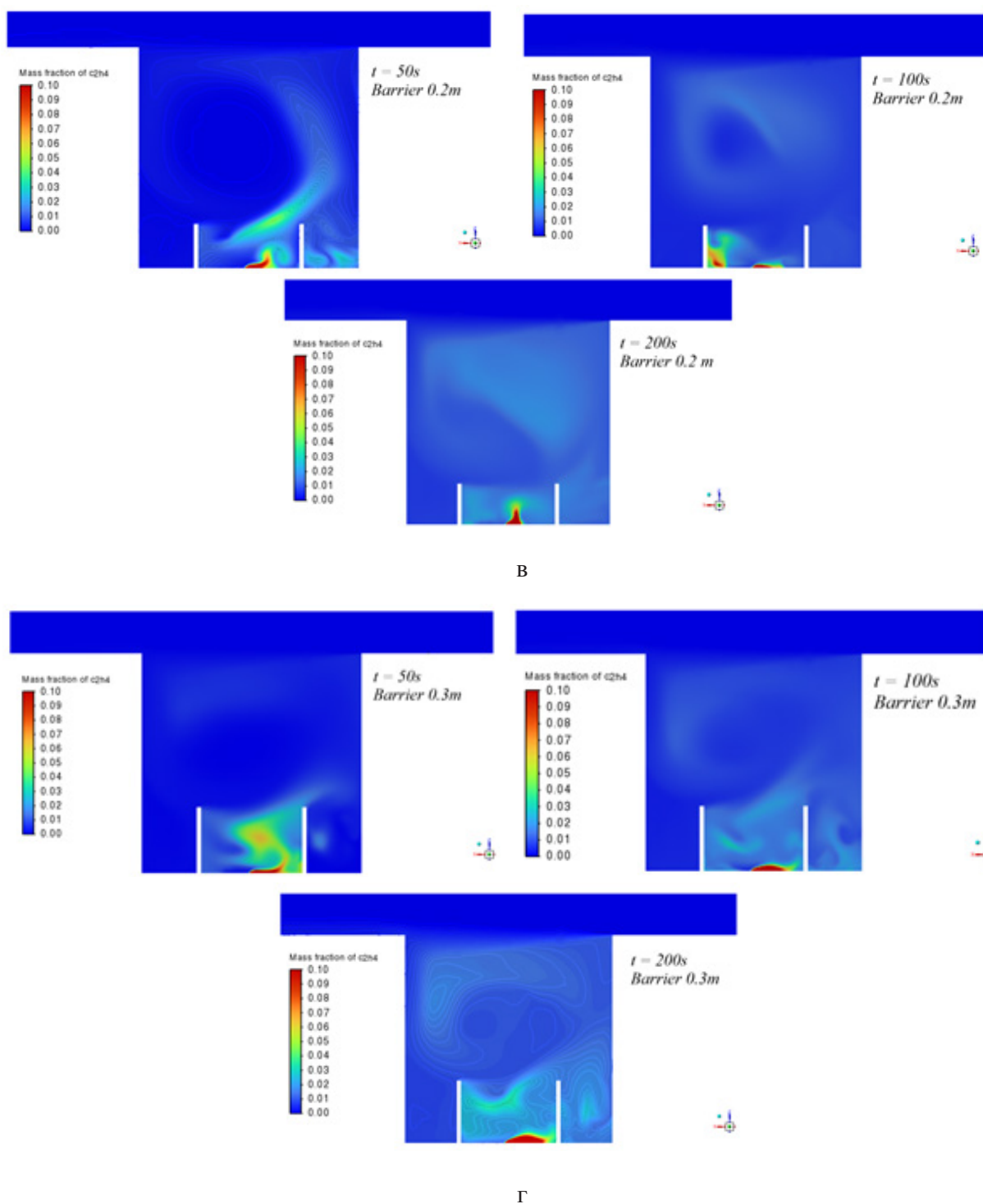


Рисунок 6 – Распределение поля концентрации вещества в различные моменты времени:  
а) без барьера, б) барьер 0.1H, в) барьер 0.2H, г) барьер 0.3H

#### Прогнозирование концентраций с помощью суррогатной модели BiLSTM

Качество работы суррогатной модели оценивалось путем прямого сопоставления прогнозируемых и расчетных по CFD временных рядов концентрации в контрольных точках для всех рассмотренных конфигураций. На рисунке 7 представлены типичные временные зависимости

концентрации, полученные из CFD-расчетов и предсказанные моделью машинного обучения, для различных положений внутри уличного каньона.

Во всех конфигурациях модель корректно воспроизводит общий уровень концентрации и характер временной изменчивости сигнала. Основные особенности нестационарной динамики, включая фазы нарастания концентрации и последующие квазипериодические колебания, совпадают с CFD-результатами по фазе и амплитуде.

Вблизи дорожного полотна и в областях, прилегающих к шумозащитному барьеру, расхождения между CFD и ML-прогнозами увеличиваются. В этих зонах модель демонстрирует сглаживание резких пиков концентрации, что особенно заметно на начальных этапах моделирования и при переходных режимах. Несмотря на это, средний уровень концентрации и общая форма временного сигнала сохраняются с приемлемой точностью.

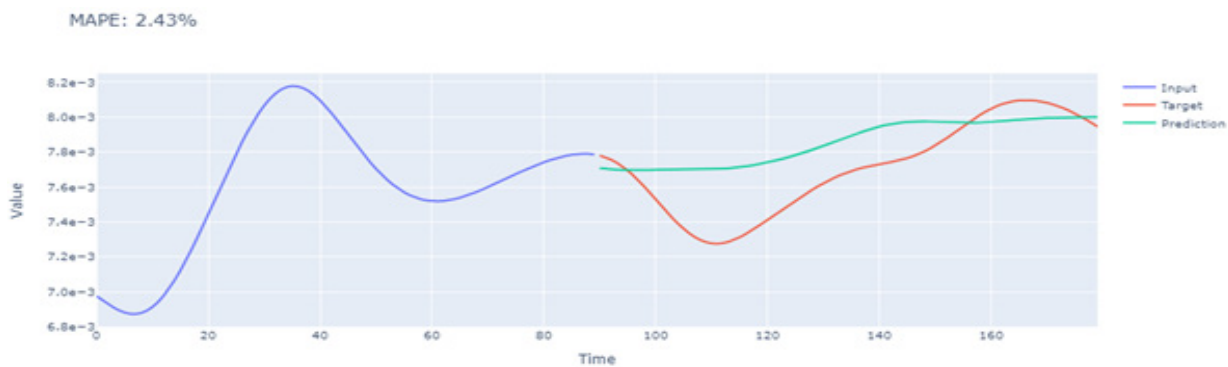
Для случая отсутствия барьера модель демонстрирует исключительно высокую точность. Временные ряды прогнозируемой концентрации практически идентичны исходным данным, визуальные расхождения минимальны. Это подтверждается низкими значениями метрик ошибок: средняя абсолютная процентная ошибка (MAPE) не превышает 6.5%, а коэффициент детерминации  $R^2$  остается положительным для всех точек, достигая 0.6 в точках с более выраженной динамикой (таблица 2). Модель успешно улавливает как общий тренд, связанный с выходом на стационарный режим, так и мелкомасштабные флуктуации, вызванные турбулентностью.

Аналогичная ситуация наблюдается при введении низкого барьера (0.1H). Несмотря на усложнение структуры потока, модель обеспечивает стабильный и точный прогноз. Наибольшие ошибки, закономерно, фиксируются в точках 1 и 2, расположенных в непосредственной близости от барьера в зоне интенсивного вихреобразования (MAPE до 14.7%). Однако даже здесь модель корректно предсказывает момент резкого роста концентрации и ее последующее осциллирующее поведение.

