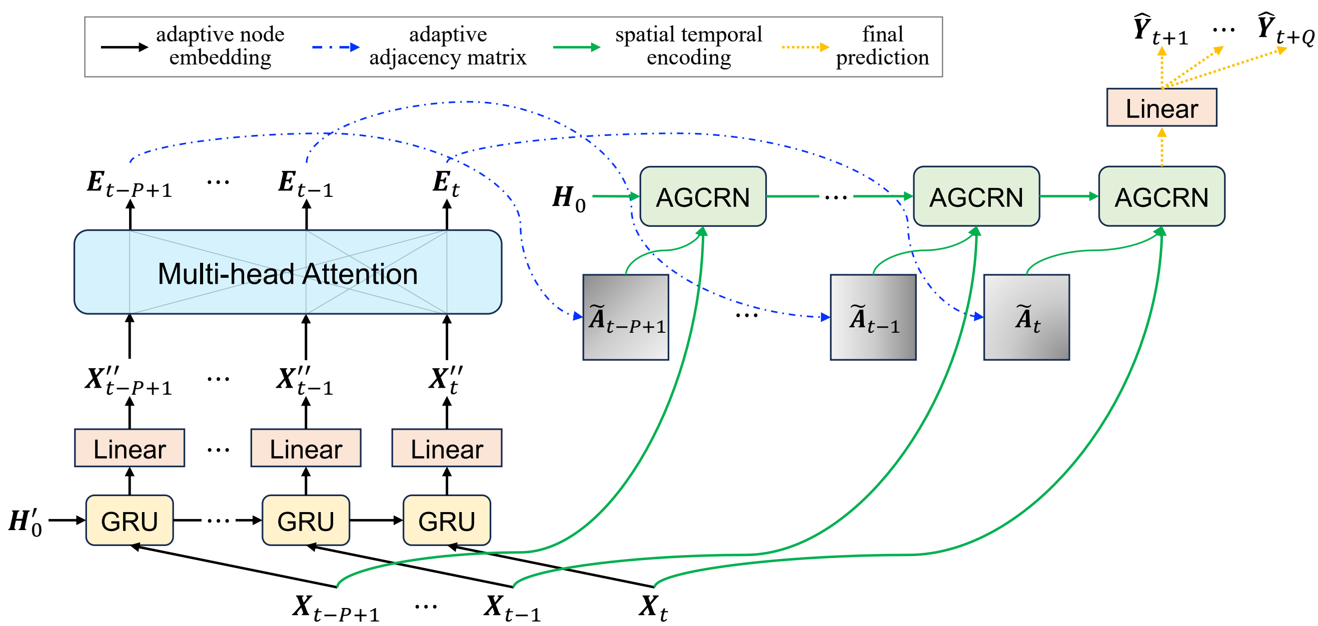
Graphical Abstract

Dynamic Spatial-Temporal Model for Carbon Emission Forecasting

Mingze Gong,Yongqi Zhang,Jia Li,Lei Chen



Highlights

Dynamic Spatial-Temporal Model for Carbon Emission Forecasting

Mingze Gong,Yongqi Zhang,Jia Li,Lei Chen

- We study carbon emission forecasting problem, which aims to inform energy planning.
- We formulate carbon emissions forecasting as a spatial-temporal prediction problem.
- We propose a novel model fusing Graph Convolutional and Recurrent Neural Network.
- Environmental factors' inclusion enhances accuracy, as verified on China dataset.
- Superior performance of the model is validated across China, US, and EU datasets.

Dynamic Spatial-Temporal Model for Carbon Emission Forecasting

Mingze Gong[a], Yongqi Zhang[b], Jia Li[a] and Lei Chen[a,*]

[a] *The Hong Kong University of Science and Technology (Guangzhou), Guangdong, China*
[b] *The Hong Kong University of Science and Technology, Hong Kong SAR, China*

ARTICLE INFO

Keywords:
Carbon emissions forecasting
Graph neural networks
Spatial-temporal prediction
Time-series prediction
Environmental predictors

ABSTRACT

Addressing the urgent need for accurate carbon emissions forecasting to support global emissions reduction goals, this paper introduces a novel approach for carbon emissions prediction that considers the spatial-temporal correlation of carbon emissions across forecasting targets. Traditional methods, primarily grounded in statistical and machine learning models, have shown certain limitations, such as a tendency towards oversimplification and inefficiency in capturing the evolving dynamic relationships between different forecasting targets. To address these, we propose an innovative Dynamic Spatial-Temporal Graph Convolutional Recurrent Network (DSTGCRN), a blend of graph convolutional network (GCN) and recurrent neural network (RNN) structures customized for spatial-temporal prediction. This approach accounts for nuanced correlations and interactions between regions, introducing the first method that formulates carbon emission prediction as a spatial-temporal time-series problem, thereby capturing dynamic correlations among regions over different time steps. The incorporation of environmental data, including temperature and the Air Quality Index (AQI), as supplementary predictors in the DSTGCRN model, has demonstrated its efficacy in the context of the China dataset. Empirical analyses conducted on datasets from China, the United States (US) and the European Union (EU) confirm the superior performance of the DSTGCRN over traditional and other spatial-temporal prediction models, highlighting the importance of the spatial-temporal frameworks for carbon emissions prediction and the model's robustness across diverse geographical contexts.

1. Introduction

Human-driven carbon emissions, primarily from energy production and fuel combustion, have surged in the 21st century, contributing significantly to global warming and its consequences such as extreme weather and sea-level rise [13, 10, 59, 29]. This escalation has been characterized by an unprecedented increase in the burning of fossil fuels, deforestation, industrial activities and land degradation [1], which releases a substantial amount of greenhouse gases into the atmosphere. The ramifications of this are manifold, affecting ecosystems, economies, and communities globally. The precision of carbon emissions forecasting has gained importance as it aids in facilitating low-carbon lifestyles, informing policy decisions, and shaping sustainable development strategies. This is underscored by several notable commitments globally: China has set a goal in 2014 to reach peak emissions by the year 2030 [13, 53]; the United States (US) has committed to reducing its greenhouse gas emissions to a level 50-52% lower than what was recorded in 2005, aiming to achieve this reduction by 2030 [40]; and the European Union (EU) plans to reduce its greenhouse gas emissions by at least 40% by 2030, using 1990 levels as a baseline [30].

