

УДК 004.8

СРАВНИТЕЛЬНЫЙ АНАЛИЗ И ОСОБЕННОСТИ ГРАФОВЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОСТРАНСТВЕННО-ВРЕМЕННОГО ПРОГНОЗИРОВАНИЯ ЗАГРЯЗНЕНИЯ ВОЗДУХА

Абубакерова З.М.

*Институт машиноведения, автоматизации и геомеханики НАН КР, Бишкек,
Кыргызская Республика.*

E-mail: abubazuu@gmail.com

Аннотация. В статье представлен обзор современных пространственно-временных графовых и гибридных нейронных сетей, применяемых для прогнозирования качества воздуха. Рассмотренные модели направлены на повышение точности прогнозов за счет учета сложных неевклидовых пространственно-временных зависимостей между станциями мониторинга и метеорологических факторов. Особое внимание уделено динамическому построению графов с использованием механистических моделей, многоуровневым архитектурам «регион-станция», адаптивным механизмам внимания, а также гибридации с декомпозицией сигнала. Проведенный анализ показывает, что эти подходы превосходят традиционные методы, обеспечивая более точное и устойчивое долгосрочное прогнозирование концентраций загрязнителей, включая редкие экстремальные события.

Ключевые слова: сверточные нейронные сети (CNN), графовые нейронные сети (GNN), прогнозирование качества воздуха, пространственно-временные данные, механизм внимания (attention mechanism)

Введение. Прогнозирование качества воздуха (ПКА) представляет собой важную инженерную задачу, напрямую связанную с охраной здоровья населения и реализацией экологической политики во всем мире. Загрязнение атмосферного воздуха, особенно мелкодисперсными частицами PM_{2.5}, признано одним из ведущих глобальных факторов риска, поскольку эти частицы способны проникать глубоко в легкие, вызывая респираторные и сердечно-сосудистые заболевания. Точные и устойчивые прогнозы концентраций загрязнителей необходимы для систем раннего предупреждения и поддержки принятия решений в городском управлении. Концентрации загрязнителей зависят от метеорологических условий, географических особенностей и пространственно-временного переноса загрязняющих веществ, на основе этих параметров и строятся прогнозные модели. Традиционные механистические модели (на основе уравнений атмосферной физики и химии) требуют множества эмпирических параметров и упрощений, что ограничивает их применимость в различных городских условиях. Многофакторная обусловленность и существенные нелинейные зависимости в процессах загрязнения воздуха приводят к необходимости применения методов машинного обучения в построении моделей этих процессов.

Ранние подходы машинного обучения и глубокого обучения, включая регрессионные модели, а также сети долгой краткосрочной памяти (Long Short-Term Memory, LSTM) и сверточные нейронные сети (convolutional neural network, CNN), позволили повысить точность прогнозов, но имеют ряд ограничений [1, 2, 3, 4]. Модели, основанные исключительно на LSTM, эффективно выявляют временные закономерности, однако слабо отражают пространственные взаимодействия между станциями мониторинга. CNN способны моделировать пространственные зависимости, но опираются на предположение о регулярной евклидовой сетке, тогда как реальные данные мониторинга имеют неравномерное и разреженное пространственное распределение.

Для преодоления этих ограничений все большее распространение получают графовые нейронные сети (Graph Neural Network, GNN), которые позволяют работать с неевклидовыми структурами данных [5]. В таких моделях станции мониторинга рассматриваются как узлы графа. Комбинируя GNN для извлечения пространственных признаков с рекуррентными нейронными сетями, такими как GRU (Gated Recurrent Unit) или LSTM для моделирования

временной динамики, исследователи разработали эффективные пространственно-временные графовые архитектуры [6].

Целью настоящего обзора является систематизация и сравнительный анализ ключевых направлений развития пространственно-временных графовых нейронных сетей (Spatial-Temporal Graph Neural Network, STGNN), применяемых для прогнозирования качества воздуха. Особое внимание уделяется стратегиям построения динамических графов, интеграции разнородных данных, а также применению механизмов внимания и декомпозиции сигнала для повышения точности и устойчивости прогнозов.

Механизм внимания в глубоких и графовых нейронных сетях. Механизм внимания (attention mechanism) – это метод глубокого обучения, позволяющий модели выделять наиболее значимые элементы входной последовательности или структуры. Изначально предложенный для задач машинного перевода [7], он стал ключевым компонентом трансформеров, гибридных пространственно-временных сетей и GNN [8].

Общая формализация: пусть дан набор скрытых состояний (векторов признаков) $h = \{\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_N\}$, где $\tilde{h}_j \in \mathbb{R}^F$ – вектор признаков j -го элемента.

Механизм вычисляет оценку выравнивания e_{ij} (релевантность) элемента \tilde{h}_j относительно некоторого запроса q_i : $e_{ij} = a(q_i, k_j)$, где k_j – ключи, $a(\cdot)$ – функция совместимости, например, многослойный перцептрон (Multilayer Perceptron, MLP) или скалярное произведение (Dot Product).

Кроме того, необходимо вычисление следующих характеристик:

1. *нормированные веса внимания*

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^T \exp(e_{ik})}, \quad (1)$$

2. *контекстный вектор:*

$$c_i = \sum_{j=1}^T \alpha_{ij} v_j, \quad (2)$$

где v_j – значения (values) элементов последовательности.

В задачах ПКА механизм внимания динамически оценивает влияние станций, временных шагов и связей, улучшая точность при гетерогенных данных.

Механизм внимания в графовых нейронных сетях. Сеть графового внимания (Graph Attention Network, GAT) – частный случай внимания в GNN, где обучаемый механизм внимания взвешивает вклады соседних узлов, в отличие от фиксированной матрицы смежности в графовой сверточной сети (Graph Convolutional Networks, GCN).

Пусть узлы i и j имеют признаки h_i и h_j . Тогда можно вычислить следующие характеристики.

1. *Оценка выравнивания (частный случай общей формулы):*

$$e_{ij} = \text{LeakyReLU}(\alpha^T [Wh_i \parallel Wh_j]), \quad (3)$$

где W – матрица преобразования, \parallel – конкатенация.

Это отвечает общей формуле $e_{ij}=a(q_i, k_j)$ при $q_i = Wh_i$ $k_j = Wh_j$.

2. *Нормированные коэффициенты:*

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(e_{ij}))}{\sum_{k \in N(i)} \exp(\text{LeakyReLU}(e_{ik}))}, \quad (4)$$

что полностью соответствует универсальной softmax-нормализации.

3. *Обновление узла:*

$$h'_i = \sigma(\sum_{j \in N(i)} \alpha_{ij} Wh_j), \quad (5)$$

где $\sigma(\cdot)$ – нелинейная функция активации.

Механизм внимания является фундаментальной составляющей всех современных STGNN, представленных в последующих разделах обзора.

