Spatio-Temporal Field Neural Networks for Air Quality Inference

Yutong Feng[1] , **Qiongyan Wang**[1] , **Yutong Xia**[2] , **Junlin Huang**[1] , **Siru Zhong**[1] , **Yuxuan Liang**[1,3*]
[1]Hong Kong University of Science and Technology (Guangzhou), China
[2]National University of Singapore, Singapore
[3]State Key Lab of Resources and Environmental Information System, China
{yfeng083,jhuang688,szhong691}@connect.hkust-gz.edu.cn,
{yutong.x,qiongyanwang,yuxliang}@outlook.com

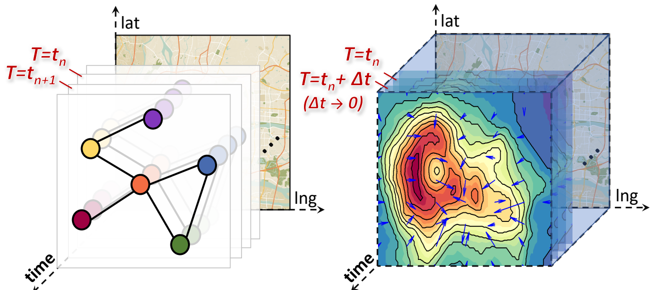
Abstract

Air quality inference aims to utilize historical data from a limited number of observation sites to infer the air quality index at unknown locations. Considering data sparsity due to the high maintenance cost of stations, good inference algorithms can effectively save the cost and refine the data granularity. While spatio-temporal graph neural networks have made excellent progress on this problem, their non-Euclidean and discrete data structure modeling of reality limits its potential. In this work, we make the first attempt to combine two different spatio-temporal perspectives, fields and graphs, by proposing a new model, Spatio-Temporal Field Neural Network, and its corresponding new framework, Pyramidal Inference. Extensive experiments validate that our model achieves state-of-the-art performance in nationwide air quality inference in the Chinese Mainland, demonstrating the superiority of our proposed model and framework.

1 Introduction

Real-time monitoring of air quality, such as PM2.5, PM10, and NO_2 concentrations, is crucial for air pollution control and protecting human health, with air pollution contributing to seven million deaths annually according to the WHO [Vallero, 2014]. However, the deployment of air quality stations in urban areas is limited due to high costs, requiring around 200,000 USD for construction and 30,000 USD annually for maintenance [Zheng *et al.*, 2013]. Additionally, these stations need significant land and dedicated personnel for upkeep, further limiting their prevalence in cities.

In the past decade, substantial research endeavors have been directed towards *air quality inference* [Han *et al.*, 2023], seeking to infer real-time air quality in locations devoid of monitoring stations by leveraging data gleaned from existing sites, as shown in Figure 2(b)-(c). With recent advancements in deep learning, Graph Neural Networks (GNN) [Kipf and Welling, 2016a] have become dominant for non-Euclidean data representation, particularly in learning complex spatial correlations among air quality monitoring stations. Integrating GNNs with temporal learning modules (e.g., RNN [Graves, 2013], TCN [Bai *et al.*, 2018], ODE [Liang *et al.*, 2022]) has led to the development of *Spatio-Temporal Graph Neural Networks* (STGNN)[Wang *et al.*, 2020; Jin *et al.*, 2023], addressing the dynamic nature of air quality data across spatial and temporal dimension. STGNNs, exemplified in studies like [Han *et al.*, 2021; Hu *et al.*, 2023b] offer superior representation extraction and flexibility in cross-domain data fusion.



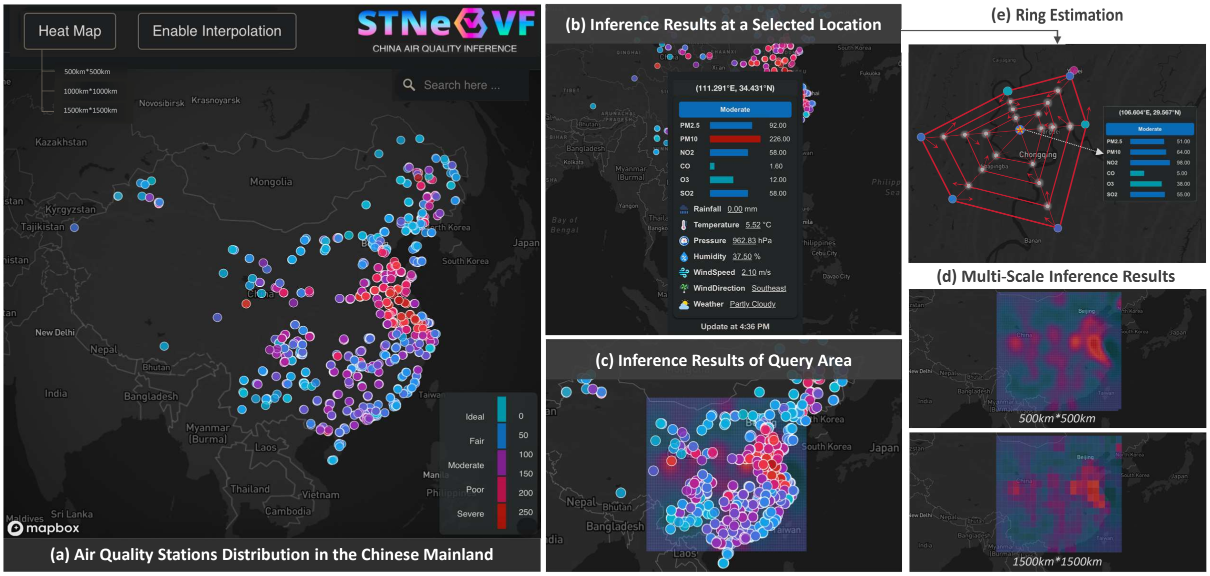
(a) Spatio-Temporal Graph (b) Spatio-Temporal Field

Figure 1: Spatio-Temporal Graph vs. Spatio-Temporal Field.