The formulation of carbon emission prediction can be effectively framed as a time-series prediction problem, a frequently proposed strategy that exploits historical data to forecast future emissions by identifying trends and patterns that influence outcomes over time. Traditional methods employed for predicting carbon emissions have largely relied on statistical and machine learning models [20, 51]. Statistical models necessitate preliminary insights regarding the distribution of the data in order to construct predictive models [33]. They identify patterns in the data such as trend

*Corresponding author
mgong081@connect.hkust-gz.edu.cn (M. Gong); yzhangee@connect.ust.hk (Y. Zhang); leejia@ust.hk (J. Li); leichen@ust.hk (L. Chen)
ORCID(s):

and seasonality, and use these patterns to predict future data points. However, they have limitations due to their inherent simplicity. These models generate a single parameter from time series or panel data estimations, which may not accurately capture the long-term nonlinear relationships between carbon emissions and their influencing factors [15]. On the other hand, machine learning models do not make strong assumptions about the underlying process that generates the data [32]. These models leverage historical data to comprehend the probabilistic relationship between past and future events [7]. Notwithstanding their capabilities, the application of both statistical and machine learning methods has been mostly restricted to independent predictions for individual regions or industries, overlooking shared factors and crucial interdependencies between multiple regions or industries. The complexity of carbon emissions prediction is further compounded by two key challenges:

- The inter-region correlation is dynamic, not constant, hence changes in one region can significantly affect the carbon emissions of others.
- Emissions within individual regions exhibit complex temporal shifts, significantly influencing the overall emissions landscape.

In response to these limitations and challenges, we propose to model the correlation of carbon emission series of multiple regions as a dynamic spatial-temporal problem. Specifically, we propose the Dynamic Spatial-Temporal Graph Convolutional Recurrent Network (DSTGCRN), a specialized integration of Graph Convolutional Network (GCN) and Recurrent Neural Network (RNN) structures. This network is designed to account for the spatial-temporal interdependencies of carbon emissions across various regions, while also capturing the dynamic changes in these correlations over multiple time steps. DSTGCRN contains two main modules. The first draws on the Gated Recurrent Unit (GRU) [9] and multi-head attention mechanism to craft temporal representations for each region at every time step. The second module utilizes these temporal representations to facilitate the GCN and RNN in capturing dynamic spatial-temporal patterns throughout the carbon emission series, adapting to different time steps effectively. In this manner, the GCN is able to capture and interpret the intricate inter-regional relationships between different regions based on their dynamic characteristics and temporal correlations, and the RNN is to account for temporal dependencies in the data, integrating past carbon emission trends of each region, and tracing the historical trajectories and transitions over time. By encompassing dynamic interactions and evolving features, our model offers a thorough solution to the complex and uncertain task of carbon emission prediction in China. The contributions of our paper can be summarized as follows:

- Conceptually, we formulate the carbon emission prediction as a spatial-temporal time-series prediction problem. This helps us to take the interaction between regions into consideration.
- Technically, we propose a new model that can leverage the dynamic correlation among regions over different time steps to do the forecasting. In this way, the dependency of carbon emissions on natural factors like weather and seasons can be captured.
- Empirically, we demonstrate that the spatial-temporal setup is important for carbon emission prediction, the proposed model significantly outperforms the traditional time-series prediction model and other spatial-temporal prediction models.

- Strategically, we enable precise energy planning by accurately predicting daily carbon emissions. Validated with data from China, the US, and the EU, our model aids regions in pursuing sustainable energy based on dynamic data insights.

2. Related workds

2.1. Carbon emissions prediction

The prediction of carbon emissions is a critical aspect of environmental science and policy-making. It holds particular relevance in regions such as China, the US and the EU, where understanding the dynamics of carbon emissions is essential due to their significant roles in the global economy and environmental landscape.

2.1.1. Statistical methods

Many studies have focused on this area, employing a range of predictive models. Autoregressive Integrated Moving Average was utilized to investigate emissions trends [31]. Grey Forecasting Models (GM), which are specifically designed to handle limited data samples [28], have also received considerable attention. Ye et al. [52] enhanced the conventional GM by incorporating lag relationships, while Gao et al. [11] further addressed the challenges of scarcity and non-linearity in carbon emissions data using GMs. Statistical regression-based models are another category that has been frequently utilized. Xu and Liao [50] harnessed such a model to forecast carbon emissions using data gathered at varying intervals. Wu et al. [47] employed regression models to predict future carbon emissions based on a variety of factors. Despite their prevalent usage, these models commonly assume linearity and stationarity in variables, presumptions that may not align with the complexities inherent in carbon emissions. Additionally, they might grapple with the management of high-dimensional, non-linear data and interdependent relationships.

2.1.2. Machine learning methods

Traditional machine learning methods like Linear Regression and Random Forest have played a significant role in many forecasting endeavors, but they have presented limitations in the context of carbon emissions forecasting [3, 21]. In contrast, the Extreme Learning Machine (ELM) and Support Vector Machine (SVM) have revealed promise in addressing these challenges. For instance, ELM has been effectively combined with SVM to enhance the machine learning process [22], and when paired with Principle Component Analysis, it has been successful in reducing the dimensionality of influencing factors, leading to improved prediction performance [38]. Neural networks, a specialized subdivision of machine learning, have exhibited their capabilities in formulating intricate patterns and interconnections within large-scale data sets. These capacities have been harnessed in the realm of carbon emissions forecasting, allowing researchers to efficiently identify the complicated interplay among various determinants of emissions. For example, [35] employed a Fast Learning Network to forecast carbon emissions and it outperforms ELM in accurately projecting emissions trends. Other Artificial Neural Network based models [15, 58, 16, 24, 2, 13] were also developed to predict carbon emissions in specific regions. Although these methods are impressive compared to traditional statistical models, they frequently encounter limitations pertaining to the geographical scope of their predictions, often constrained by the number of regions they are capable of forecasting.