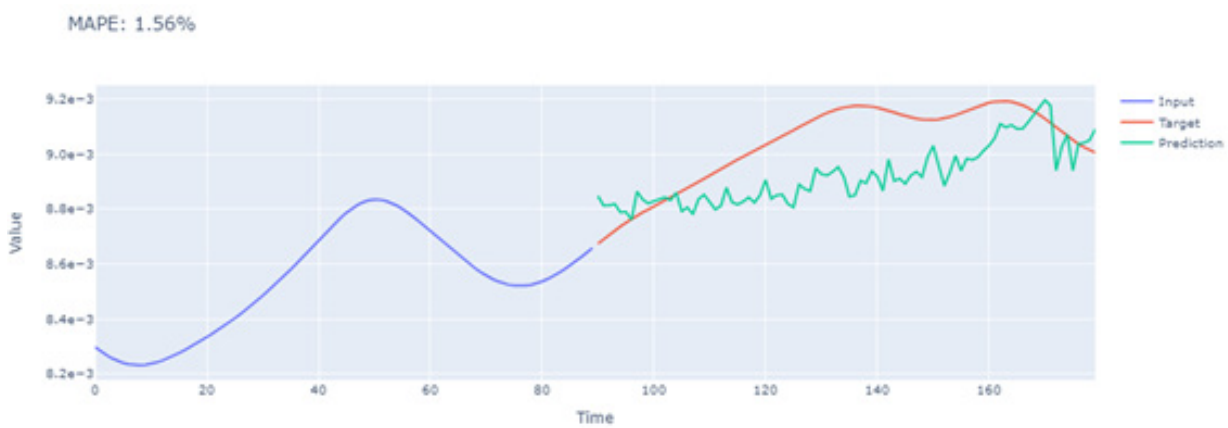
Качественный скачок в ошибке прогноза происходит при увеличении высоты барьера до 0.2H. На графиках для точек 1 и 2 (рисунок 7 (к) и рисунок 7 (л)) становятся заметны выраженные расхождения: модель запаздывает с реакцией на резкие изменения концентрации или генерирует сглаженный прогноз, не улавливая пиковых значений. Это напрямую связано с описанной выше сложной и нестационарной аэродинамикой данного случая. Процессы интенсивного перемешивания и быстрого изменения полей концентрации формируют временные ряды с высокочастотными составляющими, которые трудно предсказать на основе ограниченной истории, особенно когда модель обучена на данных, где такие экстремальные режимы представлены недостаточно полно.

При дальнейшем увеличении высоты барьера до 0.3H точность прогноза частично восстанавливается. Хотя абсолютный уровень концентраций в зоне за барьером высок, их временная изменчивость снижается благодаря стабилизации вихревой структуры. Это позволяет модели, обученной на осредненных закономерностях, давать более устойчивые прогнозы. Значения MAPE для точек 1 и 2 снижаются по сравнению со случаем 0.2H, а  $R^2$  возвращается в положительную область.

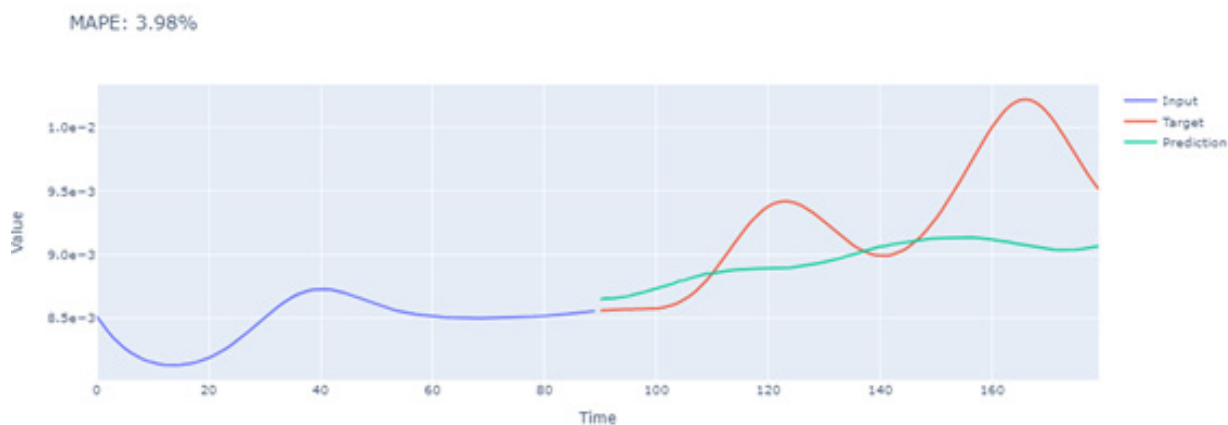
Модель обучалась на всех 5 точках одновременно – и она нашла компромиссное решение, которое хорошо работает для большинства точек (MAPE < 15%), но систематически ошибается именно в точке 2. Данное отклонение можно объяснить классической проблемой multi-output регрессии – оптимизация усредненной loss-функции, что не гарантирует приемлемого качества для каждого выхода по отдельности. Точка 2 при 0.2H – физически наиболее аномальная зона, и она «переголосована» остальными 4 точками при обучении.