Архитектурные и теоретические преимущества GAT

1. Для стабилизации обучения и повышения выразительности GAT использует K независимых механизмов внимания (голов), каждый из которых применяет собственную

линейную трансформацию и вычисляет свои коэффициенты внимания. Агрегация результатов зависит от положения слоя:

- в промежуточных слоях – *конкатенация* выходов всех голов, что дает узлу $K \cdot F'$ признаков (увеличивает размерность представления);
- на финальном (прогножном) слое – *усреднение (averaging)* по головам перед применением нелинейности (например, softmax);

Такой подход позволяет модели одновременно учитывать информацию из разных подпространств признаков, существенно повышая устойчивость и качество обучения.

2. Высокая вычислительная эффективность.

Операция самовнимания полностью подвергается параллелизму по всем ребрам (вычисление α_{ij}) и по всем узлам (агрегация признаков).

Временная сложность одного слоя с F' выходными признаками составляет $O(N/FF' + |E|/F')$, где N – число узлов, $|E|$ – число ребер. Это сопоставимо со сложностью GCN и значительно ниже, чем у ранних спектральных методов, требующих дорогостоящих операций (например, разложения Лапласиана или инверсии матриц).

GAT не зависит от глобальной структуры графа и не требует предварительного знания всей матрицы смежности.

3. Индуктивное обучение и обобщаемость.

Механизм внимания применяется локально (только к соседям первого порядка), что делает GAT независимой от глобальной структуры графа. Благодаря этому модель поддерживает *индуктивное обучение* – может обобщаться на полностью не учтенные во время обучения графы, например, набор данных PPI (Protein-Protein Interaction) с новыми графами в тестовой выборке.

4. Гибкое взвешивание соседей. В отличие от GCN, где все соседи имеют равный вес, GAT позволяет неявно присваивать *разную важность узлам* внутри одного соседства через обучаемые коэффициенты внимания.

В совокупности эти свойства делают GAT мощным и гибким инструментом для пространственно-временных задач, включая ПКА, где требуется динамическое моделирование неевклидовых зависимостей без жестких предположений о структуре данных.

GNN с вниманием для городских многомодальных данных (кейс Мадрид, АЗТ-GCN)

Среди моделей, интегрирующих механизм внимания, выделяется АЗТ-GCN (Attention Temporal Graph Convolutional Network), предложенная для прогнозирования концентрации диоксида азота (NO_2), в Мадриде [9]. Модель преодолевает ограничения традиционных подходов за счет эффективной обработки неевклидовых данных и сложных пространственно-временных зависимостей в условиях городской среды.

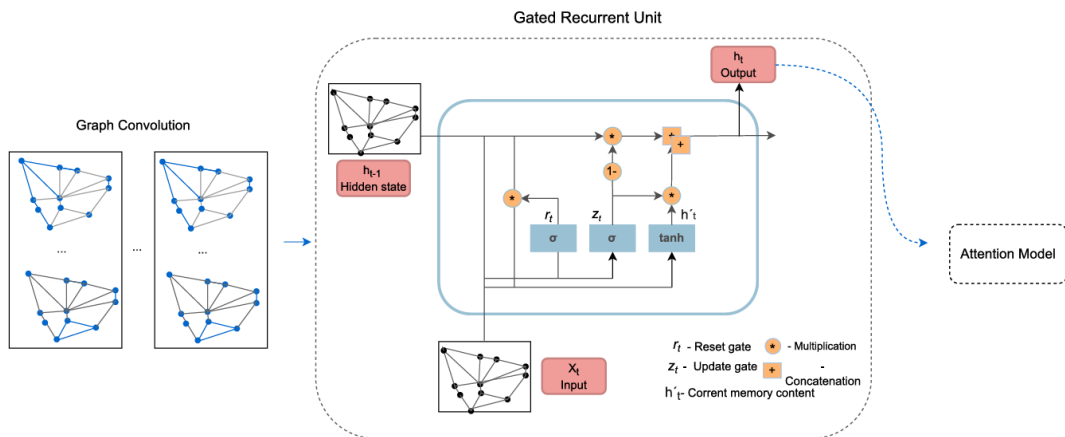


Рисунок 1– Архитектура модели АЗТ-GCN

Архитектура АЗТ-GCN. АЗТ-GCN сочетает в себе три ключевых компонента (рис. 1):

- GCN – для моделирования пространственных связей между станциями,
- GRU – для захвата временных зависимостей в последовательности,
- механизм внимания – для динамического взвешивания вклада разных временных интервалов и фокуса в наиболее релевантных признаках.

Архитектура позволяет учитывать мультимодальные входные данные, адаптивно выделять важные пространственно-временные паттерны.

Модель обучалась и тестировалась на гетерогенных данных Мадрида за периоды январь-июнь 2019 и 2022 гг.:

- концентрации NO₂ с 224 станций мониторинга,
- метеорологические параметры (температура, влажность, давление, солнечная радиация, скорость и направление ветра – последнее через One-Hot Encoding),
- данные о дорожном движении (интенсивность, занятость, нагрузка, средняя скорость) из > 4000 точек.

Данные агрегированы в сетку из 340 ячеек (покрытие ~56% города) с разными временными разрешениями (15 минут для трафика, 1ч для остальных). Входная матрица признаков X_t имеет размерность $N \times M$, где $N=24$ (станции) и $M=18$ (признаков).

Пространственная структура представлена как ненаправленный взвешенный граф:

- $N=24$ узла, соответствие с каждой станцией мониторинга NO₂,
- число ребер $N_e=276$ (полносвязный граф между всеми парами),
- веса ребер = инверсия евклидова расстояния между станциями (чем ближе – тем сильнее влияние).

Задача формулируется как регрессия узла (node regression), при этом прогноз NO₂ выполняется для каждого узла на горизонтах 1–48 часов (разбитых на интервалы: 1–12, 12–24, 24–36, 36–48 ч).

Графовая сверточная сеть – пространственная агрегация признаков. В АЗТ-GCN применяется спектральная GCN, она определяет свертки в Фурье-домене через фильтры. Свертка задается как умножение фильтра g_θ на сигнал x в Фурье-домене: $g_\theta(L) * x = U g_\theta(U^T x)$, где g_θ – фильтр с параметрами θ , L – матрица Графова Лапласиана ($L \in \mathbb{R}^{N \times N}$). $L = I_N - D^{-1/2} A D^{-1/2}$ – нормализованная матрица Лапласиана графа, где I_N – единичная матрица, D – диагональная матрица степеней, A – матрица смежности. Веса смежности вычисляются как:

$$A_{ij} = \begin{cases} \frac{1}{d_{ij}}, & i \neq j \\ 0, & otherwise \end{cases}, \text{ где } d_{ij} \text{ – расстояние между станциями } i \text{ и } j.$$

Архитектура АЗТ-GCN состоит из трех слоев графовой свертки, выходные данные которых используются в блоках GRU.