Though promising, STGNNs simply treat air quality data as a Spatio-Temporal Graph (STG), as shown in Figure 1(a). However, these models overlook a crucial property – *continuity*, which manifests across both spatial and temporal dimensions. In reality, air quality readings of stations are sampled from a continuous Euclidean space and cannot be fully encapsulated by a discrete graph structure using GNNs. Meanwhile, the temporal modules (e.g., RNN, TCN) in STGNNs exhibit the discrete nature as well, rendering them incapable of capturing continuous-time dynamics within data. To better represent the continuous and evolving nature of real-world air quality phenomena, a more powerful approach is needed, surpassing the discrete representation of STGNNs.

In this paper, we draw inspiration from Field Theory [McMullin, 2002] and innovatively formulate air quality inference from a *field* perspective, where air quality data is a physical quantity that can be conceptualized by a new concept called *Spatio-Temporal Fields* (STF), as depicted in Figure 1(b). These fields encompass three dimensions (i.e., latitude, longitude, time), assigning a distinct value to each point in spacetime. In contrast to STGs, STFs are characterized by

[*]Corresponding author. Email: yuxliang@outlook.com


Heat Map
Enable Interpolation
STNeVF
CHINA AIR QUALITY INFERENCE
500km*500km
1000km*1000km
1500km*1500km
Search here ...
Ideal
Fair
Moderate
Poor
Severe
0
50
100
150
200
250
mapbox
(a) Air Quality Stations Distribution in the Chinese Mainland
(b) Inference Results at a Selected Location
(c) Inference Results of Query Area
(d) Multi-Scale Inference Results
500km*500km
1500km*1500km
(e) Ring Estimation
Chongqing

(111.291°E, 34.431°N)

Moderate

PM2.5	92.00
PM10	226.00
NO2	58.00
CO	1.60
O3	12.00
SO2	58.00

Rainfall 0.00 mm

Temperature 5.52 °C

Pressure 962.83 hPa

Humidity 37.50 %

WindSpeed 2.10 m/s

WindDirection Southeast

Weather Partly Cloudy

Update at 4:36 PM

(106.604°E, 29.567°N)

Moderate

PM2.5	51.00
PM10	64.00
NO2	98.00
CO	5.00
O3	38.00
SO2	55.00

Figure 2: (a)-(d): User interface of our STFNN system for air quality inference. (e): An illustration of Ring Estimation.

being **regular**, **continuous**, and **unified**[1], offering a representation more aligned with reality. Under this perspective, we can transform the air quality inference problem to *reconstruct* STFs from available readings using coordinate-based neural networks, particularly Implicit Neural Representations (INR) [Sitzmann *et al.*, 2020; Xie *et al.*, 2022].

While INRs effectively handle the continuity property of air quality data, they inevitably confront two primary challenges. Firstly, the generation process of air quality data is extremely complex and influenced by various factors (such as humidity and wind speed/direction), which poses a challenge for reconstructing the underlying STFs through neural representation methods. Secondly, empirical studies [Xu *et al.*, 2019; Sitzmann *et al.*, 2020] verify that INRs always exhibit a bias towards learning low-frequency functions, which will disregard locally varying high-frequency information and higher-order derivatives even with dense supervision.

To this end, we for the first time present **Spatio-Temporal Field Neural Networks** (**STFNN**), opening new avenues for modeling spatio-temporal fields and achieving state-of-the-art performance in nationwide air quality inference in the Chinese Mainland. Targeting the first challenge, we pivot our focus from reconstructing the value of each entry in STFs to learning the derivative (i.e., gradient) of each entry. This strategic shift is inspired by learning the residual is often easier than learning the original value directly, as exemplified in ResNet [He *et al.*, 2016]. Such vector field can not only show how the pollutant concentration varies across time and space but also the direction of diffusion. To tackle the second challenge, we endeavor to augment our STFNN with local context knowledge during air quality inference at a specific location. Specifically, we combine the STGNN's capability to capture local spatio-temporal dependencies with STFNN's ability to learn global spaito-temporal unified representations. This integration results in what we term **Pyramid Inference**, a hybrid framework that leverages the strengths of both models to achieve a more comprehensive inference of air quality dynamics with both high-frequency and low-frequency components. Overall, our contributions lie in three aspects:

- *A Field Perspective*. We formulate air quality as spatio-temporal fields with the first shot. Compared to STGs, our STFs not only adeptly capture the continuity and Euclidean structure of air quality data, but also achieves a unified representation across both space and time.
- *Spatio-Temporal Field Neural Networks*. We propose a groundbreaking network called STFNN to model STF data. STFNN pioneers an implicit representation of the STF's gradient, deviating from conventional direct estimation approaches. Moreover, it preserves high-frequency information via Pyramid Inference.
- *Empirical Evidence*. We conduct extensive experiments to evaluate the effectiveness of our STFNN. The results valiadte that STFNN outperforms prior arts by a significant margin and exhibits compelling properties. A system in Figure 2 has been deployed to show its practicality in the Chinese Mainland.

2 Preliminary

Definition 1 (Air Quality Reading) We use $\mathbf{x}_t^i \in \mathbb{R}^D$ and $\mathbf{y}_t^i \in \mathbb{R}$ to denote the air quality readings and the concentration of PM2.5 from the i-th monitoring stations at time t separately. Here D encompasses various measurements, such as concentrations of other air pollutants (e.g., PM10, NO_2), and meteorological properties (e.g. humidity, weather and wind speed). $\mathbf{X}_t = \left(\mathbf{x}_t^1, \mathbf{x}_t^2, \ldots, \mathbf{x}_t^N\right) \in \mathbb{R}^{N \times D}$ denotes the observations of all stations at a specified time t. $\mathcal{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T) \in \mathbb{R}^{T \times N \times D}$ denotes the observations of all stations at all time. Similar definitions apply to $\mathbf{Y}_t$ and $\mathcal{Y}$, mirroring $\mathbf{X}_t$ and $\mathcal{X}$, respectively.

[1] It implies that the field representation accounts for variations not only across different locations in space but also over different points in time, emphasizing the comprehensive treatment of both spatial and temporal aspects within the unified framework.