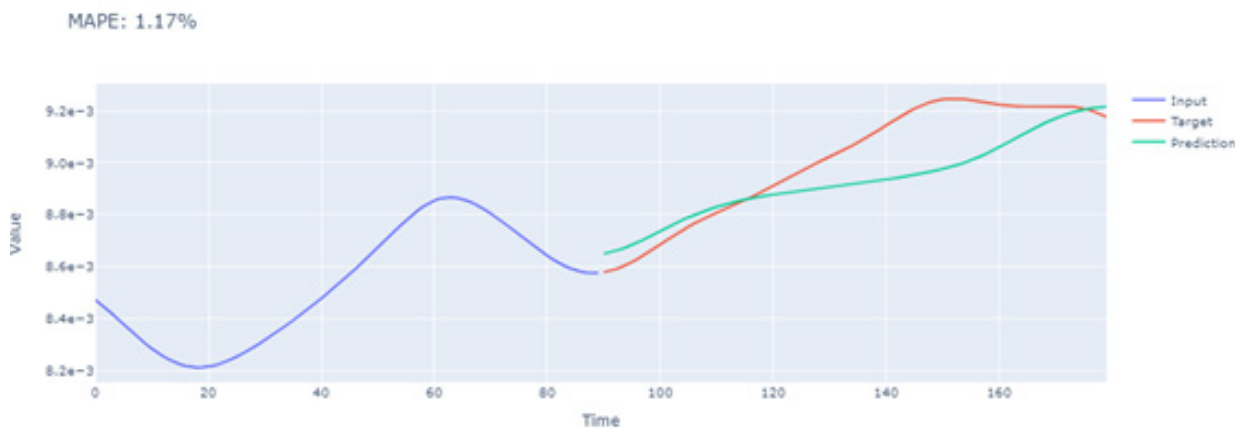
a) 0H barrier, point-1



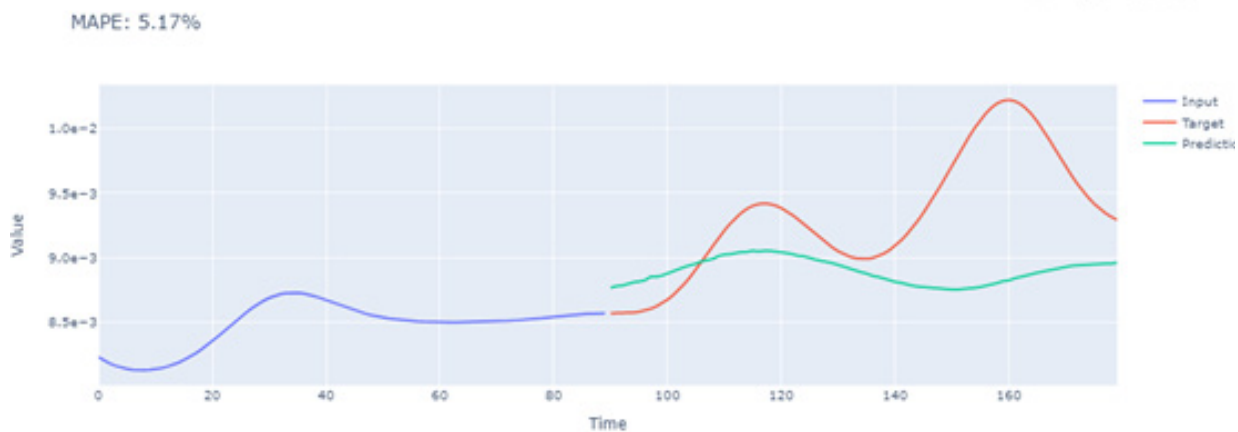
b) 0H barrier, point-2



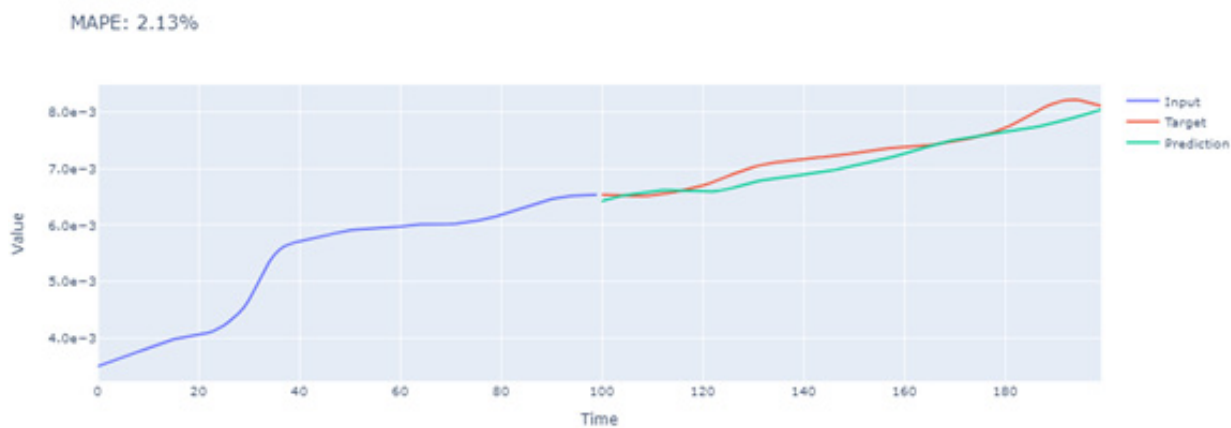
c) 0H barrier, point-3



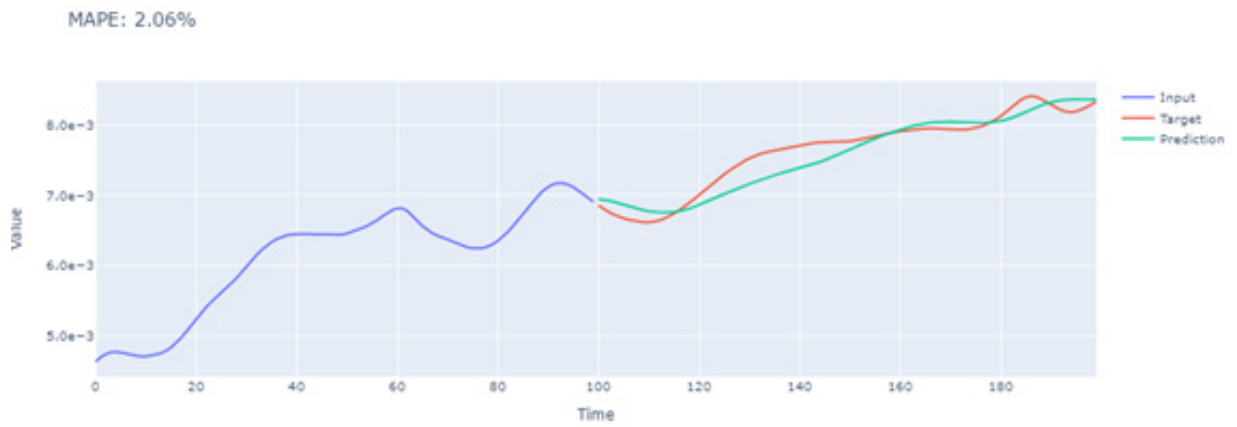
d) 0H barrier, point-4



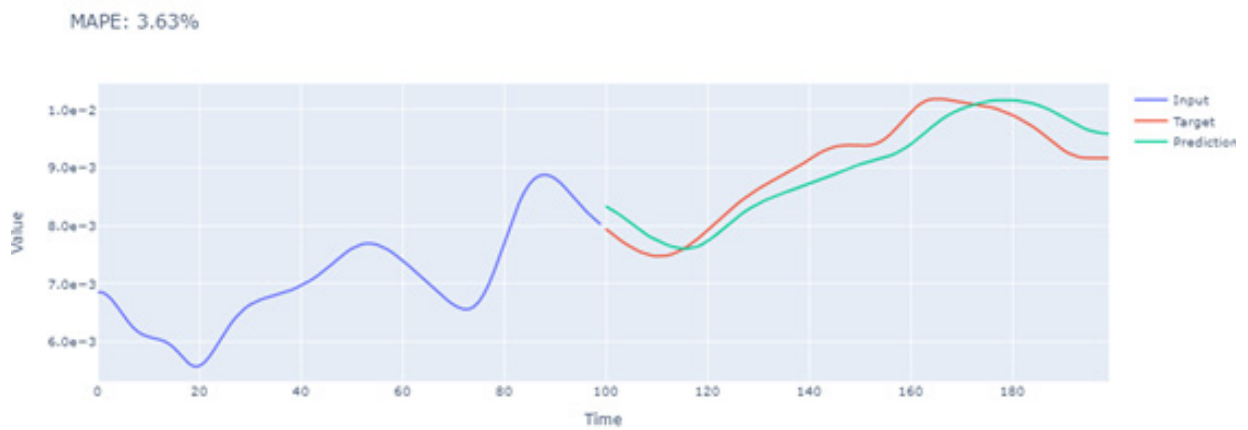
e) 0H barrier, point-5



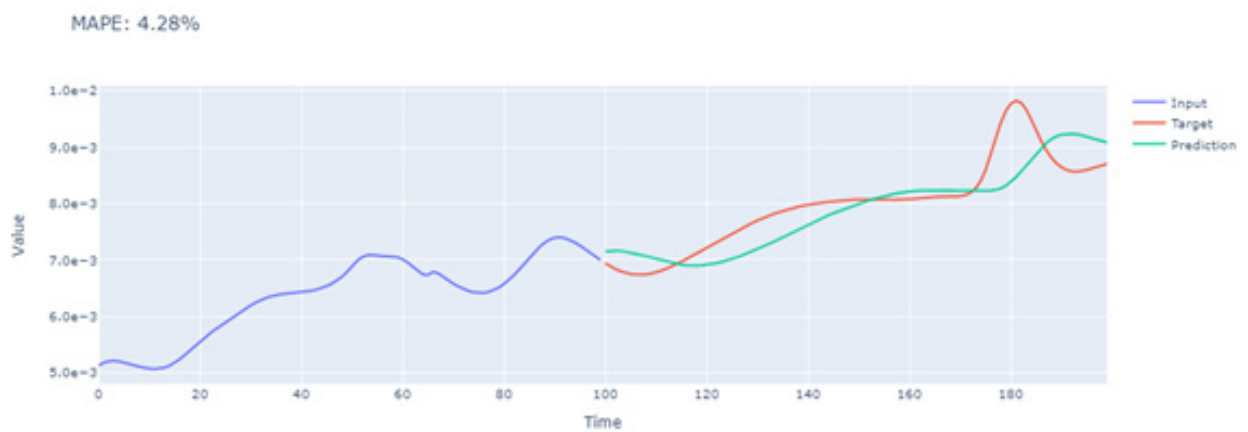
f) 0.1H barrier, point-1



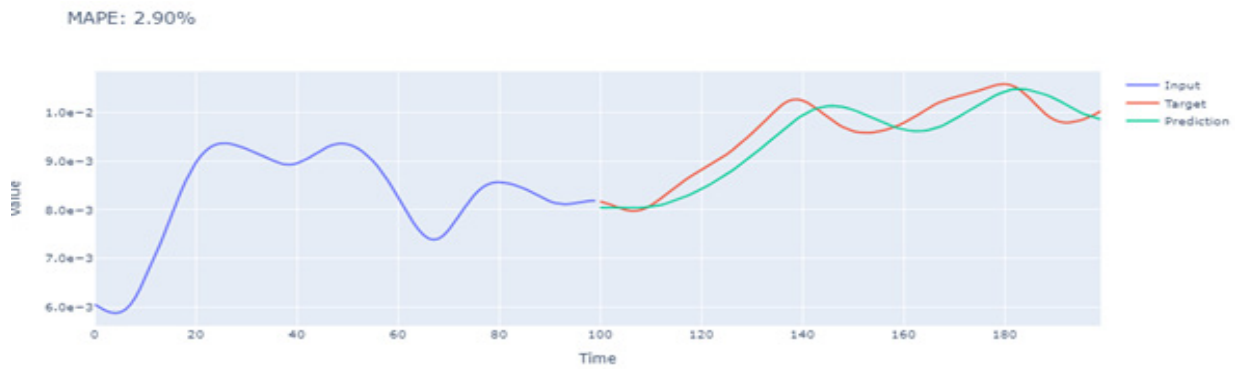
g) 0.1H barrier, point-2



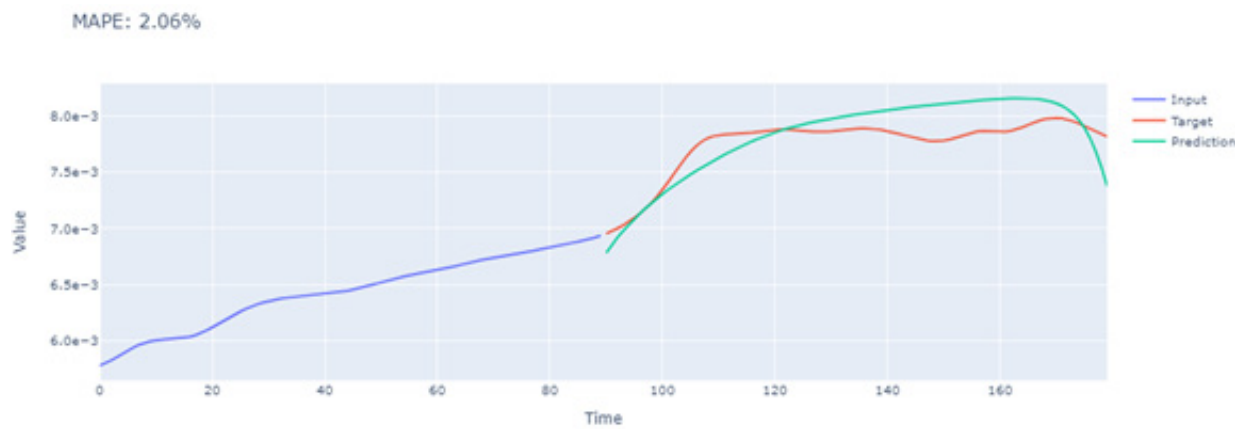
h) 0.1H barrier, point-3



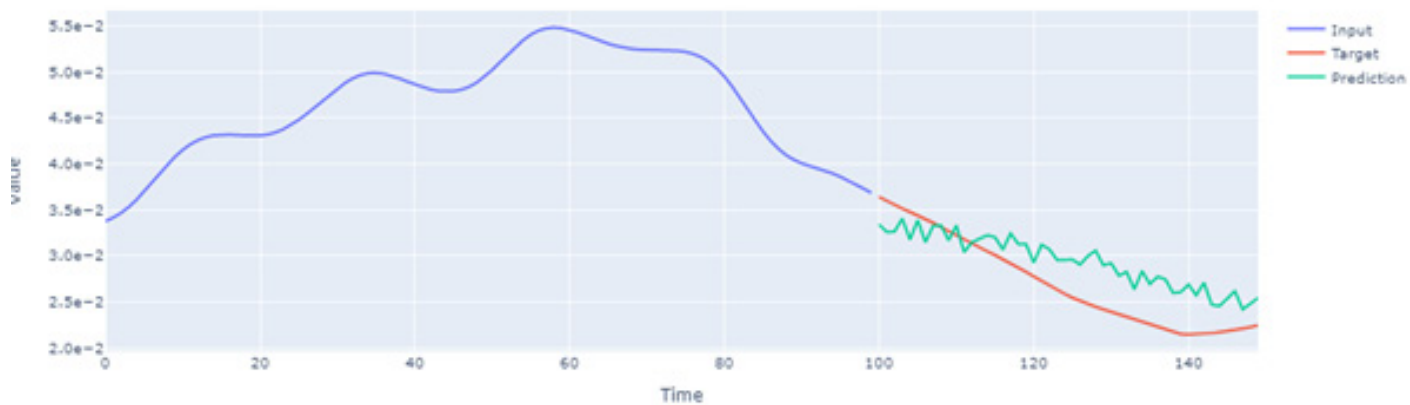
i) 0.1H barrier, point-4



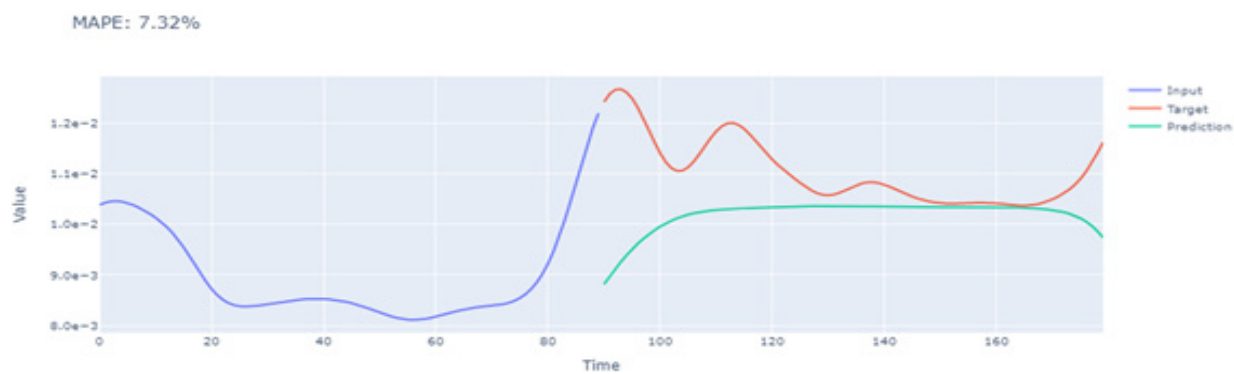
j) 0.1H barrier, point-5



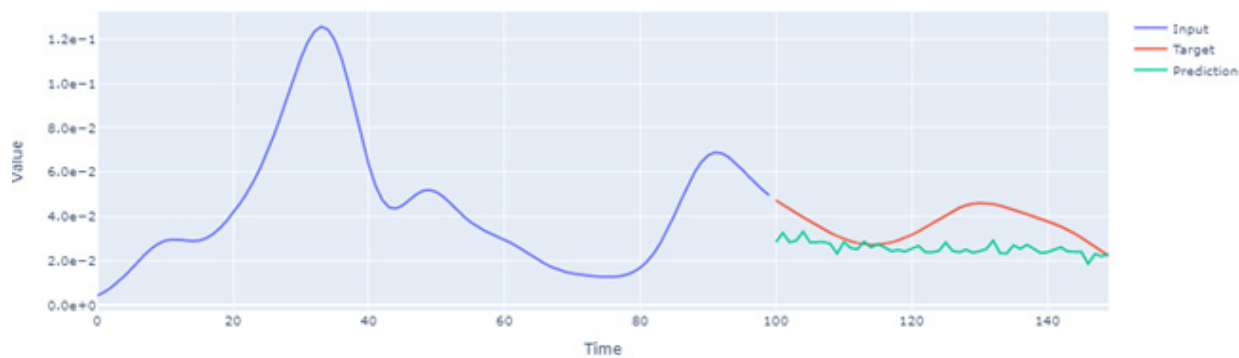
k) 0.2H barrier, point-1



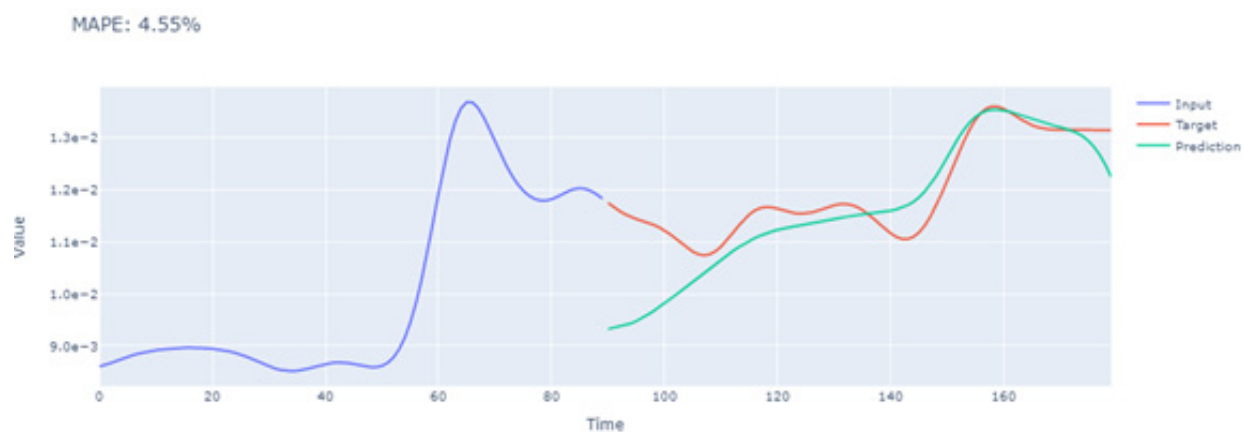
l) 0.2H barrier, point-2



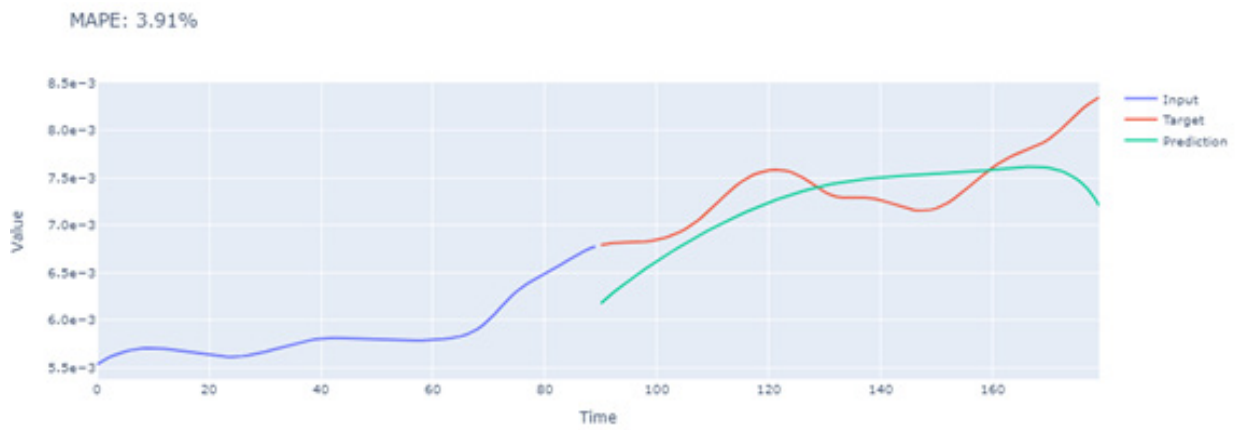
m) 0.2H barrier, point-3



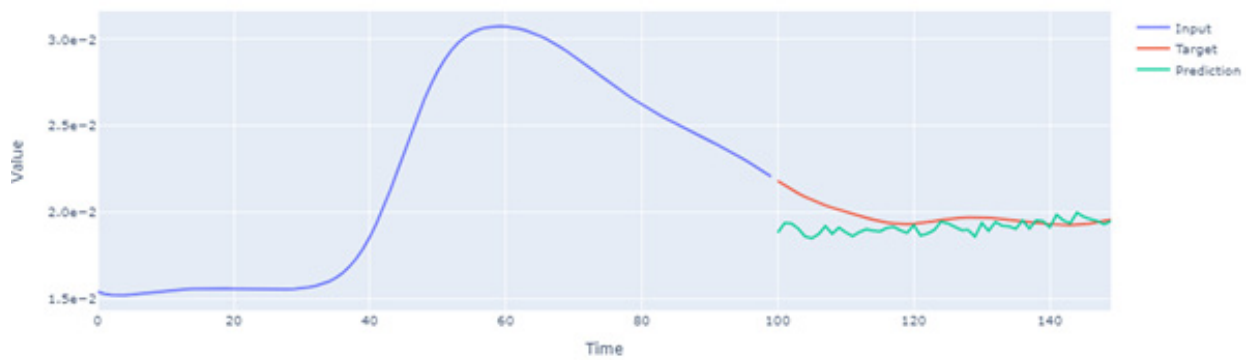
n) 0.2H barrier, point-4



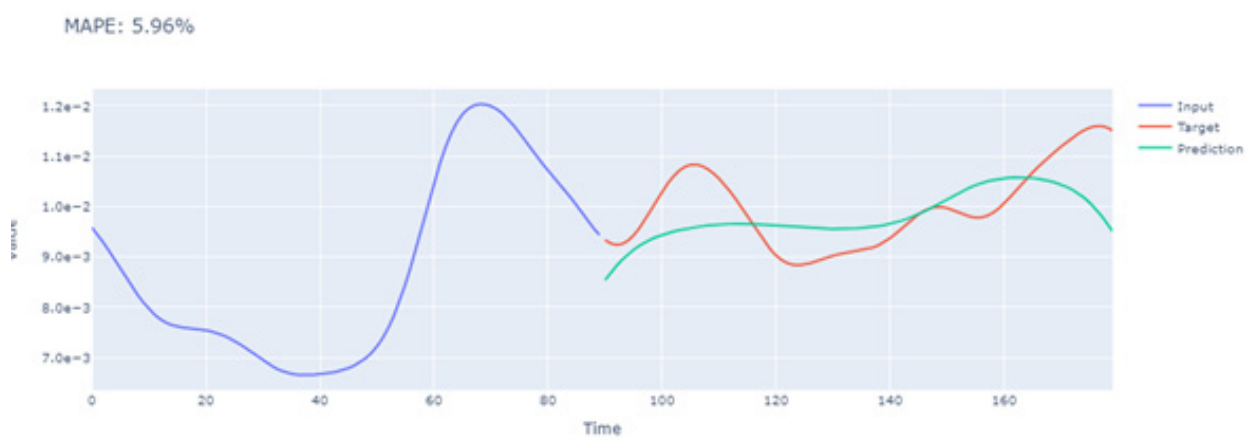
o) 0.2H barrier, point-5



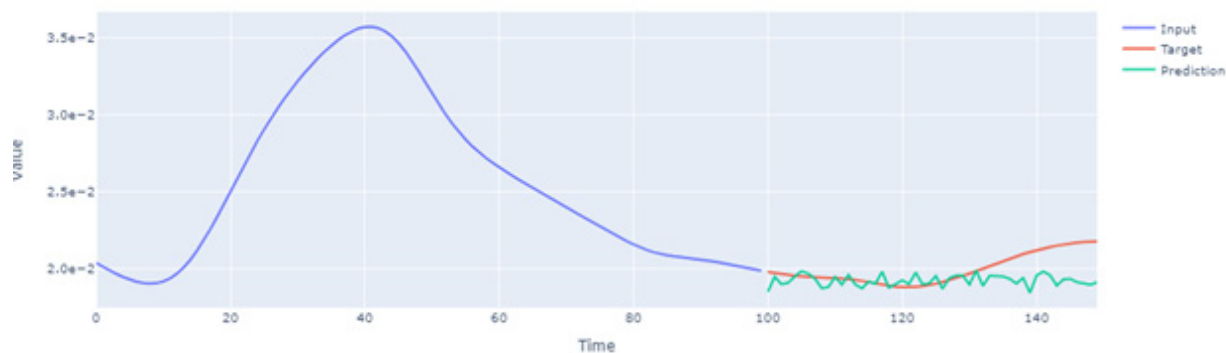
p) 0.3H barrier, point-1



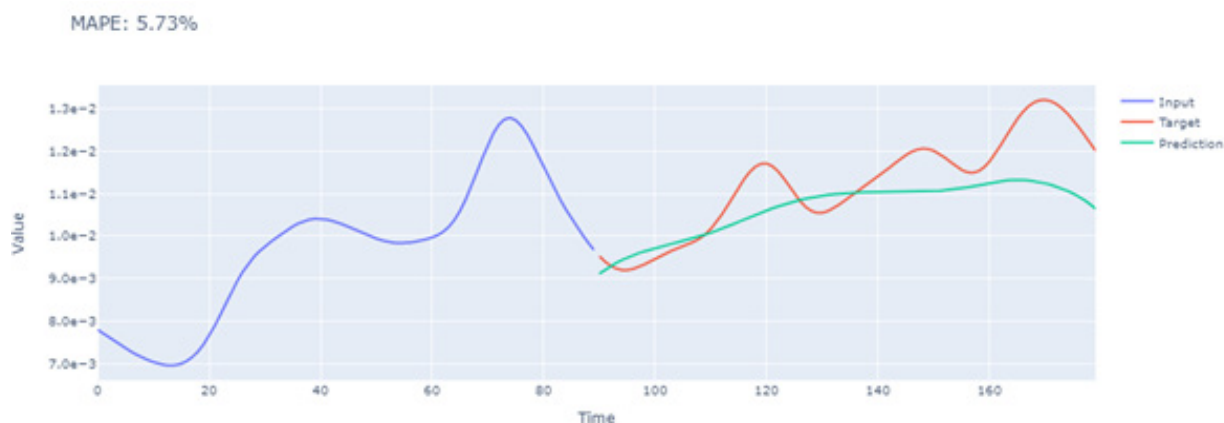
q) 0.3H barrier, point-2



r) 0.3H barrier, point-3



s) 0.3H barrier, point-4



t) 0.3H barrier, point-5

Рисунок 7 – Предсказания суррогатной модели для контрольных узлов сетки и различных высот барьера