Gated Recurrent Unit. Временная агрегация в АЗТ-GCN осуществляется с помощью GRU, которая эффективно захватывает долгосрочные временные зависимости за счет двух управляющих вентиляей:

- затвор обновления: $z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1})$ – определяет, какая часть предыдущего скрытого состояния h_{t-1} переносится в текущее h_t ;
- затвор сброса: $r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1})$ – контролирует, сколько информации из прошлого забывается при формировании кандидата на новое состояние h'_t .
- кандидат на память: $h'_t = \tanh(Wx_t + r_t \odot Uh_{t-1})$;
- итоговое скрытое состояние: $h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h'_t$.

Выходы GRU подаются в механизм *Soft Attention*, который динамически взвешивает вклады разных временных шагов h_j из последовательности T_x . Это позволяет модели фокусироваться на наиболее релевантных исторических состояниях.

Веса внимания (a_{ij}) рассчитываются с помощью функции Softmax на основе оценки выравнивания (e_{ij}), что позволяет определить контекстный вектор (c_i) как взвешенную сумму скрытых состояний.

Таким образом, модель АЗТ-GCN эффективно интегрирует многомодальные данные (концепции загрязнителей, метеопараметры, трафик) в единой графовой структуре, подтверждая важность механизмов внимания в задачах пространственно-временного прогнозирования. Производительность модели подтверждена метриками RMSE, MAE и коэффициентом корреляции Пирсона (R).

Адаптивные пространственно-временные зависимости и внимание в STAA-GCN

Одной из наиболее заметных архитектур в развитии пространственно-временных GNN стала STAA-GCN (Spatio-temporal adaptive attention graph convolution network), ориентированная на моделирование полного спектра зависимостей между станциями мониторинга для городских прогнозов качества воздуха, в частности короткосрочных концентраций PM2.5 [10].

Модель использует encoder-decoder фреймворк для многошаговых прогнозов. Ключевой вклад модели заключается в систематическом кодировании полных пространственно-временных зависимостей, включающих не только три специфических типа отношений (близость, функциональное и временное сходства), но и их комбинации, которые ранее игнорировались в подходах с отдельными графами.

В отличие от предшественников (моделирующих изолированные графы), STAA-GCN извлекает общие зависимости через Bi-level sharing стратегию для извлечения общих зависимостей (shared kernels + shared diagonal): общие сверточные ядра для комбинированных отношений; диагональные матрицы для удаления нерелевантных станций (L2-регуляризация).

Общая схема архитектуры STAA-GCN представлена на рисунке 2.

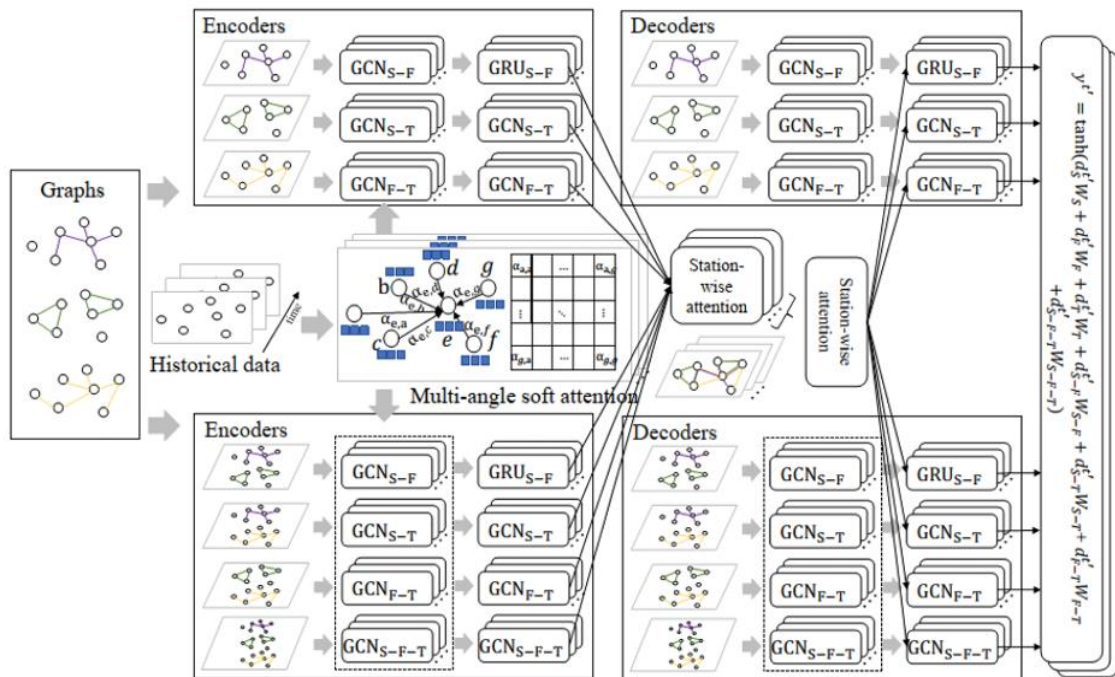


Рисунок 2 – Архитектура STAA-GCN

STAA-GCN вписывается в ландшафт моделей глубокого обучения для прогнозирования качества воздуха, где ключевыми являются: ATGCN (Attentive Temporal Graph Convolutional Network), Graph WaveNet и модели на основе DCRNN (Diffusion Convolutional Recurrent Neural Network) / STGCN (Spatial-Temporal Graph Convolutional Network) / TCN (Temporal Convolutional Networks) [1, 5, 6, 7].

STAA-GCN иллюстрирует, как адаптивное внимание и структурирование полных (включая shared) зависимостей повышают выразительность и точность в задачах городского ПКА с неравномерными данными мониторинга.

Входные данные и графы отношений в STAA-GCN. Модель использует гетерогенные входные данные:

1. Мониторинг загрязнителей (Q) (PM2.5, PM10, O₃, NO₂, SO₂) и метеопараметры (температура, влажность, скорость и направление ветра).
2. Данные P (Points of Interest, POI): информация о функциональном окружении станции (фабрики, жилые зоны и др).
3. Временные метки (W) (час/день/неделя/месяц) для учета сезонных и суточных паттернов.

Исторические данные Q за шаги T , POI P и временная информация W используются для получения прогнозируемых значений \hat{y} : $f(Q; P; W; \theta) \rightarrow \hat{y}$.