2.2. Spatial-temporal time-series prediction

In recent years, Spatial-Temporal Graph Neural Networks (STGNNs) have emerged as a powerful tool for spatial-temporal time series prediction [36]. They achieve this goal by learning time-dependent representations of the graph structure. Wu et al. [48] introduced the pioneering Multivariate Time Series Forecasting with Graph Neural Networks (MTGNN) model, utilizing Graph Neural Networks to analyze multivariate time series data from a graph-based spatio-temporal perspective. While many STGNNs have primarily focused on short-term prediction, Shao et al. [37] have developed an enhanced STGNN for multivariate time series forecasting. The reliance of many STGNNs [23, 44], on a predefined matrix, calls for an adaptive graph structure. To address this, Bai et al. [4] proposed the Adaptive Graph Convolutional Recurrent Network (AGCRN), Weng et al. [46] suggested a decomposition dynamic graph convolutional recurrent network and Wu et al. [49] introduced Graph WaveNet. Additionally, attention-based STGNNs, such as the Graph Multi-Attention Network (GMAN) proposed by Zheng et al. [57], have gained popularity due to their ability to alleviate the problem of error propagation. Guo et al. [12] utilized spatial attention to model complex spatial correlations between different locations and incorporated a spatial-temporal convolution module to capture dependencies within the data. Chen et al. [8] proposed an attention mechanism for sequence-to-sequence learning tasks, accentuating the importance of certain factors over others. In a similar vein, to fully embrace relevant factors, Tao et al. [39] advanced this domain with their Multiple Information Spatial-Temporal Attention-based Graph Convolution Networks (MISTAGCN) which is a novel attempt to assimilate various influential factors in the analysis. Nevertheless, it is noteworthy that the primary emphasis of these attention mechanisms has been predominantly on unraveling spatial relationships, with less consideration given to the extraction of temporal dependencies.

Building upon the foundations of prior studies, this work thus adds to the expanding body of research that explores the utilization of STGNNs in the domain of environmental science [34, 42, 43, 45, 54, 60]. It showcases the model's strength in forecasting carbon emissions, emphasizing its capability to represent the dynamic interactions among regions as well as the temporal variations within them.

3. Methodology

3.1. Definitions and problem setups

This section delineates the core concepts of the regional carbon emission network, environmental features, and the forecasting problem addressed in this study. It also describes the graph neural network and its spatial and temporal characteristics.

Definition 3.1 (Multisource Time Series)**.** A multisource time series is denoted as $\mathcal{Y} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_N\} \in \mathbb{R}^{N\times T}$, where $\boldsymbol{Y}_t = \{y_{1,t}, y_{2,t}, \ldots, y_{N,t}\} \in \mathbb{R}^N$'s for $t = 1 \ldots T$ are the values of the N sources (provinces/states/countries), at time step t.

Based on the definition of multisource time series, we define the problem of multisource time series forecasting as follows.

Definition 3.2 (Multisource Time Series Forecasting)**.** Denote $\boldsymbol{X}_t \in \mathbb{R}^{N\times C}$ as the observed values of all C features, including descriptors of temperature, Air Quality Index (AQI), season, month, etc., at time step t for the N regions. Note that the one-step ahead target values $\boldsymbol{Y}_t$ can also serve as a feature in $\boldsymbol{X}_t$. Then the multisource time series forecasting problem can be defined as modeling a forecasting function $f : \mathbb{R}^{N\times P\times C} \to \mathbb{R}^{N\times Q}$ that

$$\{\boldsymbol{Y}_{t+1}, \boldsymbol{Y}_{t+2}, \ldots, \boldsymbol{Y}_{t+Q}\} = f(\boldsymbol{X}_{t-P+1}, \ldots, \boldsymbol{X}_{t-1}, \boldsymbol{X}_t), \quad (1)$$

where P is the number of past time steps of historical data and Q is the number of future time steps to be forecasted.

In order to account for the correlation among multiple regions' carbon emission series, we define the regional carbon emission network as a graph in the following definition.

Definition 3.3 (Regional Carbon Emission Network)**.** A regional carbon emission network can be depicted as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{A})$, where $\mathcal{V}$ is a set of N regions. Each node in this network corresponds to a region. $\mathcal{E} = \{(i, j) : i, j \in \mathcal{V}\}$ is the set of edges modeling connections between regions, and $\boldsymbol{A}$ is the adjacency matrix where $A_{ij} = 1$ if $(i, j) \in \mathcal{E}$ otherwise $A_{ij} = 0$.

Modeling with a regional carbon emission network unveils several pivotal advantages. Firstly, by structuring emissions data in a network format, we are able to capture the intricate interdependencies between different regions, a detail that traditional models often overlook. Secondly, this networked approach grants a nuanced understanding of how emissions in one region can influence, or be influenced by, emissions in neighboring regions. Lastly, the regional carbon emission network provides a holistic, interconnected view of emissions dynamics. These advantages collectively pave the way for more informed policy-making and strategic planning.

3.2. Adaptive Graph Convolutional Recurrent Network (AGCRN) [4]

To model the spatial-temporal pattern of carbon emission, we adopt AGCRN [4] as the backbone model. Instead of using some pre-defined similarity or distance functions to model the correlation among N regions, AGCRN defines a learned node embedding matrix $\boldsymbol{E} \in \mathbb{R}^{N \times d_e}$ for N regions and each region has a specific embedding with dimension d_e. Then the data adaptive graph generation enhanced GCN at step t can be formulated as

$$\boldsymbol{Z}_t = \left(\boldsymbol{I}_N + \text{softmax}(\text{ReLU}(\boldsymbol{E} \cdot \boldsymbol{E}^\top))\right) \boldsymbol{X}_t \boldsymbol{\Theta}, \tag{2}$$

where $\boldsymbol{I}_N \in \mathbb{R}^{N \times N}$ is the identity matrix with diagonal elements being 1 and others being 0, $\cdot$ is the matrix multiplication operator, $\boldsymbol{\Theta} \in \mathbb{R}^{C \times d_e}$ is the weighting matrix. Intuitively, softmax(ReLU($\boldsymbol{E} \cdot \boldsymbol{E}^\top$) approximates the Laplacian matrix $\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ ($\boldsymbol{A}$ is the adjacency matrix and $\boldsymbol{D}$ is the degree matrix computed based on $\boldsymbol{A}$) in conventional GCN [19]. And with Eq.(2), the information of each region in $\boldsymbol{X}_t$ can be propagated to influence their neighbors based on the approximated adjacency matrix.