#### Количественный анализ ошибок прогнозирования

Систематический анализ метрик MAE, MAPE и  $R^2$  (таблица 2) позволил выявить четкие пространственные закономерности в точности модели. Наибольшая чувствительность к изменению высоты барьера наблюдается в точках, непосредственно находящихся в аэродинамической тени барьера и в зонах отрыва потока (точки 1, 2, 3). Для барьера высотой 0.2H в точке 2 фиксируется минимальное значение  $R^2 = -4.5$ . Отрицательный коэффициент детерминации указывает, что простая константная модель (среднее значение) в данном случае дает более точный прогноз, чем построенная ViLSTM. Это явный признак того, что модель не смогла выявить физически обоснованную зависимость в данных, что вызвано чрезвычайной сложностью и, вероятно, неполной представительностью обучающей выборки для данного специфического режима течения.

Следует отметить, что коэффициент детерминации  $R^2$  для нестационарных временных рядов может давать нестабильные результаты. В контексте гидрологических и аэродинамических задач эквивалентной метрикой является коэффициент эффективности Нэша–Сатклиффа (NSE), который вычисляется по формуле, идентичной  $R^2$ , и численно совпадает с приведенными в таблице 2 значениями.

Коэффициент эффективности Нэша–Сатклиффа (NSE), вычисляемый по формуле, математически эквивалентной  $R^2$ , может принимать отрицательные значения, что указывает на то,