Definition 2 (Coordinates) A coordinate $\mathbf{c} = [lng, lat, t] \in \mathbb{R}^3$ is used to represent the spatial and temporal properties of an air quality reading or a location, including longitude, latitude, and timestamp. These coordinates are categorized into two types: *source* coordinate $\mathbf{c}^{src}$, associated with readings or locations with existing air quality monitoring stations, while *target* coordinate $\mathbf{c}^{tar}$, corresponding to unobserved locations requiring inference. Notably, $\mathbf{c}_t^i$ represents the coordinate of the corresponding $\mathbf{x}_t^i$ and $\mathbf{y}_t^i$. Parallel definitions apply to $\mathbf{C}_t$ and $\mathcal{C}$ in relation to $\mathbf{X}_t$ and $\mathcal{X}$, respectively, which are not reiterated here.

Problem Definition The air quality inference problem addresses the utilization of historical data and real-time readings from a limited number of air quality monitoring stations to infer the real-time air quality *anywhere*, especially unobserved location. Traditional strategies [Hou *et al.*, 2022; Hu *et al.*, 2023a] employ graphs to illustrate the relationship between stations and locations, and the task is translated into a recovery task for the masked nodes (target locations), as shown in Figure 3 (a). In this paper, we revisit the problem from the field perspective, as shown in Figure 3 (b). Specifically, our goal is to reconstruct a spatio-temporal field G for air quality that is capable of mapping any arbitrary coordinate, especially $\mathbf{c}^{tar}$, to the corresponding concentration of PM2.5 $\mathbf{y}^{tar}$. Additional parameters, such as $\mathcal{X}$, are allowed to enhance the inference process.

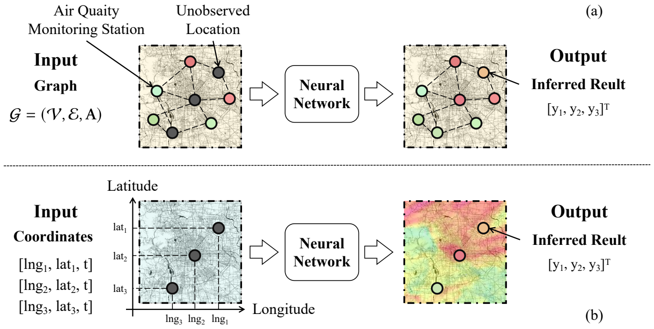


Figure 3: Paradigms for air quality inference. (a) A spatio-temporal graph perspective. (b) A spatio-temporal field perspective.

3 Methodology

3.1 Global View: Spatio-Temporal Field

A Spatio-Temporal Field (STF) is a global modeling of air quality that encompasses all stations and observation times. The STF function, denoted as $f(\cdot) : \mathbf{c} \longmapsto \mathbf{q}$, assigns a unique physical quantity $\mathbf{q}$ to each coordinate. When $\mathbf{q}$ is a scalar, $f(\cdot)$ represents a scalar field. Conversely, if $\mathbf{q}$ is a vector, with magnitude and direction, it denotes a vector field.

Specifically, our focus lies on a scalar field $G : \mathbb{R}^3 \to \mathbb{R}$ for air quality inference, where G maps the coordinates to the corresponding PM2.5 concentration. This representation facilitates a continuous and unified spacetime perspective, allowing for the inference of air quality at any location and time by inputting the coordinates.

Directly modeling G is challenging due to its intricate complexity and nonlinearity. Alternatively, it is often more feasible to learn its derivative, which refers to the gradient field of G in spacetime. We denote the gradient field as $\mathbf{F} \triangleq \nabla G$ which is a vector field. Notably, given a specific $\mathbf{F}$, an array of G solutions exists unless an initial value is specified. We use $\mathbf{y}^{src}$ and $\mathbf{y}^{tar}$ to denote the PM2.5 concentration on $\mathbf{c}^{src}$ and $\mathbf{c}^{tar}$, respectively. Our primary focus lies in inferring $\mathbf{y}^{tar}$ since the true value of $\mathbf{y}^{src}$ is known and recorded while $\mathbf{y}^{tar}$ remains undisclosed. To infer $\mathbf{y}^{tar}$, we utilize a $\mathbf{y}^{src}$ as the initial value and assume l is a piecewise smooth curve in $\mathbb{R}^3$ that point from $\mathbf{c}^{src}$ to $\mathbf{c}^{tar}$, then we have

$$\begin{aligned} \mathbf{y}^{tar} &= G\left(\mathbf{c}^{tar}\right) = G\left(\mathbf{c}^{src}\right) + \int_l \nabla G(\mathbf{r}) \cdot d\mathbf{r} \\ &= \mathbf{y}^{src} + \int_a^b \mathbf{F}\big(\mathbf{r}(z)\big) \cdot \mathbf{r}'(z) dz \end{aligned} \tag{1}$$

where $\cdot$ is the dot product, and $\mathbf{r} : [a, b] \to l$ represents the position vector. The endpoints of l are given by $\mathbf{r}(a)$ and $\mathbf{r}(b)$, with $a < b$. Seeking an Implicit Neural Representation (INR) becomes the objective to fit $\mathbf{F}$ as $\mathbf{F}$ is usually intricate to the extent that it cannot be explicitly formulated.