With the learned embedding matrix, AGCRN integrates the approximated adjacency matrix into a GRU module such that both the spatial and temporal correlations in the carbon emission series can be captured. Formally, AGCRN can be described with the following equations:

$$\begin{aligned}
\tilde{\boldsymbol{A}} &= \text{softmax}(\text{ReLU}(\boldsymbol{E} \cdot \boldsymbol{E}^T)) \\
\boldsymbol{R}_t &= \sigma\left(\tilde{\boldsymbol{A}}[\boldsymbol{X}_t, \boldsymbol{H}_{t-1}]\boldsymbol{E}\boldsymbol{W}_r + \boldsymbol{E} \cdot \boldsymbol{b}_r\right) \\
\boldsymbol{U}_t &= \sigma\left(\tilde{\boldsymbol{A}}[\boldsymbol{X}_t, \boldsymbol{H}_{t-1}]\boldsymbol{E}\boldsymbol{W}_u + \boldsymbol{E} \cdot \boldsymbol{b}_u\right) \\
\hat{\boldsymbol{H}}_t &= \tanh\left(\tilde{\boldsymbol{A}}[\boldsymbol{X}_t, \boldsymbol{U}_t \odot \boldsymbol{H}_{t-1}]\boldsymbol{E}\boldsymbol{W}_h + \boldsymbol{E} \cdot \boldsymbol{b}_h\right) \\
\boldsymbol{H}_t &= \boldsymbol{R}_t \odot \boldsymbol{H}_{t-1} + (1 - \boldsymbol{R}_t) \odot \hat{\boldsymbol{H}}_t
\end{aligned} \tag{3}$$

where $\sigma(\cdot)$ is the sigmoid function, $[\cdot]$ is the concatenation operator, $\odot$ is the element-wise dot product operator, $\boldsymbol{R}_t, \boldsymbol{U}_t \in \mathbb{R}^{N \times d_e}$ are the reset gate and update gate, respectively, $\boldsymbol{E} \in \mathbb{R}^{N \times d_e}$, $\boldsymbol{W}_r, \boldsymbol{W}_u, \boldsymbol{W}_h \in \mathbb{R}^{d_e \times (C + d_h) \times d_h}$ and $\boldsymbol{b}_r, \boldsymbol{b}_u, \boldsymbol{b}_h \in \mathbb{R}^{d_e \times d_h}$ are learnable parameters, $\boldsymbol{X}_t$ and $\boldsymbol{H}_t \in \mathbb{R}^{N \times d_h}$ are the input/output at the t-th time step. Similar to GRU, all the learnable parameters in Eq.(3) can be trained with backpropagation through time.

AGCRN offers a robust framework for the analysis of spatial-temporal patterns, thanks to its unique architecture.

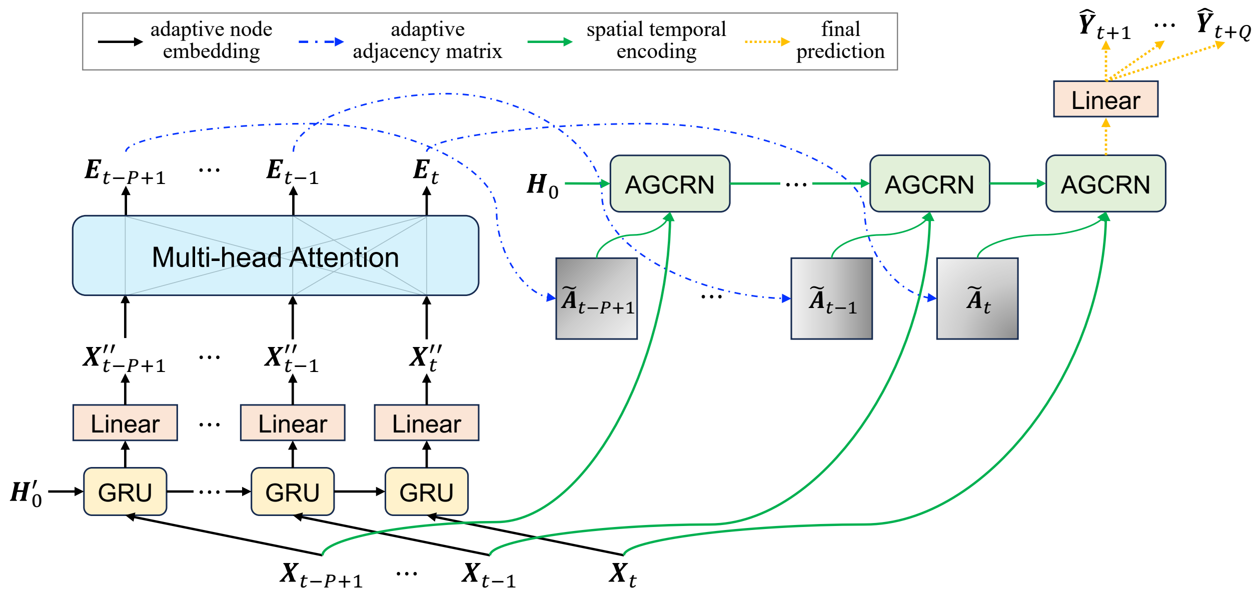


Figure 1: Computation graph of the proposed DSTGCRN method. In the black lines, the raw input is put into the GRU module and multi-head attention module to capture the temporal dependency within the historical data over the past P time steps. In the blue lines, the output embeddings for the N regions at the past P time steps are used to approximate the adaptive adjacency matrix $\tilde{A}_t'$s to model the dynamic correlation among regions. In the green line, the adaptive graph convolutional recurrent network (AGCRN) module takes input of raw inputs and the adaptive adjacency matrix to model the spatial-temporal patterns of the given data samples. Finally, in the red lines, we forecast the carbon emission values of the N regions in the next Q times steps based on the last step hidden state in the AGCRN module. (Color in print).