что простое использование среднего значения наблюдений обеспечивает более точный прогноз, чем рассматриваемая модель. Аналогичное поведение характерно для  $R^2$ , вычисленного по той же формуле, в частности в библиотеке scikit-learn. Значение  $NSE = -4.5$  в точке 2 для конфигурации барьера 0.2H свидетельствует о том, что модель не смогла выявить устойчивую зависимость в данных для указанного режима течения, что обусловлено экстремальной нестационарностью динамики концентрации в зоне аэродинамического влияния барьера.

В то же время точки, расположенные за барьером относительно направления ветра (точка 4), или на значительном удалении от зоны его непосредственного влияния (точка 5), демонстрируют гораздо более устойчивые результаты. Ошибки прогноза растут незначительно с увеличением высоты барьера, а значения ( $R^2$ ) остаются в допустимом диапазоне. Это свидетельствует о том, что модель успешно экстраполирует изученные закономерности переноса для относительно спокойных зон потока, даже при изменении глобальной конфигурации.

Интересной находкой является нелинейная зависимость ошибки от высоты барьера. Наиболее проблемной с точки зрения прогнозирования оказалась не максимальная (0.3H), а средняя (0.2H) высота. Это подтверждает гипотезу о том, что ключевым фактором, лимитирующим точность суррогатной ML-модели в данной задаче, является не абсолютный уровень сложности физического явления, а степень его нестационарности и наличие в данных резких, плохо коррелирующих изменений.

Таблица 2 – Метрики ошибок для модели BiLSTM

Средняя абсолютная процентная ошибка (MAPE)				