На основе этих данных строятся три специфических графа отношений для дальнейшего извлечения зависимостей:

- граф пространственной близости (G_S) отражает физическое расстояние между станциями; связность реализуется, если расстояние меньше порогового значения α_S , а веса ребер рассчитываются как обратная величина физического пространственного расстояния

$$A_{S_{ij}} = \begin{cases} \text{dist}(v_i, v_j) < \alpha_S; \\ 0, \text{otherwise} \end{cases};$$

- граф функционального сходства (G_P) отражает корреляцию между станциями, окруженными схожими POI; веса = обратное Евклидово расстояние между POI-векторами станций

$$A_{P_{ij}} = \begin{cases} \text{sim}(P_i, P_j) > \alpha_F; \\ 0, \text{otherwise} \end{cases};$$

- граф сходства временных паттернов (G_T) использует потенциальные временные корреляции, например, среднемесячные значения; веса ребер – коэффициент корреляции Пирсона

$$A_{T_{ij}} = \begin{cases} \text{sim}(t_i, t_j) > \alpha_T. \\ 0, \text{otherwise} \end{cases}.$$

Модули кодировщика: извлечение полных пространственно-временных зависимостей. Архитектура включает в себя семь параллельных кодировщиков для обработки всех типов отношений, трех специфических (S , F , T) и четырех общих комбинаций (S - F , S - T , F - T , S - F - T).

Мягкое внимание с нескольких углов (Multi-angle Soft Attention) решает проблему неравномерных и динамических зависимостей между станциями (из-за диффузии загрязнителей). Входные признаки станций $X_t \in \mathbb{R}^{N \times d}$ проецируются в M семантических пространствах, затем вычисляются веса влияния (усредненные по углам), формируя адаптивную матрицу весов для каждой временной метки:

$$\alpha_{i,j}^t = \frac{1}{M} \sum_{m=1}^M \text{softmax} \left((x_i^{m,t})' \right) = \frac{1}{M} \sum_{m=1}^M \frac{V_i^m \cdot \sigma(W_i^m \cdot x_i^t + b_i^m)}{\sum_{j=1}^N V_j^m \cdot \sigma(W_j^m \cdot x_j^t + b_j^m)},$$

где W_i^m, V_i^m, b_i^m – обучаемые параметры, а σ – функция активации tanh.

Итоговая матрица весов $E^t \in \mathbb{R}^{N \times N}$ формируется из вычисленных весов между всеми станциями.

Блок специфической графовой свертки (Specific Graph Convolutional Unit). Для специфических отношений $r_{\text{spe}} \in \{S, F, T\}$ применяется вершинная свертка, основанная на произведении Адамана (Hadamard product, \odot). Матрица смежности $A_{r_{\text{spe}}}$ графа объединяется с матрицей весов E^t (от GAT).

Агрегация информации k -соседей:

$$\tilde{X}_{r_{\text{spe}}}^t = \sum_{k=1}^K \left((D_{r_{\text{spe}}}^t)^{-1} A_{r_{\text{spe}}} \odot E^t \right)^k X^t W_{r_{\text{spe}}}^k,$$

где $W_{r_{\text{spe}}}^k$ – обучаемый параметр, K – глубина агрегации (в экспериментах $K = 2$), а $(D_{r_{\text{spe}}}^t)^{-1}$ – обратная диагональная матрица, обеспечивающая нормализацию.

Двухуровневая стратегия совместного использования (Bi-level Sharing Strategy) – ключевая инновация для извлечения общих зависимостей. Ключевыми особенностями этой стратегии являются следующие:

А) Совместные/разделяемые ядра свертки $W_{r_{\text{sct}}}^k$ на графе для извлечения общих признаков из различных пространств, например, для комбинации трех отношений $r_{\text{sct}} \in \{S - F - T\}$, в качестве входа используются все три специфических графа G_S, G_P, G_T , но с общим ядром (Shared Kernel):

$$\begin{cases} X_S^t = \sum_{k=1}^K ((D_S^t)^{-1} A_S \odot E^t)^k X^t W_{r_{\text{sct}}}^k \\ X_F^t = \sum_{k=1}^K ((D_F^t)^{-1} A_F \odot E^t)^k X^t W_{r_{\text{sct}}}^k ; \\ X_T^t = \sum_{k=1}^K ((D_T^t)^{-1} A_T \odot E^t)^k X^t W_{r_{\text{sct}}}^k \end{cases}$$

В) Разреженные диагональные матрицы весов W_D^t с L1-регуляризацией применяются, поскольку не все станции имеют общие комбинированные зависимости, для обнуления ($\tilde{X}_S^t = X_S^t W_D^t$) слабых/нерелевантных станций Shared Diagonal Matrix). Итоговое представление общих отношений получается усреднением по всем уровням:

$$\tilde{X}_{r_{\text{sct}}}^t = (\tilde{X}_S^t + \tilde{X}_F^t + \tilde{X}_T^t)/3.$$

Блок временной автокорреляции (Temporal Autocorrelation Unit) использует GRU для захвата корреляций по каждой из семи пространственных зависимостей

$$r \in \{r_{\text{spe}} \cup r_{\text{sct}} \cup r_{\text{sca}}\}.$$

Скрытое состояние h^t на шаге t рассчитывается на основе текущей пространственной зависимости \tilde{X}_r^t и предыдущего скрытого состояния h^{t-1} :

$$\begin{cases} r^t = \sigma(W^r [h^{t-1} \parallel \tilde{X}_r^t] + b^r) & \text{(Reset Gate)} \\ z^t = \sigma(W^z [h^{t-1} \parallel \tilde{X}_r^t] + b^z) & \text{(Update Gate)} \\ \tilde{h}^t = \tanh(W^{\tilde{h}} [r^t \odot h^{t-1} \parallel \tilde{X}_r^t] + b^{\tilde{h}}) \\ h^t = (1 - z^t) \odot h^{t-1} + z^t \odot \tilde{h}^t \end{cases}$$

Слияние внимания на уровне станции (Station-level Attention Fusion) сливает семь скрытых состояний \tilde{h}_m^t в одно полное пространственно-временное представление: веса внимания $\alpha_m^{i,t}$ определяют относительную важность каждого типа зависимости для конкретной станции в данный момент:

$$\alpha_m^{i,t} = \text{softmax}(h_m^{i,t}) = \frac{\exp(h_m^{i,t})}{\sum_{m \in M} \exp(h_m^{i,t})}, \quad h_m^{i,t} = V_m^t \tanh(W_m^t \cdot (\tilde{h}_m^{i,t})^T + b_m^t).$$

Итоговое скрытое состояние $h^{i,t}$ для станции i является взвешенной суммой всех семи зависимостей:

$$h^{i,t} = \sum_{m \in M} \alpha_m^{i,t} h_m^{i,t}.$$

Модуль извлечения признаков (Декодер) использует внимание на уровне времени (Temporal-level attention) для адаптивного захвата корреляций между целевым прогнозом и историческими скрытыми состояниями $\{h^1, \dots, h^T\}$. Контекстный признак c_m^t используется для обновления состояния декодера d_m^t , на каждом будущем шаге прогнозирования t' :

$$c_m^t = \sum_T \alpha_m^t \cdot h^t.$$

Механизм гейтинга (Gating mechanism) взвешивает и объединяет множественные пространственно-временные зависимости $d_m^{t'}$, полученных декодером, для генерации окончательного многошагового прогноза $\hat{y}^{t'}$:

$$\hat{y}^{t'} = \tanh(d_S^{t'} W_S + d_F^{t'} W_F + d_T^{t'} W_T + \dots + d_{S-F-T}^{t'} W_{S-F-T}),$$

где W_m – обучаемые параметры весов для каждого типа зависимости.