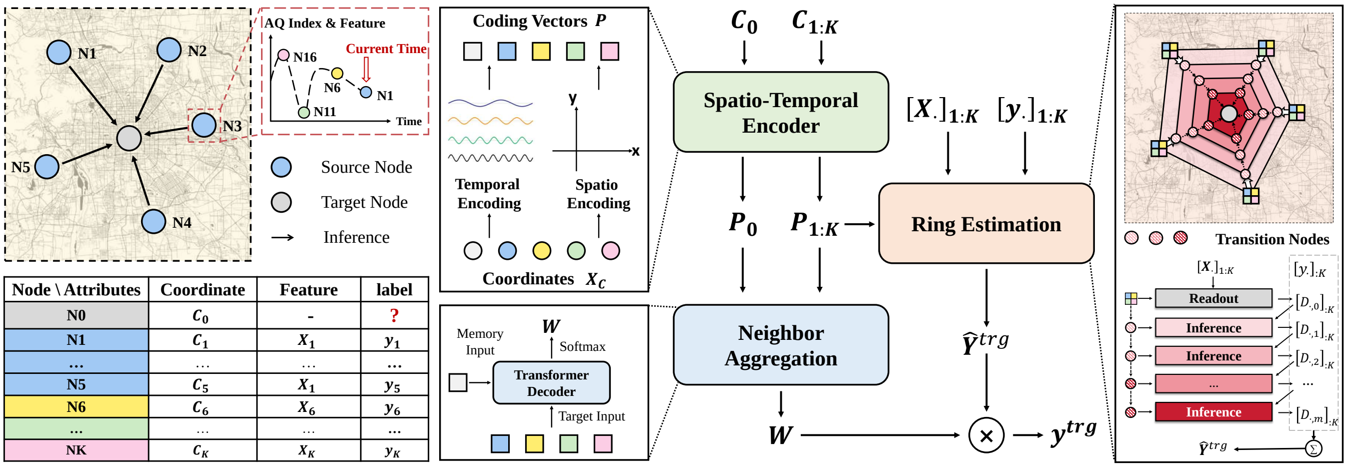
3.2 Local View: Spatio-Temporal Graph

The formulation presented in Eq. (1) ensures the recoverability of $\mathbf{c}^{tar}$ across arbitrary coordinates through curve integration, utilizing solely a single initial value. However, this approach yields an excessively coarse representation of the entire STF, resulting in the loss of numerous local details and high-frequency components The impact is particularly pronounced when $\mathbf{y}^{src}$ is situated at a considerable distance from $\mathbf{y}^{tar}$ as the increase in the length of l introduces a significant cumulative error. In response to this limitation, we leverage the potent learning capabilities of STGNN to capture local spatio-temporal correlations effectively. We employ the local spatio-temporal graph (STG) to model the spatio-temporal dependencies of the given coordinates and their neighboring air monitoring stations with their histories. The corresponding design can be found in the foundational work [Song *et al.*, 2020], and it is not necessary to reiterate it here.

3.3 Hybrid Framework: Pyramidal Inference

We intend to integrate the continuous and uniform global modeling of spacetime provided by STF with the local detailing capabilities of STG, thereby establishing a hybrid framework that leverages the strengths of both approaches. Within the local STG, the estimation of $\mathbf{y}^{tar}$ is achieved by leveraging information from its neighboring nodes through Eq. (1). By calculating estimates of $\mathbf{y}^{tar}$ from these neighbors and assigning a learnable weight w_i to each estimation ($\sum_{i=1}^{K} w_i = 1$), we enhance the precision of the inference result for a specific coordinate. This operation can be formulated as

$$\begin{aligned} \hat{\mathbf{y}}^{tar} &= \sum_{i=1}^{K} \left[w_i \cdot \left(\mathbf{y}_i^{src} + \int_{l_i} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} \right) \right] \\ &= \sum_{i=1}^{K} \left[w_i \cdot \left(\mathbf{y}_i^{src} + \int_{a_i}^{b_i} \mathbf{F}\big(\mathbf{r}(z)\big) \cdot \mathbf{r}'(z) dz \right) \right] \end{aligned} \tag{2}$$



Node \ Attributes	Coordinate	Feature	label
N0	C_0	-	?
N1	C_1	X_1	y_1
...	...	...	...
N5	C_5	X_1	y_5
N6	C_6	X_6	y_6
...	...	...	...
NK	C_K	X_K	y_K

Figure 4: Implementation of STFNN

where $\hat{\mathbf{y}}^{tar}$ is the joint estimation of $\mathbf{y}^{tar}$ by neighbors. w_i and $\mathbf{y}_i^{src}$ represent the weight and PM2.5 concentration of the i_{th} neighbor, respectively. l_i is the integral path in $\mathbb{R}^3$ that points from the coordinates of the i_{th} neighbor to the target coordinate.

We call the inference strategy represented by Eq. (2) **Pyramidal Inference**, which is the framework of the STFNN we propose. To demonstrate its sophistication, we deconstruct Eq. (2) into two steps:

$$\hat{\mathbf{Y}}^{tar} = \begin{bmatrix} \hat{\mathbf{y}}_1 \\ \vdots \\ \hat{\mathbf{y}}_K \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^{src} + \int_{\mathcal{C}_i} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} \\ \vdots \\ \mathbf{y}_K^{src} + \int_{\mathcal{C}_K} \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} \end{bmatrix} \tag{3}$$

and

$$\hat{\mathbf{y}}^{tar} = [\ w_1 \quad \dots \quad w_K\ ] \cdot \hat{\mathbf{Y}}^{tar} = \mathbf{W}^T \cdot \hat{\mathbf{Y}}^{tar}, \tag{4}$$

where $\hat{\mathbf{y}}_i$ is the estimate of $\mathbf{y}^{tar}$ by the i_{th} neighbor, $\mathbf{W}$ and $\hat{\mathbf{Y}}^{tar}$ represent the vector form of w_i and $\hat{\mathbf{y}}_i$, respectively. These two operations can be viewed as follows. In Eq. (3), a curve integral is used over the gradient field to estimate $\mathbf{y}^{tar}$ from each neighbor in a continuous and spatio-temporally uniform way, which takes advantage of the STF. In Eq. (4), the information from the neighbors is aggregated through $\mathbf{W}$, which considers the spatio-temporal dependencies between the nodes and takes advantage of the STG. In this way, the Pyramidal Inference framework combines the two different spacetime perspectives into a distinctive new paradigm.

4 Implementation

Upon introducing the formulation of Pyramidal Inference, we proceed to its implementation through a meticulously designed model architecture, depicted in Figure 4. The model comprises three pivotal components:

- **Spatio-Temporal Encoding.** This component transforms coordinates into coded vectors endowed with representational meaning, enhancing the network's ability to comprehend and leverage the spatio-temporal characteristics of coordinates.
- **Ring Estimation:** Implementation of Eq. (3), mapping the encoded vector to the gradient of the STF. This process yields each neighbor's estimate of the PM2.5 concentration for the target coordinates through a path integral.
- **Neighbor Aggregation:** Implementation of Eq. (4), utilizing the coded vectors of neighbors and target coordinates as inputs to derive the estimated weights for each neighbor concerning the target coordinates.

In the following parts, we will provide a detailed exposition of each module, elucidating their functionalities step by step.

4.1 Spatio-Temporal Encoding

We revisit the previously introduced local STG, which amalgamates nodes across different time steps into a unified graph, potentially obscuring the inherent temporal properties of individual nodes. In essence, this local STG places nodes from diverse time steps into a shared environment without discerning their temporal distinctions. However, this issue can be mitigated through meticulous positional coding of nodes [Gehring *et al.*, 2017; Song *et al.*, 2020].