№	No barrier	0.1 barrier	0.2 barrier	0.3 barrier
1	2.08%	5.91%	10.70%	6.65%
2	1.80%	14.70%	17.56%	14.15%
3	3.22%	4.00%	12.60%	3.07%
4	1.81%	4.59%	3.07%	7.62%
5	2.70%	3.47%	3.99%	7.10%
Средняя абсолютная ошибка (MAE)				
1	0.0001984	0.000285	0.01561	0.00237
2	0.0001942	0.002661	0.0267	0.00683
3	0.0002086	0.0002132	0.003559	0.000168
4	0.0009714	0.001019	0.0001686	0.00222
5	0.0002340	0.0001147	0.0001681	0.000335
$R^2$ – метрика				
1	0.45	0.47	-0.15	0.26
2	0.32	0.52	-4.5	0.29
3	0.32	0.71	0.51	0.63
4	0.46	0.61	0.71	0.23
5	0.60	0.72	0.63	0.61

### Заключение

В ходе настоящего исследования была решена задача разработки и валидации гибридного подхода, сочетающего методы вычислительной гидродинамики (CFD) и машинного обучения (ML), для прогнозирования пространственно-временного распределения загрязняющих веществ в идеализированных городских уличных каньонах с учетом влияния инженерных элементов инфраструктуры – шумозащитных барьеров различной высоты.

CFD-моделирование позволило детально изучить аэродинамические особенности и процессы дисперсии примеси в различных геометрических конфигурациях. Установлено, что высота барьера оказывает существенное нелинейное влияние на структуру течения и характер

накопления загрязняющего вещества. В частности, барьер высотой  $0.2H$  создает наиболее сложный и нестационарный режим течения с интенсивным вихреобразованием, что приводит к значительным пространственно-временным флуктуациям концентрации.