Функция потерь и регуляризация. Поскольку прогнозирование качества воздуха является задачей регрессии, основной функцией потерь является среднеквадратичная ошибка (MSE):

$$\mathcal{L}_P = \sum_{t=1}^{\tau} \|y^{T+t} - \hat{y}^{T+t}\|_2^2.$$

Для снижения риска переобучения вводятся дополнительные регуляризации:

- регуляризация множественных отношений (Basis Decomposition) применяется для уменьшения числа параметров в графовой части; ядро свертки W_r^l раскладывается на общую базисную матрицу V_b^l и линейные коэффициенты $a_b^{l,r}$:

$$W_r^l = \sum_{b=1}^B a_b^{l,r} V_b^l,$$

- регуляризация разреженности (L1) применяется к сумме весов диагональной матрицы, чтобы эффективно приблизить их к порогу обнуления.
- ограничение согласованности (Consistency Constraint, L2) между общими зависимостями. Например, для S-F: $L_{CS-F} = |\tilde{X}_S^t - \tilde{X}_F^t|^2$.

Общая функция потерь (L) включает ошибку прогнозирования и все регуляризационные компоненты (с учетом того, что декодеры имеют аналогичные модули, потери L_E и L_D удваиваются):

$$L = L_P + \gamma(L_{ES_{S-F}} + L_{DS_{S-F}}) + \dots + \iota(L_{EC_{S-F-t}} + L_{DC_{S-F-t}}).$$

Роль каждого модуля: абляция. Абляционные эксперименты показывают критическую роль каждого компонента: удаление station-level attention, multi-angle attention, shared diagonal matrix или consistency constraint приводит к заметному росту Mean Absolute Error (MAE) на несколько процентов в зависимости от датасета. Модель достигает state-of-the-art (высшие показатели) на датасетах Пекина, Тяньцзиня и Лондона по RMSE, MAE и MAPE.

Сложность и ресурсы. Сложность STAA-GCN составляет $O(N^2KM+NL)$, где N – число станций (квадратичная зависимость от N), K – глубина свертки, M – углы внимания, L – слои. Для средних городов (десятки станций) это приемлемо, для очень крупных сетей требуется оптимизация.

Ограничения и переносимость. Shared-зависимости могут «шуметь» в следующих случаях: редкие или некачественные POI, малое число станций. Модель чувствительна к пороговым значениям расстояний и сезонным изменениям. Перенос на города с иной схемой зонирования и редкой сетью станций может потребовать адаптации.

Долгосрочный прогноз PM2.5 с физически осмысленными ребрами (Multi-gated GNN + GAT)

Особое направление исследований связано с прогнозом на средние и долгие горизонты (72–96 часов). Для этих задач была предложена модель STGNN, ориентированная на учет физических механизмов переноса загрязнителей через многографовые структуры [11]. Данная работа фокусируется на Южной Корее, где высокие концентрации PM2.5 часто связаны с сезонными трансграничными факторами (включая влияние Китая).

Основные особенности модели:

- построение множественных ребер, сочетающих физический коэффициент атмосферной диффузии (на основе модифицированной модели Гауссова плюма с учетом ветра, расстояния и направления) и статистическое сходство временных рядов PM2.5 (по евклидову расстоянию последовательностей);

- гибридный модуль Multi-gated GNN + GAT + GRU, где Multi-gated GNN агрегирует признаки соседей (сумма и среднее) через гейтированный линейный блок (Gated Linear Unit, GLU), GAT динамически взвешивает важность узлов, а GRU обрабатывает временные последовательности.

Структура данных: 3 года (2015-2017), 8760 временных точек, 101 станция, 45 – признаков, включая:

- почасовые средние концентрации PM2.5 в Южной Корее;
- PM2.5 из Китая за 24 часа до текущего момента ($t-24$);
- метеоданные ECMWF (European Centre for Medium-Range Weather Forecasts): температура, влажность, давление, ветер (U, V), высота пограничного слоя и др.;
- загрязнители (AOD, PM1, PM10, CO, NO₂, SO₂, O₃ и др.)
- география: координаты, землепользование, плотность населения;
- временные признаки: месяц и час, которые были дополнительно закодированы косинусно для учета сезонности и суточных пиков.

Архитектура Multi-gated GNN. Кодирование временных признаков (*Cosine Transformer*) осуществляется на основе косинуса для учета сезонной и почасовой периодичности:

$$Month' = \cos(30^\circ(x_{month} - 3)), Hour' = |\cos(15^\circ(x_{hour} - 9))|.$$

Генератор многографовой матрицы (Multi-Graph Matrix Generator) создает гибридные ребра (диффузия + сходство). Диффузионное ребро отражает физический перенос PM2.5 воздушными массами (учитывает скорость, направление ветра и расстояние), то есть реальное распространение загрязнителя от источника к приемнику. Ребро сходства отражает статистическую схожесть временных рядов PM2.5 между станциями, что компенсирует случаи, когда физическая диффузия слаба или станции расположены неравномерно.

Ребро коэффициента диффузии PM2.5 ($Q_{i \rightarrow j}^t$) основано на модифицированной формуле Гауссова плюма (Gaussian plume). Веса ребра отражают количество PM2.5, переносимого от станции i к станции j .

$$Q_{i \rightarrow j}^t = \text{SoftMax} \left[\text{Standards} \left(\frac{1}{|a| \cdot u \cdot d} \right) \right],$$

где $|a|$ – абсолютная разница между направлением ветра на станции i и направлением от i к j ; u – горизонтальная скорость ветра на станции i ; d – Евклидово расстояние между станциями i и j ; Standards – стандартизация (Z-score); SoftMax – нормализация в диапазоне.

Ребро коэффициента сходства PM2.5 ($Q_{i \rightarrow j}^t$) учитывает сходство временных рядов PM2.5 между двумя станциями, что помогает компенсировать низкую предсказательную способность диффузионного ребра в районах с неконцентрированным расположением станций:

$$Q_{i \rightarrow j}^t = \text{SoftMax} \left[\text{Standards} \left(\frac{1}{D_{tw}(i, j)} \right) \right],$$

где $D_{tw}(i, j)$ – Евклидово расстояние между временными последовательностями PM2.5 станций i и j .

Многогейтовая/многовентильная GNN и GRU (Multi-gated GNN) – это гибрид для совместного извлечения пространственно-временных признаков.