We use $\mathbf{p} \in \mathbb{R}^{10}$ to denote the coding vector, essential for accurately describing the spatio-temporal characteristics of a node or coordinate. This vector is expressed as the concatenation $\mathbf{p} = [\mathbf{p}_S, \mathbf{p}_T]$, where $\mathbf{p}_S$ represents spatial coding and $\mathbf{p}_T$ represents temporal coding. In the spatial dimension, a node's properties can be captured by its absolute position, represented by z-normalized longitude lng_z and latitude lat_z, forming $\mathbf{p}_S = [lng_z, lat_z]$. For encoding temporal information $\mathbf{p}_T$, sinusoidal functions with different periods are employed, reflecting the periodic nature of temporal phenomena. We utilize a set of periods $\mathbf{T} = \{1a, 7a, 30.5a, 365a\}$, with a as the scaling index, to represent days, weeks, months, and years. The temporal coding $\mathbf{p}_T$ is then represented as

$$\mathbf{p}_{(T,i)} = \begin{cases} sin\left(2\pi t/\mathbf{T}_{int(i/2)+1}\right) & i \bmod 2 = 0 \\ cos\left(2\pi t/\mathbf{T}_{int(i/2)+1}\right) & i \bmod 2 \neq 0 \end{cases} \tag{5}$$

where $\mathbf{p}_{(T,i)}$ denotes the value of the i_{th} dimension ($1 \leq i \leq 8$) of $\mathbf{p}_T$, $int(i/2)$ denotes dividing i by 2 and rounding down, and $\mathbf{T}_{int(i/2)+1}$ is the $int(i/2) + 1$ period of $\mathbf{T}$.

4.2 Ring Estimation

Motivation. We employ the continuous approach of curve integration over the gradient to determine the PM2.5 concentration at the target coordinate in Eq. (3). However, due to inherent limitations in numerical accuracy within computing systems, achieving true continuity in STF becomes unattainable. Therefore, we have adopted an incremental approach. We set the integral path to be a straight line from the neighbors' coordinate $\mathbf{c}^{src}$ to the target coordinate $\mathbf{c}^{tar}$ for convenience, then the unit direction vector $\vec{\mathbf{r}}$ in the path can be written as $\vec{\mathbf{r}} \triangleq (\mathbf{c}^{tar} - \mathbf{c}^{src}) \big/ (\|\mathbf{c}^{tar} - \mathbf{c}^{src}\|)$. After that, we replace the integral operation with a summation operation and modify Eq. (3) to a discrete form

$$\hat{\mathbf{y}}^{tar} = \sum_{i=1}^{K} \left[w_i \cdot \left(\mathbf{y}_i^{src} + \sum_{j=1}^{m} \mathbf{D}_{i,j} \cdot \vec{\mathbf{r}_i} \right) \right] \tag{6}$$

where m represents the step size of the summation and $\mathbf{c}_i^{src}$ is the coordinate of the i_{th} node in the local STG. $\mathbf{D}_{i,j} \in \mathbb{R}^3$ represents the *difference* at the j_{th} step of the i_{th} node, which is the discrete approximation of the gradient. Our objective is to build a module for estimating $\mathbf{D}_{i,j}$, which is the only unknown in Eq. (6).

Overview. Towards this objective, we introduce a pivotal module named Ring Estimation, designed for the joint estimation of the differences $[\mathbf{D}_{\cdot,j}]_{:K} = [\mathbf{D}_{1,j}, \cdots, \mathbf{D}_{K,j}] \in \mathbb{R}^{K\times 3}$ at the j_{th} step of all neighbors. We posit that simultaneous estimation of $[\mathbf{D}_{\cdot,j}] : K$ enhances inference efficiency and captures correlations between them, thereby reducing estimation errors compared to individually estimating $\mathbf{D}_{i,j}$ for a single neighbor K times. Specifically, the Ring Estimation module divides the polygon surrounded by K neighbors around the target coordinate into m ring zones from the outermost to the innermost, with total mK of *transition* nodes (coordinates) uniformly spaced along the inference path. The inner edge of the j_{th} ring zone serves as the outer edge for the $(j+1)_{th}$ zone. Like the target coordinate, the transition nodes lack features and labels (PM2.5 concentration). They serve as the intermediary states and springboards in the process of estimating $\hat{\mathbf{y}}^{tar}$. By increasing the value of m, the Ring Estimation block facilitates the inference in an approximately continuous manner.

4.3 Neighbor Aggregation

It is advisable to assign varying weights $\mathbf{W} = [w1, \ldots, w_K]$ to the estimations of the target nodes based on the different spatio-temporal scenarios in which they are situated. To this end, we present the Neighbor Aggregation module, which takes into account the coding of the coordinates of the neighbors and the target coordinate and employs end-to-end learning to compute the estimation weights w_i of each neighbor on the target node. To obtain $\mathbf{W}$, we first multiplied the output by $W_N \in \mathbb{R}^{10\times 1}$ to transform it into a Logit score. Then, we applied the $Softmax$ operation to ensure that the weights sum up to one. In this end, the formulation of the Neighbor Aggregation can be written as

$$\mathbf{W} = Softmax\left(W_N \cdot Decoder(\mathbf{P}^{src}, \mathbf{P}^{tar})\right) \tag{7}$$

5 Experiments

In this section, we delve into our experimental methodology aimed at evaluating the performance and validating the efficacy of the STFNN. Specifically, our experiments are designed to explore the following research questions, elucidating key aspects of our approach and its applicability in real-world scenarios:

- **RQ1**: How does STFNN's approach, focusing on inferring concentration gradients for indirect concentration value inference, outperform traditional methods in terms of accuracy and effectiveness?
- **RQ2**: What specific contributions do the individual components of STFNN make to its effectiveness in inferring air pollutant concentrations?
- **RQ3**: How do variations in each hyperparameter impact the overall performance of STFNN?
- **RQ4**: Do the three-dimensional hidden states learned by the model accurately represents the gradient of the spatio-temporal field?
- **RQ5**: Can our model demonstrate proficient performance in inferring concentrations of various air pollutants, including NO_2?