На основе данных нестационарных CFD-расчетов была разработана суррогатная модель на основе архитектуры ViLSTM с механизмом внимания. Модель продемонстрировала высокую эффективность в прогнозировании временной эволюции концентрации в большинстве контрольных точек, особенно для случаев отсутствия барьера и при наличии барьера малой высоты ( $0.1H$ ). Средняя абсолютная процентная ошибка (MAPE) в этих конфигурациях не превышала 15%, а коэффициент детерминации  $R^2$  оставался положительным, что свидетельствует о хорошей обобщающей способности модели.

Однако при увеличении высоты барьера до  $0.2H$  точность прогноза снизилась, особенно в зонах непосредственного аэродинамического влияния барьера, где наблюдались резкие и слабо коррелированные изменения концентрации. В наиболее сложной точке (точка 2) модель показала отрицательное значение  $R^2$ , что указывает на недостаточную репрезентативность обучающих данных для данного режима и ограниченную способность архитектуры улавливать экстремально нестационарные процессы.

Тем не менее, для барьера высотой  $0.3H$  точность прогноза частично восстановилась благодаря стабилизации вихревой структуры и снижению временной изменчивости полей концентрации. Это подтверждает, что ключевым фактором, определяющим эффективность суррогатной ML-модели, является не абсолютная сложность физического явления, а степень его нестационарности и наличие в данных резких, плохо предсказуемых изменений.

Проведенное исследование подтвердило перспективность гибридного CFD–ML подхода для задач оценки и прогнозирования качества воздуха в городской среде. Предложенная методология позволяет сочетать физическую достоверность численного моделирования с вычислительной эффективностью машинного обучения, что открывает возможности для оперативного анализа сценариев, оптимизации городской инфраструктуры и поддержки принятия экологически ориентированных решений. Суррогатная модель обеспечивает ускорение прогнозирования по сравнению с CFD-расчетом на 7–8 порядков: время инференса составляет порядка 1–5 мс, тогда как одна CFD-симуляция занимает около 54 часов на CPU (32 потока, 128 GB RAM). Таким образом, предложенный гибридный подход позволяет проводить многовариантный анализ сценариев при вычислительных затратах, недостижимых для прямого численного моделирования.

В качестве направлений дальнейших исследований целесообразно расширение обучающей выборки за счет включения большего числа вариаций геометрических и метеорологических параметров, применение более сложных архитектур глубокого обучения, а также интеграция предложенного подхода в системы мониторинга и управления качеством городского воздуха в режиме, близком к реальному времени.

### Информация о финансировании

Работа поддержана грантом Министерства науки и высшего образования Республики Казахстан (BR28712901).

### REFERENCES

- 1 World Health Organization. Ambient (outdoor) air pollution. World Health Organization (2023). URL: [https://www.who.int/ru/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/ru/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
- 2 Turner, M.C., Andersen, Z.J., Baccarelli, A., Diver, W.R., Gapstur, S.M., Pope, C.A., Prada, D., Samet, J., Thurston, G., and Cohen, A. Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations. *CA: A Cancer Journal for Clinicians*, 70 (6), 460–479 (2020). <https://doi.org/10.3322/caac.21632>

- 3 Singh, J., Payra, S., and Verma, S. Air quality and human health. In: *Geospatial Analytics for Environmental Pollution Modeling*, pp. 317–352 (Cham: Springer, 2023). [https://doi.org/10.1007/978-3-031-45300-7\\_13](https://doi.org/10.1007/978-3-031-45300-7_13)
- 4 Hänninen, O., Lehtomäki, H., Korhonen, A., Kokkola, T., Hartikainen, A., Sippula, O., Haverinen-Shaughnessy, U., Leviäkangas, P., and Rumrich, I.K. Health risks related to air pollution by transport categories and vehicle types: Comparison by mortality indicators. *Environment International*, 202, 109657 (2025). <https://doi.org/10.1016/j.envint.2025.109657>
- 5 Lolli, S. Urban PM2.5 concentration monitoring: A review of recent advances in ground-based, satellite, model, and machine learning integration. *Urban Climate*, 63, 102566 (2025). <https://doi.org/10.1016/j.uclim.2025.102566>
- 6 Natterer, E.S., Rao, S.R., Tejada Lapuerta, A., Engelhardt, R., Hörl, S., and Bogenberger, K. Machine learning surrogates for agent-based models in transportation policy analysis. *Transportation Research Part C: Emerging Technologies*, 180, 105360 (2025). <https://doi.org/10.1016/j.trc.2025.105360>
- 7 Xia, F., Cheng, X., Lei, Z., Liu, Y., Wang, J., and Zhang, H. Heterogeneous impacts of local traffic congestion on local air pollution within a city: Utilizing taxi trajectory data. *Journal of Environmental Economics and Management*, 122, 102896 (2023). <https://doi.org/10.1016/j.jeem.2023.102896>
- 8 Santiago, J.L., Sanchez, B., Quaassdorff, C., de la Paz, D., Martilli, A., Martín, F., and Lechón, Y. Performance evaluation of a multiscale modelling system applied to particulate matter dispersion in a real traffic hot spot in Madrid (Spain). *Atmospheric Pollution Research*, 11 (1), 141–155 (2020). <https://doi.org/10.1016/j.apr.2019.10.001>
- 9 Lin, C., Wang, Y., Ooka, R., Flageul, C., Kim, Y., Kikumoto, H., Wang, Z., and Sartelet, K. Modeling of street-scale pollutant dispersion by coupled simulation of chemical reaction, aerosol dynamics, and CFD. *Atmospheric Chemistry and Physics*, 23, 1421–1436 (2023). <https://doi.org/10.5194/acp-23-1421-2023>
- 10 Vahidi, S., Yeganeh, M., and Ghasaban, M. The effect of the morphology of highly polluted urban areas neighborhood on suspended particles of PM2.5 and PM10 air pollutants (Case study: Ahvaz, Iran). *Energy Nexus*, 20, 100527 (2025). <https://doi.org/10.1016/j.nexus.2025.100527>
- 11 Fellini, S., Majumdar, D., Salizzoni, P., and van Reeuwijk, M. Three-dimensional pollutant dispersion in tree-lined urban canyons: Combined wind-tunnel and LES analysis. *Atmospheric Environment*, 367, 121748 (2026). <https://doi.org/10.1016/j.atmosenv.2025.121748>
- 12 Mousavi, D., Ghaemian, M., and Brinkerhoff, J. Multiscale CFD analysis of urban air pollution dome and ventilation enhancement via an urban chimney. *Atmospheric Environment*, 337, 120783 (2024). <https://doi.org/10.1016/j.atmosenv.2024.120783>
- 13 Huertas, J.I., and Prato, D.F. CFD modeling of near-roadway air pollution. *Environmental Modeling & Assessment*, 25 (1), 129–145 (2020). <https://doi.org/10.1007/s10666-019-09666-w>
- 14 Kikumoto, H., and Ooka, R. Large-eddy simulation of pollutant dispersion in a cavity at fine grid resolutions. *Building and Environment*, 127, 127–137 (2017). <https://doi.org/10.1016/j.buildenv.2017.11.005>
- 15 Reiminger, N., Jurado, X., Vazquez, J., Wemmert, C., and Meunier, G. Methodologies to assess mean annual air pollution concentration combining numerical results and wind roses. *Sustainable Cities and Society*, 59, 102221 (2020). <https://doi.org/10.1016/j.scs.2020.102221>
- 16 Martín, F., Janssen, S., Rodrigues, V., Sousa, J., Santiago, J.L., Rivas, E., Stocker, J., Jackson, R., Russo, F., Villani, M.G., Tinarelli, G., Barbero, D., José, R.S., Pérez-Camanyo, J.L., Santos, G.S., Bartzis, J., Sakellaris, I., Horváth, Z., Környei, L., Cuvelier, C. Using dispersion models at microscale to assess long-term air pollution in urban hot spots: A FAIRMODE joint intercomparison exercise for a case study in Antwerp. *Science of the Total Environment*, 925, 171761 (2024). <https://doi.org/10.1016/j.scitotenv.2024.171761>
- 17 Bi, J., Knowland, K.E., Keller, C.A., and Liu, Y. Combining machine learning and numerical simulation for high-resolution PM2.5 concentration forecast. *Environmental Science & Technology*, 56 (3), 1544–1556 (2022). <https://doi.org/10.1021/acs.est.1c05578>
- 18 Jurado, X., Reiminger, N., Benmoussa, M., Vazquez, J., and Wemmert, C. Deep learning methods evaluation to predict air quality based on computational fluid dynamics. *Expert Systems with Applications*, 203, 117294 (2022). <https://doi.org/10.1016/j.eswa.2022.117294>
- 19 Schulte, N., Snyder, M., Isakov, V., Heist, D., and Venkatram, A. Effects of solid barriers on dispersion of roadway emissions. *Atmospheric Environment*, 97, 286–295 (2014). <https://doi.org/10.1016/j.atmosenv.2014.08.026>