It incorporates a learned node embedding matrix $\boldsymbol{E}$, which is crucial in capturing the unique characteristics inherent to each node. This is a pivotal element in spatial pattern modeling, given the potential for each node to display unique attributes that influence the target variable. Moreover, this learned $\boldsymbol{E}$ is deployed to approximate spatial dependencies as presented in Eq.(2), thereby enabling the model to discern spatial correlations. It further integrates this approximated adjacency matrix into a GRU module which is capable of modeling the temporal evolution. Finally, all parameters within the AGCRN Eq.(3) can be optimized end-to-end, thereby enhancing the model's adaptability and flexibility. As a result, AGCRN is proficient in capturing both spatial and temporal correlations, rendering it a highly effective tool for the modeling of such patterns.

3.3. Construction of DSTGCRN

Even though AGCRN [4] is able to model the spatial and temporal correlation of multisource time series of the carbon emission in different regions, it still has limitations. Since the embedding matrix $\boldsymbol{E}$ is shared across different time steps, the approximated correlation is unchanged all over time. However in reality, the correlation of carbon emissions between different regions can change over time. For instance, the carbon emissions may be influenced by weather, temperature, weekly or seasonal factors, etc. Such a change over time will influence the carbon trading and emission across different regions. Hence, how to dynamically model the correlation among regions over different time steps is essential but very challenging.

We propose a novel method, as illustrated in Figure 1, which introduces modules that generate dynamic embeddings for each raw input, factoring in their respective features. These dynamic embeddings facilitate the creation of an adaptive adjacency matrix for each time step, capturing evolving inter-region correlations. This dynamic adjacency matrix and the raw inputs are then integrated by the AGCRN module to model spatial-temporal patterns. The output of this module undergoes a linear transformation to produce the final prediction, yielding a method that effectively models region correlations over variable time steps. Compared to our proposed framework, a naive solution would be to introduce a time-specific node embedding matrix at each time step, e.g., $\boldsymbol{E}_t \in \mathbb{R}^{N \times d_e}$. However, this will introduce about $T \times N \times d_e$ additional embedding parameters, potentially leading to an over-fitting issue. In addition, the embedding matrix $\boldsymbol{E}_t$ for the future time steps will not be available, leading to poor generalization performance. Hence instead, we introduce a dynamic node embedding module in this part. This module receives the raw features $\boldsymbol{\mathcal{X}}_t = \{\boldsymbol{X}_{t-P+1}, \ldots, \boldsymbol{X}_{t-1}, \boldsymbol{X}_t\} \in \mathbb{R}^{P \times N \times C}$ and outputs a set of dynamic embeddings $\boldsymbol{\mathcal{E}}_t \in \mathbb{R}^{P \times N \times d_e}$. Here, d_e represents the size of the vector space in which nodes are embedded. The node embeddings capture the essential characteristics of the nodes and are used to construct the adjacency matrix, which in turn is used to represent the graph structure to be used in GCN.

3.3.1. Sequential encoding with GRU

To capture the temporal dependency among the input features, such as the periodic changes in weather, temperature, etc., we first use GRU, a variant of RNN to adaptively learn from P historical time steps of $\boldsymbol{\mathcal{X}}_t = \{\boldsymbol{X}_{t-P+1}, \ldots, \boldsymbol{X}_{t-1}, \boldsymbol{X}_t\}$. Denote $\boldsymbol{H}'_{t-1} \in \mathbb{R}^{N \times d_e}$ as the hidden state of previous time step and d_e as the hidden dimension, then the update function of GRU is given as follows:

$$
\begin{aligned}
\boldsymbol{R}'_t &= \sigma\left([\boldsymbol{X}_t, \boldsymbol{H}'_{t-1}]\boldsymbol{W}'_r + \boldsymbol{b}'_r\right) \\
\boldsymbol{U}'_t &= \sigma\left([\boldsymbol{X}_t, \boldsymbol{H}'_{t-1}]\boldsymbol{W}'_u + \boldsymbol{b}'_u\right) \\
\hat{\boldsymbol{H}}'_t &= \tanh([\boldsymbol{X}_t, \boldsymbol{U}'_t \odot \boldsymbol{H}'_{t-1}]\boldsymbol{W}'_h + \boldsymbol{b}'_h)) \\
\boldsymbol{H}'_t &= \boldsymbol{R}'_t \odot \boldsymbol{H}'_{t-1} + (1 - \boldsymbol{R}'_t) \odot \hat{\boldsymbol{H}}'_t
\end{aligned}
\tag{4}
$$

Here, the weighting matrices $\boldsymbol{W}'_r, \boldsymbol{W}'_u, \boldsymbol{W}'_h \in \mathbb{R}^{(C+d_h) \times d_h}$ and bias vectors $\boldsymbol{b}'_r, \boldsymbol{b}'_u, \boldsymbol{b}'_h \in \mathbb{R}^{d_h}$ are learnable and are much smaller than the size $T \times N \times d_e$.

The GRU processes each sequence one-time step at a time. At each time step t, it accepts an input $\boldsymbol{X}_t \in \mathbb{R}^{N \times C}$, and it updates its hidden state $\mathbf{H}'_t$ based on this input and the previous hidden state $\mathbf{H}'_{t-1}$, thereby capturing temporal dependencies. The output $\mathbf{H}'_t$ at each time step is a new representation of the input at that time step which incorporates information from previous time steps. This design allows the model to generate dynamic node embeddings, where each node's embedding changes over time based on its historical data. The node embedding at each time step t is represented by the hidden state $\mathbf{H}'_t \in \mathbb{R}^{N \times d_h}$, where d_h is the hidden dimension.

Upon completion of the sequence processing on the input sequence $\boldsymbol{\mathcal{X}}_t$, the GRU produces a sequence of hidden states $\boldsymbol{H}'_{t-P+1}, \ldots, \boldsymbol{H}'_{t-1}, \boldsymbol{H}'_t$. These hidden states are then concatenated along the time dimension to form $\boldsymbol{\mathcal{X}}'_t = \{\boldsymbol{H}'_{t-P+1}, \ldots, \boldsymbol{H}'_{t-1}, \boldsymbol{H}'_t\} \in \mathbb{R}^{P \times N \times d_h}$, effectively capturing the temporal dependencies across the entire sequence.