Построение признаков узла и ребра (S_i^t) строятся путем конкатенации прогноза PM2.5 на предыдущем шаге (\hat{X}_i^{t-1}) и текущих атрибутов (P_i^t):

$$S_i^t = \hat{X}_i^{t-1} \oplus P_i^t.$$

Выходная матрица признаков ребра ($e_{i \rightarrow j}^t$) получается путем поэлементного умножения (\odot) признаков узла на веса обоих ребер и пропуска через многослойный перцептрон (M_t):

$$e_{i \rightarrow j}^t = M_t(S_i^t \odot Q_{i \rightarrow j}^t \odot Q'_{i \rightarrow j}{}^t, S_j^t).$$

Механизм агрегации и гейтирования (GLU и Multi-gated GNN) агрегирует признаки, вычисляя сумму (S_{sum}^t) и среднее (S_{avg}^t) признаков всех связанных узлов. Далее применяется GLU, который модулирует линейную проекцию входных данных с помощью сигмоидного гейта.

Графовая сеть внимания параллельно используется для автоматического захвата динамической важности узлов. Коэффициент внимания a_{ij}^t рассчитывается при помощи LeakyReLU и Softmax.

Рекуррентный блок (GRU) используется для обработки временных последовательностей и решения проблемы долгосрочной зависимости. Входной вектор x_j^t для GRU формируется путем конкатенации результатов трех модулей: Multi-gated GNN (\hat{h}_j^t), GAT (\hat{h}_j^t) и текущего состояния узла (S_j^t); GRU использует гейты обновления (z_j^t) и сброса (r_j^t) для расчета скрытого состояния h_j^t ; потеря - L1 (MAE).

Результаты и значимость. Модель превосходит 12 базовых методов (включая LSTM, GRU, GCN+GRU, GAT+GRU и PM2.5GNN+GRU), особенно на горизонтах 72 и 96 часов. Для прогноза на 96 часов вперед по RMSE - +2.613% и по R^2 - +5.263% по сравнению с лучшей базовой моделью.

Абляция подтвердила важность: Cosine Transformer, ребра сходства PM2.5 (особенно для удаленных станций), агрегация (сумма+среднее) и GLU-гейтирование.

Представленная модель STGNN демонстрирует, как физически осмысленные многографовые ребра в комбинации с Multi-gated GNN, GAT и GRU позволяют преодолеть проблемы локальной зависимости и улучшить глобальную точность долгосрочных прогнозов, включая экстремальные события. Это одно из наиболее перспективных направлений для практического применения в системах мониторинга и управления качеством воздуха.

Иерархическая (мульти-гранулярная) STGNN с динамическим графом по HYSPLIT (MGST-GNN)

Дальнейшее развитие STGNN-подходов привело к интеграции физических моделей в структуру графовых нейросетей. Примером является Multi-Granularity ST-GNN – гибридная иерархическая графовая нейросеть «регион-станция», которая интегрирует физическую модель траекторий воздушных масс HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory) для динамического построения графа [12]. Основная цель – преодолеть ограничения статических пространственных связей в большинстве ST-GNN.

Ключевая идея. Связи между узлами (ребра и их веса) формируются не по расстоянию или фиксированному направлению ветра, а по реальным траекториям переноса загрязнителей, рассчитанным HYSPLIT за 24–48 часов. Это делает граф динамическим, направленным и физически осмысленным.

Методология и архитектура модели: многогранулярное представление данных. MGST-GNN использует иерархическую структуру графа, состоящую из двух уровней:

- глобальный граф – вершины соответствуют регионам (например, 13 регионов города Цзинань или 10 городов дельты Янцзы),
- частичные подграфы – узлы соответствуют станциям внутри региона и содержат четыре группы признаков:
 - концентрация загрязнителей (PM2.5, PM10, SO₂, CO₂, CO, O₃ и др.);
 - метеоданные (Т, влажность, давление, осадки, ветер);
 - временные признаки: час, день недели, месяц (one-hot encoding);
 - геоморфология (высота и тип землепользования в радиусе 10 км).

Динамическое построение ребер с помощью HYSPLIT. HYSPLIT моделирует перенос, диффузию и осаждение воздушных масс. Для каждой станции или региона вычисляются траектории воздушных потоков в течение заданного горизонта (например, 48 часов). Если из узла v_a пять траекторий проходят через узел v_b в часы 1, 3, 4, 8 и 48, то вектор атрибутов ребра e_{ab} принимает вид:

$$e_{ab} = [\dots, 0, 1, 0, 1, 1, 0, 0, 1, \dots],$$

где единицы отражают время, когда загрязненная масса достигает целевого узла.

Эта схема позволяет сети учитывать не просто соседство, а фактическое воздействие источников загрязнения с учетом ветра, осадков и рельефа.

Механизм передачи сообщений (Message Passing) строится на основе пространственных зависимостей, которые моделируются через классический механизм передачи сообщений, включающий агрегацию M и обновление описания ϕ :

$$m_a = \psi(M_a(x_s, x_a, e_{a,s})), \quad x'_a = \phi(x_a, m_a, u_a),$$

где x_a – признаки узла, $e_{a,s}$ – атрибуты ребра, u_a – глобальный контекст региона.

Передача информации происходит в обоих направлениях:

- восходящая агрегация (Partial \rightarrow Global): усреднение признаков станций;
- нисходящее обновление (Global \rightarrow Partial): глобальный LSTM передает контекст вниз через полносвязный слой.

Таким образом, сеть способна одновременно учитывать локальные и межрегиональные зависимости, формируя согласованную картину распространения загрязнений.

Кодировщик-декодер с временным вниманием. LSTM в кодировщике принимает в качестве входных данных последовательность атрибутов узла станции $X_{a,i}$ за τ_{in} часов. Конечное скрытое состояние LSTM кодировщика передается в декодер.

В декодирующей стадии вводится механизм временного внимания для изучения динамической временной корреляции между будущим и историческим временем. Оценка выравнивания (ρ) вычисляется на основе скрытого состояния энкодера h_t и предыдущего скрытого состояния декодера h'_{t-1} , c'_{t-1} :

$$\rho_{t'}^t = v^T \tanh(W[h'_{t-1}; c'_{t-1}] + Uh_t + b), \quad \lambda_{t'}^t = \frac{\exp(\rho_{t'}^t)}{\sum_{t=1}^T \exp(\rho_{t'}^t)}$$

Контекстный вектор $c_{t'}$ формируется как взвешенная сумма скрытых состояний h_t , что позволяет декодеру выделять наиболее релевантные исторические зависимости для долгосрочного прогноза: $c_{t'} = \sum_{t=1}^T \lambda_{t'}^t h_t$.