20 Venkatram, A., Isakov, V., Deshmukh, P., and Baldauf, R. Modeling the impact of solid noise barriers on near-road air quality. *Atmospheric Environment*, 141, 462–469 (2016). <https://doi.org/10.1016/j.atmosenv.2016.07.005>

21 Li, B., Qiu, Z., and Zheng, J. Impacts of noise barriers on near-viaduct air quality in a city: A case study in Xi'an. *Building and Environment*, 196, 107751 (2021). <https://doi.org/10.1016/j.buildenv.2021.107751>

22 Wu, Y., and Chen, H. Optimizing block morphology for reducing traffic pollutant concentration in adjacent external spaces of street canyons: A machine learning approach. *Building and Environment*, 242, 110587 (2023). <https://doi.org/10.1016/j.buildenv.2023.110587>

23 Mora-Mariano, D., and Flores-Tlacuahuac, A. Bayesian LSTM framework for the surrogate modeling of process engineering systems. *Computers & Chemical Engineering*, 181, 108553 (2023). <https://doi.org/10.1016/j.compchemeng.2023.108553>

**<sup>1\*</sup>Исатов А.А.,**

профессор, ORCID ID: 0000-0002-1937-8615,

\*e-mail: [alibek.issakhov@gmail.com](mailto:alibek.issakhov@gmail.com)

**<sup>2</sup>Нығметова Ә.,**

докторант, ORCID ID: 0009-0004-5395-4569,

e-mail: [a.nygmetova@kbtu.kz](mailto:a.nygmetova@kbtu.kz)

**<sup>1</sup>Абылкасымова А.Б.,**

қауымдастырылған профессор, ORCID ID: 0000-0002-5967-6959,

e-mail: [abylkassymova.aizhan@gmail.com](mailto:abylkassymova.aizhan@gmail.com)

**<sup>1</sup>Аханова Н.Е.,**

профессор, ORCID ID: 0000-0003-2767-8711,

e-mail: [n.akhanova@kbtu.kz](mailto:n.akhanova@kbtu.kz)

<sup>1</sup>Қазақстан-Британ техникалық университет, Алматы қ., Қазақстан

<sup>2</sup>Халықаралық ақпараттық технологиялар университеті, Алматы қ., Қазақстан

## ML-CFD МОДЕЛЬДЕУІН ҚОЛДАНА ОТЫРЫП, ҚАЛАЛЫҚ КАНЬОНДАРДАҒЫ ШЫҒАРЫНДЫЛАРДАН АУАНЫҢ ЛАСТАНУЫН БОЛЖАУ

### Аңдатпа

Мақалада шу кедергілерін ескере отырып, қалалық көше каньондарындағы ластаушы заттардың таралуын болжаудың гибриді тәсілі ұсынылған. Әдістеме егжей-тегжейлі CFD модельдеуін және назар аудару механизмі бар BiLSTM нейрондық желі архитектурасына негізделген суррогат модельді біріктіреді. 0,1H, 0,2H және 0,3H тосқауыл биіктігі бар конфигурациялар зерттелді. CFD есептеулері тосқауыл биіктігінің аэродинамикаға және ластаушы заттардың жиналу аймақтарының қалыптасуына сызықтық емес әсер ететінін көрсетті, ал ең күрделі стационарлық емес мінез-құлық 0,2H биіктікте байқалды. Суррогат модель кедергісіз және 0,1H тосқауыл жағдайларында концентрация эволюциясын сәтті болжады, орташа абсолюттік пайыздық қателіктің 15%-дан аз екенін көрсетті. Алайда 0,2H тосқауыл үшін процестің жоғары дәрежедегі стационарлық емес сипатына байланысты қарқынды турбуленттілік аймақтарында дәлдік төмендеді. Бұл тәсіл физикалық дәлдікті сақтай отырып, есептеу шығындарын айтарлықтай азайтады, сондықтан қалалық экологиядағы шешім қабылдауды қолдау жүйелері үшін перспективалы болып табылады. Модель CFD есептеулерімен салыстырғанда болжауды 7–8 есе жылдам қамтамасыз етеді: қорытындылау уақыты 1–5 мс болса, бір CFD модельдеуін орындау CPU-де шамамен 54 сағатты алады.

**Түйін сөздер:** ауаның ластануы, көше каньондары, шу кедергілері, есептеу сұйықтық динамикасы (CFD), машиналық оқыту, BiLSTM, суррогат модельдеу, концентрацияны болжау.

**<sup>1\*</sup>Issakhov A.A.,**

Professor, ORCID ID: 0000-0002-1937-8615,

\*e-mail: alibek.issakhov@gmail.com

**<sup>2</sup>Nygmetyova A.,**

PhD student, ORCID ID: 0009-0004-5395-4569,

e-mail: a.nygmetyova@kbtu.kz

**<sup>1</sup>Abylkassymova A.B.,**

Associate Professor, ORCID ID: 0000-0002-5967-6959,

e-mail: abylkassymova.aizhan@gmail.com

**<sup>1</sup>Akhanova N.Ye.**

Professor, ORCID ID: 0000-0003-2767-8711,

e-mail: n.akhanova@kbtu.kz

<sup>1</sup>Kazakh-British Technical University, Almaty, Kazakhstan

<sup>2</sup>International Information Technology University, Almaty, Kazakhstan

## FORECASTING AIR POLLUTION FROM EMISSIONS IN URBAN CANYONS USING ML-CFD MODELING

### Abstract

This paper presents a hybrid approach for predicting pollutant dispersion in urban street canyons, taking into account noise barriers. The methodology combines detailed CFD modeling and a surrogate model based on the BiLSTM neural network architecture with an attention mechanism. Configurations with barrier heights of 0.1H, 0.2H, and 0.3H were studied. CFD calculations revealed a nonlinear effect of barrier height on aerodynamics and the formation of pollutant accumulation zones, with the most complex non-stationary behavior observed at a height of 0.2H. The surrogate model successfully predicts concentration evolution for both the barrier-free and 0.1H barrier cases, demonstrating an average absolute percentage error of less than 15%. For a 0.2H barrier, accuracy decreases in zones of intense turbulence due to the highly non-stationary nature of the process. This approach significantly reduces computational costs while maintaining physical accuracy, which is promising for decision support systems in urban ecology. The model provides accelerated forecasting compared to CFD calculations by 7–8 orders of magnitude, so the inference time showed 1–5 ms, while one CFD simulation takes about 54 hours on the CPU.

**Keywords:** air pollution, street canyons, noise barriers, computational fluid dynamics (CFD), machine learning, BiLSTM, surrogate modeling, concentration prediction

*Received February 12, 2026; revised April 16, June 1, 2026; accepted June 3, 2026*