The concatenated output $\boldsymbol{\mathcal{X}}'_t$ is subsequently passed through a linear layer and a Rectified Linear Unit (ReLU) activation function to introduce non-linearity. This transformation is defined as follows:

$$
\boldsymbol{\mathcal{X}}''_t = \mathrm{ReLU}(\boldsymbol{\mathcal{X}}'_t \boldsymbol{W}_h), \tag{5}
$$

where $\boldsymbol{W}_h \in \mathbb{R}^{d_h \times d_e}$ is a transformation matrix. The transformed output $\boldsymbol{\mathcal{X}}''_t$ belongs to $\mathbb{R}^{P \times N \times d_e}$.

3.3.2. Enhanced temporal communication with Multihead Attention

The multi-head attention mechanism is introduced to enhance temporal communication in the DSTGCRN due to its proven capability in capturing dependencies in sequence-based tasks, regardless of the distance between data points in the sequence [41]. Given the nature of carbon emissions data, which exhibit complex temporal shifts and interdependencies over time, it is crucial to identify and learn from these dependencies for accurate prediction.

Our model employs the output tensor of the preceding $\mathcal{X}_t''$ as the source for the queries $\mathbf{Q}$, keys $\mathbf{K}$, and values $\mathbf{V}$. Given this, the output of a single attention head is computed as:

$$\begin{aligned}&\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V})\\&=\text{softmax}\left(\frac{(\mathcal{X}_t''\boldsymbol{W}_Q)(\mathcal{X}_t''\boldsymbol{W}_K)^T}{\sqrt{s}}\right)\mathcal{X}_t''\boldsymbol{W}_V,\end{aligned}$$

where $\mathbf{Q}=\mathcal{X}_t''\boldsymbol{W}_Q$, $\mathbf{K}=\mathcal{X}_t''\boldsymbol{W}_K$, and $\mathbf{V}=\mathcal{X}_t''\boldsymbol{W}_V$, $\sqrt{s}$ is a scaling factor, usually defined as $\sqrt{d_e}$. Specifically, the computation above begins by determining attention scores that reflect the relationships between data points. These scores undergo an adjustment to maintain a uniform scale. Subsequently, the scores are converted into probabilities that gauge the significance of each data point. The resulting output accentuates the most pertinent data points, illuminating crucial patterns and tendencies within the sequence.

The Multihead Attention mechanism is composed of a set of $i = 1,...,H$ concurrent attention heads, each defined by Attention$(\mathbf{Q}_i,\mathbf{K}_i,\mathbf{V}_i)$. For each attention head i, we define learned linear transformations as $\boldsymbol{W}_{Q_i}$, $\boldsymbol{W}_{K_i}$, and $\boldsymbol{W}_{V_i}$ for the queries, keys, and values, respectively. The output for each head represented as head$_i$, is computed using these transformations on the input sequence. These outputs, head$_1,\dots,$head$_H$, each attend to different features in the input sequence. They are then concatenated and transformed by a learned weight matrix to produce the final output $\mathcal{E}_t = \{\boldsymbol{E}_{t-P+1},\dots,\boldsymbol{E}_{t-1},\boldsymbol{E}_t\}$, which belongs to the space $\mathbb{R}^{P\times N\times d_e}$ and serves as an enhanced temporal embedding for the input sequences $\mathcal{X}_t = \{\boldsymbol{X}_{t-P+1},\dots,\boldsymbol{X}_{t-1},\boldsymbol{X}_t\}$.

The Multihead Attention mechanism in our model captures various temporal patterns within sequences. For example, while one attention head might pinpoint short-term fluctuations in carbon emissions, another could highlight longer-term trends. These diverse perspectives, when combined, offer an integrated view of the temporal dynamics of the emissions data. Through the Multihead Attention, our model detects both subtle temporal variations within sequences and dominant patterns across regions. The resulting dynamic embeddings carry forward the enriched information about temporal dependencies and inter-region correlations to the next layer in the model.

3.3.3. *Dynamic Spatial-Temporal Graph Convolutional Recurrent Network (DSTGCRN)*

In this part, to fully account for the dynamic correlations between regions and the temporal dependencies within each region, we incorporate the dynamic embeddings into the AGCRN module to form a new model called dynamic spatial-temporal graph convolutional recurrent network (DSTGCRN). Rather than using a single shared embedding matrix $\boldsymbol{E}$, we introduce $\mathcal{E}_t = \{\boldsymbol{E}_{t-P+1},\dots,\boldsymbol{E}_{t-1},\boldsymbol{E}_t\}$ as the input embedding at different time steps.

Specifically, we replace $\boldsymbol{E}$ in Eq.(3) with $\boldsymbol{E}_t$. In this way, the approximated adjacency matrix will be changed to

$$\tilde{\boldsymbol{A}}_t = \text{softmax}\left(\text{ReLU}(\boldsymbol{E}_t\cdot\boldsymbol{E}_t^\top\right), \quad (6)$$

a dynamic matrix in $\mathbb{R}^{N\times N}$ changing over different time steps.

Such a dynamic adjacency matrix $\tilde{\boldsymbol{A}}_t$ offers substantial advantages in forecasting carbon emissions across different regions. As influencing factors of carbon emissions exhibit both spatial and temporal variations [55], the dynamic nature of $\tilde{\boldsymbol{A}}_t$ enables the model to accurately capture these temporal dynamics and spatial changes. This is particularly relevant given the significant variations in economic development levels, industrial structures, and energy consumption habits among different provinces.

Given the output $\boldsymbol{H}_t \in \mathbb{R}^{N \times d_h}$ capturing the spatial-temporal patterns in the historical data, we then directly obtain the carbon emission prediction for the next Q steps of all nodes with a linear transformation layer

$$\hat{\boldsymbol{Y}}_{t+i} = \boldsymbol{H}_t \cdot \boldsymbol{W}_{o,i} \tag{7}$$

where $\boldsymbol{W}_{o,i} \in \mathbb{R}^{d_h \times 1}$ for $i = 1 \ldots Q$ are specific weighting matrix for the i-th time step in figure. Simultaneously generating the next Q time steps can be more efficient than recurrently outputting future values. The overall workflow of $\hat{\boldsymbol{Y}}_t$'s computation is graphically shown in Figure 1 and clarified in Algorithm 1.