Прогнозирование. Вход для LSTM декодера на шаге t' включает: предыдущий прогнозируемый результат \hat{o}_{t-1} ; прогностические данные о погоде $wd_{t'}$; временные признаки $tf_{t'}$; контекстный вектор $c_{t'}$: $h'_{t'} = \text{LSTM}(h'_{t-1}, [\hat{o}_{t-1}; wd_{t'}; tf_{t'}; c_{t'}])$,

выход скрытого состояния $h'_{t'}$ подается в GNN для получения прогнозируемой концентрации загрязнителя на будущий шаг τ_{out} .

MGST-GNN показала наилучшую точность прогнозирования PM2.5 на горизонте до 48 часов: MAE = 6.15-52.06 (Цзинань), 7.16-54.73 (Янцзы); SMAPE = 0.07-0.46. Метеоданные; исключение погоды, времени или геоморфологии заметно ухудшает качество. Главные преимущества MGST-GNN: физически обоснованный динамический граф (HYSPLIT); многогранулярность (локальные + региональные зависимости); сочетание интерпретируемости физической модели и гибкости глубокого обучения; значительный прирост точности на долгосрочном горизонте.

Пространственно-временное моделирование с частотной декомпозицией на примере VMD-GAT-BiLSTM

Эта модель представляет собой **пространственно-временную гибридную модель глубокого обучения VMD-GAT-BiLSTM**, сочетающую Variational Mode Decomposition (VMD), Graph Attention Network (GAT) и Bi-directional LSTM (BiLSTM) [13]. Модель

предназначена для точного прогноза концентрации PM2.5 на горизонтах от 1 до 48 часов (эффективна при долгосрочном прогнозировании).

Основная идея и архитектура. Модель VMD-GAT-BiLSTM реализует каскадное объединение трех взаимодополняющих подходов (рис.3):

- VMD выполняет декомпозицию исходных временных рядов концентрации PM2.5 на серии стабильных/устойчивых подпоследовательностей (Intrinsic Mode Functions, IMFs), с разными частотными диапазонами. Это позволяет снизить влияние шумов и нестационарности данных на точность модели.
- GAT извлекает пространственные зависимости между станциями для каждой IMF, используя механизм самовнимания, динамически взвешивает вклад соседних станций (без фиксированных весов). Используя 2 слоя GAT (оптимально).
- BiLSTM моделирует временные зависимости в прямом и обратном направлении, улучшая моделирование долгосрочной динамики.

Данные и область исследования:

- 30 станций в Пекине (Китай), почасовые данные 2017–2020 г.;
- загрязнители: PM2.5, PM10, CO, NO₂, O₃;
- метеофакторы: температура, точка росы, давление воздуха, осадки, компоненты ветра (U/V);
- матрица смежности A строится комбинацией: географическое расстояние + коэффициент Пирсона (корреляция рядов между станциями). Эти два фактора комбинируются для определения весов ребер, что позволяет учитывать как физическую близость, так и статистическую связь качества воздуха.

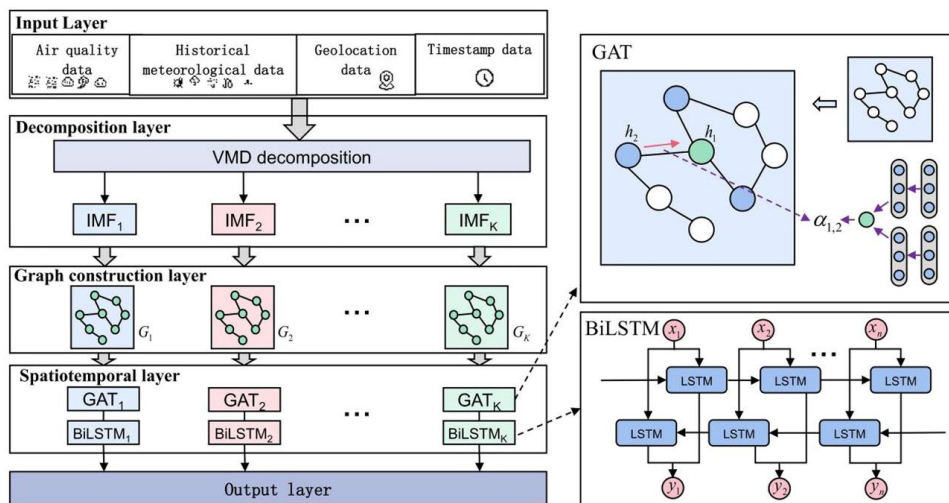


Рисунок 3 – Общая структура модели VMD-GAT-BiLSTM

Итоговый прогноз получается путем агрегирования результатов всех IMFs.

Механизмы модели. Слой декомпозиции (Decomposition Layer). VMD формулирует вариационную задачу разложения сигнала $x(t)$ на K подпоследовательностей $s_i(t)$ с уникальными центральными частотами ω_i , минимизируя функцию:

$$J = \sum_{i=1}^K \left(\frac{1}{2} \| s_i \|^2 + \mu \omega_i \right), \quad \sum_{i=1}^K s_i(t) = u(t).$$

Оптимизация осуществляется методом Лагранжа, а результатом служат IMFs. Используется Pelican Optimization Algorithm (POA) для минимизации энтропии и повышения стабильности декомпозиции.

Слой построения графа (Graph Construction Layer). Матрица смежности A рассчитывается по комбинированному признаку географической дистанции и корреляции:

$$A_{ij} = \begin{cases} \beta \exp\left(-\frac{\text{dist}(i,j)^2}{2\sigma_1^2}\right) + (1-\beta) \exp\left(-\frac{(1-\rho(i,j))^2}{2\sigma_2^2}\right), & a_{ij} \geq \theta, \\ 0, & \text{otherwise,} \end{cases}$$

где β – вес между физическим и статистическим факторами, σ_1, σ_2 – средние значения дистанций (dist) и корреляций (ρ), $\rho(i,j)$ – коэффициент Пирсона (Pearson coefficient) между станциями мониторинга i и j , а θ – порог связи.

Пространственно-временной слой (Spatiotemporal Layer). Каждая подпоследовательность обрабатывается в два этапа: (1) GAT – для извлечения пространственных паттернов с двумя внимательными слоями (оптимальная глубина по результатам эксперимента); (2) BiLSTM – для захвата временных зависимостей и контекста. Такое каскадное соединение позволяет модели улавливать взаимосвязи разных временных масштабов и пространственных зон

Выходной слой (Output Layer): предсказания для всех IMFs агрегируются по формуле: $\hat{y} = \sum_{i=1}^K x_i$, что дает итоговую оценку концентрации PM2.5.

Эксперименты и результаты:

- реализация: PyTorch, Adam ($\eta = 1e-4$), MSE в качестве функции потерь.
- метрики: RMSE, MAE, Index of Agreement (IA).
- базовые модели: GRU, BiLSTM, CNN-LSTM, GCN-LSTM, STGCN, Transformer.

VMD-GAT-BiLSTM превосходит все базовые модели по всем метрикам на горизонтах 1–48 ч. Абляция показала: замена VMD на EMD сильно ухудшает качество; оптимально – именно 2 слоя GAT и $K=6$.