5.1 Experimental Settings

Datasets

The study obtained a nationwide air quality dataset [Liang *et al.*, 2023] from January 1st, 2018, to December 31st, 2018. This dataset includes air quality and meteorological data. The input data can be divided into two classes: continuous and categorical data. Continuous data includes critical parameters such as air pollutant concentrations (e.g., PM2.5, CO), temperature, wind speed, and others. Categorical data encompasses weather, wind direction, and time. In the event of an unanticipated occurrence, such as a power outage, some data may be unavailable.

5.2 Baselines for Comparison

We compare our STFNN with the following baselines that belong to the following four categories:

- **Statistical models**: **KNN** [Guo *et al.*, 2003] utilizes non-parametric, instance-based learning, inferring air quality by considering data from the nearest neighbors. **Random Forest (RF)** [Fawagreh *et al.*, 2014] aggregates interpolation from diverse decision trees, each trained on different dataset subsets, providing robust results.
- **Neural Network based models**: **MCAM** [Han *et al.*, 2021] introduces multi-channel attention blocks capturing static and dynamic correlations.
- **Neural Processes based models**: **SGNP**, a modification of Sequential Neural Processes (SNP) [Singh *et al.*, 2019], incorporates a cross-set graph network before aggregation, enhancing air quality inference. **STGNP** [Hu *et al.*, 2023a] employing a Bayesian graph aggregator for context aggregation considering uncertainties and graph structure.

Model	Year	#Param(M)	Mask Ratio = 25%				Mask Ratio = 50%				Mask Ratio = 75%			
			MAE	**Δ**	RMSE	MAPE	MAE	**Δ**	RMSE	MAPE	MAE	**Δ**	RMSE	MAPE
KNN	1967	-	30.50	+146.0%	65.40	1.36	30.25	+145.5%	72.23	0.71	34.07	+194.0%	74.55	0.64
RF	2001	-	29.22	+135.6%	68.95	0.76	29.71	+141.2%	71.61	0.75	29.82	+157.3%	70.99	0.74
MCAM	2021	0.408	23.94	+93.1%	36.25	0.95	25.01	+103.0%	37.94	0.92	25.19	+117.3%	37.82	1.04
SGNP	2019	0.114	23.60	+90.3%	37.58	0.83	24.06	+95.3%	37.08	0.93	21.68	+87.1%	33.68	0.84
STGNP	2022	0.108	23.21	+87.2%	38.13	0.62	21.95	+78.2%	37.13	0.67	19.58	+68.9%	31.95	0.69
VAE	2013	0.011	28.49	+129.8%	67.11	0.94	28.92	+134.7%	69.67	0.94	29.00	+150.2%	69.11	0.93
GAE	2016	0.073	12.63	+1.9%	23.80	0.46	12.78	+3.7%	24.11	0.46	12.57	+8.5%	23.73	0.46
GraphMAE	2022	0.073	12.40	-	23.20	0.46	12.32	-	23.11	0.46	11.59	-	21.51	0.43
STFNN	-	0.208	**11.14**	**-10.2%**	**19.75**	**0.39**	**11.32**	**-8.1%**	**19.91**	**0.42**	**11.27**	**-2.8%**	**19.86**	**0.41**

Table 1: Model comparison on the nationwide dataset. The parameter count, denoted as #Param, is in the order of million (M). The symbol Δ represents the reduction in MAE compared to GraphMAE. The mask ratio represents the proportion of unobserved nodes to all nodes.

- **AutoEncoder based models**: **VAE** [Kingma and Welling, 2022] applies variational inference to air quality inference, utilizing reconstruction for target node inference. **GAE** [Kipf and Welling, 2016b] reconstructs node features within a graph structure, while **GraphMAE** [Hou *et al.*, 2022] introduces a masking strategy for innovative node feature reconstruction.

5.3 Hyperparameters & Setting

To mitigate their impact, instances exceeding a 50% threshold of missing data at any given time were prudently omitted from our analysis. Our dataset was carefully partitioned into three segments: a 60% training set, a 20% validation set, and a 10% test set. During training, in each epoch, we randomly select stations with ratio α, mask their features and historical information, and let them act as the target node. We ignore all locations where PM2.5 (or NO_2 in the case of **RQ5**) is missing. The model is trained with an Adam optimizer, starting with a learning rate of 1E-3, reduced by half every 40 epochs during the 200 training epochs. The batch size for training is set to 32. The hidden dimension of MLP and all the Transformer-Decoder networks is fixed at 64. For Ring Estimation, the neighbor number is set to 6, incorporating the past 6 timesteps, and the iteration step m is defined as 16.

5.4 Model Comparison (RQ1)

In addressing RQ1, we conduct a meticulous comparative analysis among models based on the evaluation metrics of MAE, RMSE, and MAPE. The empirical outcomes derived from this analysis are systematically presented across the expansive spectrum of the nationwide air quality dataset, meticulously documented within Table 1.

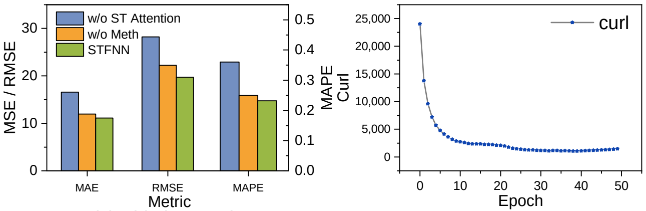
The results indicate that STFNN consistently demonstrates enhanced efficacy across various evaluation metrics, surpassing existing baseline models. Table 1 shows that our approach reduces MAE under three different mask ratios (25%, 50%, and 75%) in comparison to GraphMAE, establishing a new State-of-the-Art (SOTA) in nationwide PM2.5 concentration inference in the Chinese Mainland. In our view, there are three main reasons for this. First, the gradient field is a better representation of reality. Second, the spatial and temporal modules of STFNN capture both types of information together, avoiding bias or information loss. Finally, our Pyramidal Inference framework captures global and local spatio-temporal properties, which helps us model the pollutant concentration field more accurately.

5.5 Ablation Study (RQ2)

To assess the contributions of individual components to the performance of our model and address RQ2, we conducted ablation studies. The findings from these studies are presented in Figure 5 (a).