Finally, we use the L2 loss function as an objective and optimize all the learnable parameters in the GRU, Multihead Attention, and DSTGCRN modules. The loss function for each sample at time step t is formulated as

$$\mathcal{L} = \sum_{i=1}^{Q} (\boldsymbol{Y}_{t+i} - \hat{\boldsymbol{Y}}_{t+i})^2. \tag{8}$$

The minimization of Eq.(8) is solved by back-propagation optimizers with automatic gradient like Adam [18] in our implementation.

4. Experimental settings

4.1. Data sources and preprocessing

The datasets utilized in this study are sourced from Carbon Monitor (`https://carbonmonitor.org/`), and encompass carbon emissions data from China, the US and the EU. Detailed methodologies for the collection and processing of this data are outlined in [25, 26, 27]. The China dataset specializes in the period from January 1st, 2019, to December 31st, 2022. The US dataset covers the span from January 1st, 2019, to December 31st, 2021, while the EU dataset extends from January 1st, 2019, to April 30th, 2023. Each dataset contains critical variables that facilitate a deep understanding of the carbon emissions landscape in the respective regions. These variables include the specific locations (provinces/states/countries) where the data was collected, the date of emission recordings, the source sectors from which the emissions originated, and the corresponding emission values.

To facilitate time-series analysis, each of the aforementioned datasets has been individually restructured into a matrix format, representing daily carbon emissions over the respective time periods and geographical subdivisions.

Algorithm 1 DSTGCRN method computation

Require: Batch of $\boldsymbol{X}_{t-P+1}, \ldots, \boldsymbol{X}_t$.

1: **Initialize:**
- *GRU module* for temporal dependency.
- *Attention module* to focus on key time steps.
- *AGCRN module* for spatial-temporal patterns.

2: **Procedure:**
- Process input through GRU; apply linear layer and ReLU function.
- Use the Attention module on the output of the previous step over P time steps.
- Create output embeddings for N regions at P time steps.
- Approximate adaptive adjacency matrix with preceding output.
- Utilize AGCRN module with raw input and adjacency matrix.
- Forecast emissions using linear transformation.

3: **Output:** Predictions $\hat{\boldsymbol{Y}}_{t+1}, \ldots, \hat{\boldsymbol{Y}}_{t+Q}$.

Concurrently, we sourced temperature data from the National Oceanic and Atmospheric Administration (https://www.ncei.noaa.gov/) and AQI data from the China National Environmental Monitoring Centre (http://www.cnemc.cn/). We then calculated the daily average temperature and AQI at each monitoring site, associating these data points with the corresponding provinces in China. This initiative aims to ascertain whether incorporating these environmental variables can enhance the accuracy of carbon emissions forecasts in China, offering a richer perspective on the dynamics influencing carbon emissions in the region.

4.2. Baselines and Evaluation Metrics

We compare the DSTGCRN model against eight baseline methods, including traditional statistical models, contemporary machine learning models, and spatial-temporal models. The traditional statistical models employed in this study encompass the Vector Autoregressive (VAR) model [17]-a multivariate extension of the Autoregressive (AR) model-for the multivariate dataset of China, its simpler counterpart, the AR model [56], for the univariate datasets of the US and EU where only emissions data is considered, and the Support Vector Regression (SVR) model [5]. The modern machine learning models include Multilayer Perceptron (MLP) and Fully Connected Long Short Term Memory (FCLSTM) [14]. For the spatial-temporal models, we include AGCRN [4], GMAN [57], MISTAGCN [39] and MTGNN [48]. The traditional statistical and modern machine learning methods were chosen for their popularity and computational efficiency. However, they are known to have limitations in adequately capturing complex non-linear and spatio-temporal relationships. These spatial-temporal models are renowned for their performance in traffic forecasting. In this study, we adapted them to the problem of emissions forecasting, leveraging their ability to analyze spatial and temporal data. Specifically, we chose the MTGNN model, a pioneer in the field of graph spatial-temporal forecasting, and the MISTAGCN model, which is known for its ability to utilize multiple information sources in forecasting, such as weather conditions and time-of-day variables.

We adopt Rooted Mean Square Error (RMSE), Mean Average Error (MAE), Rooted Mean Square Percentage Error (RMSPE) and Mean Absolute Percentage Error (MAPE) metrics to evaluate the performance of different methods. They are calculated as below:

- MAE: $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$;
- MAPE: $\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|$;
- RMSE: $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$;
- RMSPE: $\sqrt{\frac{100\%}{n}\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{y_i}\right)^2}$;
- R^2: $1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum i=1^{n}(y_i-\bar{y})^2}$.

Here, for the i-th observation at time step t, y_i represents the actually observed emission, $\hat{y}_i$ represents the predicted emission, $\bar{y}$ represents the average of observed emission over all the time steps and n represents the total number of observations.

4.3. Hyper-parameter setups

The experiments are conducted on a Linux machine equipped with an NVIDIA GeForce RTX 3060 graphics card with 12GB of GPU memory. The basic parameters and settings of our experiments are summarized in Table 1. To optimize the model's performance, we utilized HyperOpt [6], a Python library for optimizing the hyperparameters of machine learning models through probabilistic optimization, conducting 80 trials to fine-tune the hyperparameters.

Table 1
Basic settings in experiments and values of hyperparameters for DSTGCRN.

Settings	Value
Lag	7d
Horizon	1d/3d
Loss function	L2Loss
Optimizer	Adam
Percentage of training data	80%
Percentage of validation data	10%
Percentage of test data	10%
Epochs	300
Early stop patience	60
Batch size	32
Learning rate	0.001
Hidden dimension	32
Embedding dimension	64
Runs	5

5. Results and discussions

In this section, we aim to test the effectiveness of DSTGCRN in predicting carbon emissions, compared to traditional statistical models, modern machine learning models, and other spatial-temporal models. We seek to validate the model's ability to capture complex spatial-temporal interdependencies and temporal shifts in the data and to assess its performance across broadened prediction horizons. In this regard, we propose the following research questions, to which the subsequent subsections will provide answers respectively:

- How does DSTGCRN's forecasting accuracy for carbon emissions in China compare to existing models, and how well does it adapt and perform in other geographical and environmental contexts?
- In what ways do feature engineering techniques and the inclusion of environmental predictors enhance the DSTGCRN model's forecasting accuracy?
- How do mutable relationships between provinces in China over distinct time intervals contribute to understanding the dynamic nature of carbon emissions?
- To what extent do the individual components of the DSTGCRN model influence its overall forecasting performance?