Модель хорошо обобщается и на другие загрязнители (PM10, NO₂, O₃, CO).

VMD-GAT-BiLSTM обеспечивает сбалансированное моделирование пространственно-временных зависимостей и частотной структуры данных. За счет объединения VMD (устранение нестационарности), GAT (гибкое взвешивание пространственных взаимосвязей) и BiLSTM (учет временной долгосрочной динамики) модель достигает наилучших результатов по всем метрикам и является перспективным направлением для моделирования долгосрочного прогноза качества воздуха в городской среде.

Заключение: анализ перспективных архитектур STGNN и рекомендации.

Проведенный обзор STGNN-архитектур подтверждает их превосходство над классическими моделями (LSTM, GRU, CNN-LSTM) и ранними графовыми подходами (STGCN, GCN-LSTM) в задаче ПКА. Использование графовых структур позволяет естественным образом представлять пространственные отношения между точками наблюдения, а внедрение механизмов внимания и адаптивной фильтрации значительно улучшает способность моделей улавливать динамику загрязняющих веществ в условиях высокой вариативности данных.

Ключевые направления развития STGNN-архитектур. Основное повышение точности прогнозов достигнуто за счет интеграции трех ключевых инновационных стратегий:

1. динамическое построение графа – переход от статических графов (по расстоянию) к динамическим, физически обоснованным, связям за счет интеграции профессиональных механистических моделей, таких как HYSPLIT;
2. механизмы внимания – применение GAT и многоугольных/многоуровневых механизмов внимания (Multi-angle, Station-level attention) для динамического взвешивания вкладов соседних узлов решает проблему фиксированных весов в классических GCN;
3. гибридизация с обработкой сигнала – сочетание GNN с методом обработки временного ряда, таким как вариационное модовое разложение (VMD) сигнала на относительно стабильные компоненты (IMFs), что снижает влияние шума и нестационарности.

Моделью, демонстрирующей наиболее впечатляющее превосходство в критически важной области долгосрочного прогнозирования и прогнозирования редких событий, оказалась Гибридная модель Multi-gated GNN/GAT.

Представленная архитектура демонстрирует превосходство благодаря синергии двух ключевых инновационных компонентов: множественные типы ребер (на основе коэффициента атмосферной диффузии и метрики сходства временных рядов PM_{2.5}), обеспечивающие реалистичное моделирование межузловых зависимостей; комбинации Multi-gated GNN и GAT, обеспечивающей эффективное улавливание комплексных пространственно-временных паттернов и снижение эффектов локальности классических GCN-подходов.

Для задач раннего предупреждения, управления рисками и экологического планирования критически важна именно стабильность на расширенных горизонтах и при пиковых концентрациях. Поэтому дальнейшее развитие методов ПКА должно опираться на интеграцию физически обоснованных графов, адаптивного внимания и гибридной обработки сигналов. Архитектуры класса Multi-gated GNN/GAT представляют наиболее продвинутый и готовый к практическому внедрению подход для надежного долгосрочного прогнозирования качества воздуха.

Литература

1. Li P., Zhang T., Jin Y. A Spatio-Temporal Graph Convolutional Network for Air Quality Prediction // Sustainability. - 2023. - Vol. 15. No. 9. - Art. 7624. - DOI: 10.3390/su15097624.
2. Лыченко, Н. М. Модели прогноза качества воздуха на основе методов машинного обучения с учетом временных и пространственных зависимостей / Н. М. Лыченко // Проблемы автоматизации и управления. – 2024. – № 3(51). – С. 136–145. – EDN PJSEPL.
3. Devasekhar V., Natarajan P. Prediction of Air Quality and Pollution using Statistical Methods and Machine Learning Techniques // International Journal of Advanced Computer Science and Applications. - 2023. - Vol. 14. - No. 4.
4. Лыченко, Н. М. ПРИМЕНЕНИЕ LSTM-нейронных сетей ДЛЯ КЛАССИФИКАЦИИ индекса качества воздуха Г. БИШКЕК / Н. М. Лыченко, А. В. Сороковая // Проблемы автоматизации и управления. – 2020. – № 1(38). – С. 70–80. – DOI 10.5281/zenodo.3904130. – EDN QQZNF1.
5. Li D., Yu H., Geng Y.-a., Li X., Li Q. DDGNet: A dual-stage dynamic spatio-temporal graph network for PM_{2.5} forecasting // В сб.: 2021 IEEE International Conference on Big Data (Big Data). - Piscataway, NJ: IEEE, 2021. - С. 1679-1685. - DOI: 10.1109/BigData52589.2021.9671941
6. Bai S., Kolter J. Z., Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling [Электронный ресурс]. - arXiv:1803.01271, 2018. - Режим доступа: <https://arxiv.org/pdf/1803.01271> (дата обращения: 14.09.2025).
7. Bahdanau D., Cho K., Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate // В сб.: *International Conference on Learning Representations (ICLR)*. – 2015. – Режим доступа: <https://arxiv.org/abs/1409.0473> (дата обращения: 8.11.2025).
8. Veličković P., Cucurull G., Casanova A., Romero A., Liò P., Bengio Y. Graph Attention Networks // В сб.: *International Conference on Learning Representations (ICLR)*. – 2018. – Режим доступа: <https://arxiv.org/abs/1710.10903> (Дата обращения: 8.11.2025).
9. Iskandaryan D., Ramos F., Trilles S. Graph Neural Network for Air Quality Prediction: A Case Study in Madrid // IEEE Access. - 2023. - Vol. 11. - P. 2730-2742. - DOI: 10.1109/ACCESS.2023.3234214.

10. Liu H., Han Q., Sun H., Sheng J., Yang Z. Spatiotemporal adaptive attention graph convolution network for city-level air quality prediction // *Scientific Reports*. - 2023. - Vol. 13. - Art. 13335. - DOI: 10.1038/s41598-023-39286-0.
11. Kim D.-Y., Jin D.-Y., Suk H.-I. Spatiotemporal graph neural networks for predicting mid-to-long-term PM2.5 concentrations // *Journal of Cleaner Production*. - 2023. - Vol. 425. - Art. 138880. - DOI: 10.1016/j.jclepro.2023.138880.
12. Liao H., Yuan L., Wu M. A spatiotemporal graph neural network with multi granularity for air quality prediction [Электронный ресурс]. - Режим доступа: [paper2.pdf](#) (дата обращения: 17.09.2025).
13. Wang X., Zhang S., Chen Y., He L., Ren Y., Zhang Z., Li J., Zhang S. Air quality forecasting using a spatiotemporal hybrid deep learning model based on VMD-GAT-BiLSTM // *Scientific Reports*. - 2024. - Vol. 14. - Art. 17841. - DOI: 10.1038/s41598-024-68874-x.