Effects of meteorological features. To analyze the impact of meteorological features on the accuracy of the final model, we removed them from the raw data. Therefore, the gradient was obtained solely from the spatio-temporal coordinates of neighboring stations fed into the gradient encoder. The figure demonstrates that removing the meteorological features resulted in some improvement in the model's mean absolute error (MAE), which still outperformed all baseline models.

Effects of Neighbor Aggregation. To investigate the impact of a dynamic and learnable implicit graph structure on the model, we substituted the model's Neighbor Aggregation module with IDW and SES, a non-parametric approach inspired by Zheng et al [Zheng *et al.*, 2013]. This approach employs implicit graph relations that are static. The results depicted in Figure 5 (a) demonstrate that utilizing the Neighbor Aggregation module results in a significant decrease in MAE.



(a) Ablation Study (b) Variation of curl

Figure 5: (a) ablation study (b) the variation of curl

5.6 Hyperparameters Study (RQ3)

In this section, we comprehensively explore the effects of various hyperparameters on the model's performance, thereby addressing RQ3.

Effects of Hidden & FFD Dimension. We adjusted the hidden layer dimension of the Spatio-Temporal Encoding module and the forward propagation of the Transformer-Decoder structure used by the Ring Estimation and Neighbor Aggregation modules from 16 to 64. The results in Figure 6 (a-b) show that adjusting the hidden layer dimension has little effect on the absolute values of MAE and RMSE, but it significantly decreases MAPE.

Effects of Step Size. We vary the value of the accumulation step size of the Ring Estimation in $\{2, 4, 8, 16\}$. The result is shown in Figure 6 (c). It has been observed that as the step size increases, the model's performance initially declines before improving. Additionally, when comparing 2 steps to 16 steps, we note that the model's training time per round is approximately 50% longer for 16 steps.

Effects of Neighbors Number. We vary the number of neighbors from 2 to 8, and the result is shown in Figure 6 (d). We observe that the performance of our model improved as the number of neighbors increased. Notably, our proposed STFNN exhibited excellent performance even with a small number of neighbors. Due to its ability to learn global spatio-temporal patterns, the model can use global information for inference even in scenarios where there are only a few neighbors present.

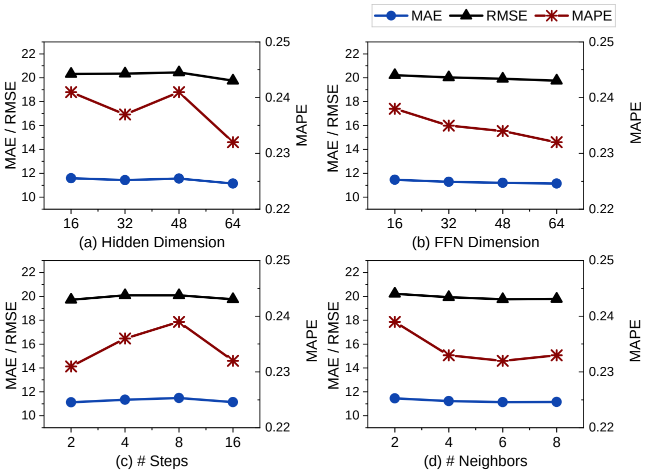


Figure 6: Hyperparameter Study

5.7 Interpretability (RQ4)

To confirm that the network's learned vector is the gradient of the spatio-temporal field, we calculate the curl variation with the number of training epochs. We use Yang et al.'s method [Yang *et al.*, 2023] to quantify the curl and present the experimental results in Figure 5 (b). It is evident that the curl of the vector field obtained by the network decreases as the training progresses, indicating successful learning of the gradient field.

5.8 Generalizability (RQ5)

Our model not only excels in inferring PM2.5 but also establishes a new benchmark, achieving the SOTA in inferring the concentration of NO_2, as shown in Table 2. This noteworthy outcome underscores the versatility of our model across distinct air quality parameters. In comparison to the baseline, our model demonstrates a significant advantage, showcasing its capability to handle diverse pollutants effectively and outperforming established methods in inference for NO_2.

Model	Mask Ratio = 25%		Mask Ratio = 50%		Mask Ratio = 75%	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
KNN	18.10	62.51	18.47	64.22	20.18	62.86
RF	16.90	61.25	17.60	64.91	17.36	63.70
MCAM	18.25	27.80	17.75	27.42	21.17	29.41
SGNP	17.66	25.43	19.17	26.36	16.57	24.11
STGNP	16.43	27.85	15.62	26.06	15.70	26.23
VAE	29.85	112.81	31.43	119.59	30.82	117.33
GAE	12.80	30.16	12.77	30.16	12.77	30.00
GraphMAE	12.76	30.25	12.60	29.53	12.48	29.30
STFNN	**11.34**	**23.65**	**11.52**	**24.93**	**11.97**	**25.81**

Table 2: Experiment result on NO2

6 Related Works

Traditional methods [Hasenfratz *et al.*, 2014; Jumaah *et al.*, 2019] rely on linear spatial assumptions. However, these models only consider simple spatial relationships and do not adapt to complex changes in air quality. In recent times, there's been a growing interest in studying Spatio-Temporal Graph to understand the intricate relationship that involves both spatial and temporal for air quality inference. STGNNs [Jiang *et al.*, 2021; Salim and Haque, 2015; Wang *et al.*, 2020; Sun *et al.*, 2020; Wang *et al.*, 2021], which integrate the strengths of GNNs, have emerged as the leading approach for uncovering intricate relationships in STG data. Some follow-ups [Li *et al.*, 2017; Yu *et al.*, 2017; Geng *et al.*, 2019] introduce temporal components such as Recurrent Neural Networks (RNN) [Graves, 2013] and Temporal Convolutional Networks (TCN) [Bai *et al.*, 2018] to better address the spatio-temporal dependencies. However, The limitation of STGNNs lies in their lack of consideration for contiguity and Euclidean spatial structures.

7 Conclusion

In this work, we introduced a novel perspective for air quality inference, framing it as a problem of reconstructing Spatio-Temporal Fields (STFs) to better capture the continuous and unified nature of air quality data. Our proposed Spatio-Temporal Field Neural Network (STFNN) breaks away from the limitations of Spatio-Temporal Graph Neural Networks (STGNNs) by focusing on implicit representations of gradients, offering a more faithful representation of the dynamic evolution of air quality phenomena.

Acknowledgements

This work is supported by a grant from State Key Laboratory of Resources and Environmental Information System. This study is also funded by the Guangzhou-HKUST(GZ) Joint Funding Program (No. 2024A03J0620).

References

[Bai *et al.*, 2018] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[Fawagreh *et al.*, 2014] Khaled Fawagreh, Mohamed Medhat Gaber, and Eyad Elyan. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering: An Open Access Journal*, 2(1):602–609, 2014.

[Gehring *et al.*, 2017] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. Convolutional sequence to sequence learning. In *International conference on machine learning*, pages 1243–1252. PMLR, 2017.

[Geng *et al.*, 2019] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3656–3663, 2019.

[Graves, 2013] Alex Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.

[Guo *et al.*, 2003] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. Knn model-based approach in classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, pages 986–996. Springer, 2003.

[Han *et al.*, 2021] Qilong Han, Dan Lu, and Rui Chen. Fine-grained air quality inference via multi-channel attention model. In *IJCAI*, pages 2512–2518, 2021.

[Han *et al.*, 2023] Jindong Han, Weijia Zhang, Hao Liu, and Hui Xiong. Machine learning for urban air quality analytics: A survey. *arXiv preprint arXiv:2310.09620*, 2023.

[Hasenfratz *et al.*, 2014] David Hasenfratz, Olga Saukh, Christoph Walser, Christoph Hueglin, Martin Fierz, and Lothar Thiele. Pushing the spatio-temporal resolution limit of urban air pollution maps. In *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 69–77. IEEE, 2014.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Hou *et al.*, 2022] Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. Graphmae: Self-supervised masked graph autoencoders, 2022.

[Hu *et al.*, 2023a] Junfeng Hu, Yuxuan Liang, Zhencheng Fan, Hongyang Chen, Yu Zheng, and Roger Zimmermann. Graph neural processes for spatio-temporal extrapolation. *arXiv preprint arXiv:2305.18719*, 2023.

[Hu *et al.*, 2023b] Junfeng Hu, Yuxuan Liang, Zhencheng Fan, Li Liu, Yifang Yin, and Roger Zimmermann. Decoupling long-and short-term patterns in spatiotemporal inference. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[Jiang *et al.*, 2021] Renhe Jiang, Du Yin, Zhaonan Wang, Yizhuo Wang, Jiewen Deng, Hangchen Liu, Zekun Cai, Jinliang Deng, Xuan Song, and Ryosuke Shibasaki. Dl-traff: Survey and benchmark of deep learning models for urban traffic prediction. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 4515–4525, 2021.

[Jin *et al.*, 2023] Guangyin Jin, Yuxuan Liang, Yuchen Fang, Jincai Huang, Junbo Zhang, and Yu Zheng. Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. *arXiv preprint arXiv:2303.14483*, 2023.

[Jumaah *et al.*, 2019] Huda Jamal Jumaah, Mohammed Hashim Ameen, Bahareh Kalantar, Hossein Mojaddadi Rizeei, and Sarah Jamal Jumaah. Air quality index prediction using idw geostatistical technique and ols-based gis technique in kuala lumpur, malaysia. *Geomatics, Natural Hazards and Risk*, 10(1):2185–2199, 2019.

[Kingma and Welling, 2022] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022.

[Kipf and Welling, 2016a] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Kipf and Welling, 2016b] Thomas N. Kipf and Max Welling. Variational graph auto-encoders, 2016.

[Li *et al.*, 2017] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

[Liang *et al.*, 2022] Yuxuan Liang, Kun Ouyang, Yiwei Wang, Zheyi Pan, Yifang Yin, Hongyang Chen, Junbo Zhang, Yu Zheng, David S Rosenblum, and Roger Zimmermann. Mixed-order relation-aware recurrent neural networks for spatio-temporal forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[Liang *et al.*, 2023] Yuxuan Liang, Yutong Xia, Songyu Ke, Yiwei Wang, Qingsong Wen, Junbo Zhang, Yu Zheng, and Roger Zimmermann. Airformer: Predicting nationwide air quality in china with transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14329–14337, 2023.

[McMullin, 2002] Ernan McMullin. The origins of the field concept in physics. *Physics in Perspective*, 4:13–39, 2002.

[Salim and Haque, 2015] Flora Salim and Usman Haque. Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and internet of things. *International Journal of Human-Computer Studies*, 81:31–48, 2015.

[Singh *et al.*, 2019] Gautam Singh, Jaesik Yoon, Youngsung Son, and Sungjin Ahn. Sequential neural processes, 2019.

[Sitzmann *et al.*, 2020] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

[Song *et al.*, 2020] Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 914–921, 2020.

[Sun *et al.*, 2020] Junkai Sun, Junbo Zhang, Qiaofei Li, Xiuwen Yi, Yuxuan Liang, and Yu Zheng. Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks. *IEEE Transactions on Knowledge and Data Engineering*, 34(5):2348–2359, 2020.

[Vallero, 2014] Daniel A Vallero. *Fundamentals of air pollution*. Academic press, 2014.

[Wang *et al.*, 2020] Senzhang Wang, Jiannong Cao, and S Yu Philip. Deep learning for spatio-temporal data mining: A survey. *IEEE transactions on knowledge and data engineering*, 34(8):3681–3700, 2020.

[Wang *et al.*, 2021] Huandong Wang, Qiaohong Yu, Yu Liu, Depeng Jin, and Yong Li. Spatio-temporal urban knowledge graph enabled mobility prediction. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 5(4):1–24, 2021.

[Xie *et al.*, 2022] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, volume 41, pages 641–676. Wiley Online Library, 2022.

[Xu *et al.*, 2019] Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*, 2019.

[Yang *et al.*, 2023] Xianghui Yang, Guosheng Lin, Zhenghao Chen, and Luping Zhou. Neural vector fields: Generalizing distance vector fields by codebooks and zero-curl regularization. *arXiv preprint arXiv:2309.01512*, 2023.

[Yu *et al.*, 2017] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

[Zheng *et al.*, 2013] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1436–1444, 2013.