5.1. Overall comparison

Table 2 presents a comprehensive comparison of the forecasting performance of DSTGCRN with other baseline models across three different datasets: China, the US, and the EU. In the case of the dataset from China, traditional time series models, including the statistical model VAR and deep learning model FC-LSTM, while they have illustrated their aptitude in some applications, are found to be less capable of capturing the complex patterns and dependencies in the context of carbon emissions prediction. Surprisingly, some GCN and attention-based models like MISTAGCN and GMAN performed even worse than simpler traditional methods such as SVR and basic machine learning methods like MLP. This underperformance might be attributed to the inherent complexity of these models, which could introduce noise that obscures meaningful signals in the data, thereby reducing predictive accuracy. Conversely, AGCRN and MTGNN outperform other methods when evaluated using MAE and RMSE metrics, albeit MTGNN exhibits higher errors in percentage metrics, suggesting a potential shortfall in accurately representing proportional variations. Furthermore, these models tend to overlook the fluctuating correlations between provinces, s, which can be vital in accurately predicting carbon emissions.

Our proposed model, DSTGCRN, surpasses all baseline models in performance and robustness. Despite a marginal rise in standard deviation for extended horizons, its high reliability remains intact. Unlike other models, which display an escalation in errors with a longer horizon, DSTGCRN maintains its superior performance, a testament to its ability to effectively capture dynamic features and temporal factors.

Table 2
Overall comparison of DSTGCRN with all baselines after 5 runs for China, the US, and the EU datasets. The values represent the mean metrics. The best performance of each metric is indicated by boldface numbers. The STD row shows the standard deviation of DSTGCRN. The Improvement row shows the performance increase of DSTGCRN compared to the second-best performing model. Models denoted with an asterisk (*) utilized not only emissions data but also incorporated temperature and AQI information. All experiments are conducted with a lag of 7 days, and the prediction horizons are Q=1 and Q=3, representing 1 day and 3 days ahead prediction, respectively.

		Q=1					Q=3				
Dataset	Model	MAE	MAPE	RMSE	RMSPE	R^2	MAE	MAPE	RMSE	RMSPE	R^2
China	VAR*	0.113	12.803%	0.133	15.171%	-2.434	0.122	13.817%	0.144	16.398%	-3.035
	SVR	0.036	4.570%	0.044	5.530%	0.363	0.040	5.011%	0.048	6.077%	0.249
	MLP	0.045	5.795%	0.070	8.325%	0.715	0.045	5.847%	0.070	8.365%	0.708
	FC-LSTM	0.075	9.972%	0.112	13.362%	0.263	0.079	10.388%	0.116	13.987%	0.223
	AGCRN*	0.014	1.788%	0.025	2.889%	0.964	0.016	1.966%	0.027	3.165%	0.956
	GMAN*	0.069	9.755%	0.094	14.079%	0.211	0.071	9.843%	0.112	14.548%	0.204
	MISTAGCN*	0.062	8.275%	0.092	11.208%	-0.958	0.065	8.757%	0.095	11.959%	-1.186
	MTGNN*	0.016	4.557%	0.027	16.983%	-0.411	0.022	3.758%	0.038	8.140%	0.610
	DSTGCRN*	**0.012**	**1.539%**	**0.021**	**2.407%**	**0.975**	**0.009**	**1.186%**	**0.016**	**1.931%**	**0.983**
	Improvements	16.1%	14.0%	16.7%	16.7%	1.2%	40.6%	39.6%	40.3%	39.0%	2.8%
	STD	0.001	0.001	0.001	0.002	0.004	0.002	0.002	0.003	0.003	0.005
US	AR	0.038	16.021%	0.045	18.895%	-3.123	0.039	15.900%	0.046	18.674%	-3.251
	SVR	0.011	5.222%	0.013	6.215%	0.313	0.012	5.757%	0.015	6.821%	0.176
	MLP	0.018	7.021%	0.030	9.413%	0.535	0.017	6.994%	0.030	9.371%	0.529
	FC-LSTM	0.023	9.500%	0.040	12.167%	0.312	0.024	9.896%	0.040	12.898%	0.297
	AGCRN	0.003	1.104%	0.005	1.485%	0.989	0.004	1.584%	0.008	2.192%	0.975
	GMAN	0.020	10.097%	0.028	14.561%	0.273	0.021	10.460%	0.039	15.764%	0.247
	MISTAGCN	0.018	8.739%	0.029	12.407%	-0.173	0.018	8.924%	0.028	12.558%	-0.141
	MTGNN	0.004	3.618%	0.008	9.834%	0.741	0.007	4.835%	0.015	12.422%	0.688
	DSTGCRN	**0.003**	**1.071%**	**0.005**	**1.415%**	**0.991**	**0.004**	**1.480%**	**0.006**	**1.975%**	**0.981**
	Improvements	7.8%	3.1%	13.4%	4.9%	0.2%	12.5%	7.0%	24.6%	11.0%	0.6%
	STD	0.000	0.001	0.000	0.001	0.001	0.001	0.003	0.001	0.003	0.006
EU	AR	0.049	23.707%	0.061	31.629%	-0.725	0.051	23.919%	0.064	32.013%	-0.823
	SVR	0.021	5.891%	0.028	7.397%	0.743	0.025	6.959%	0.032	8.670%	0.662
	MLP	0.032	11.114%	0.069	15.930%	0.453	0.032	11.037%	0.069	15.808%	0.458
	FC-LSTM	0.031	11.997%	0.064	17.081%	0.426	0.032	12.266%	0.068	17.178%	0.383
	AGCRN	0.003	1.081%	0.006	1.513%	0.994	0.003	1.214%	0.007	1.742%	0.992
	GMAN	0.036	12.968%	0.046	18.896%	0.186	0.033	12.293%	0.076	19.504%	0.261
	MISTAGCN	0.027	9.060%	0.058	13.489%	0.469	0.042	11.529%	0.092	15.525%	0.050
	MTGNN	0.007	3.711%	0.018	7.159%	0.765	0.011	8.294%	0.023	17.186%	-1.087
	DSTGCRN	**0.002**	**0.780%**	**0.005**	**1.114%**	**0.997**	**0.003**	**1.009%**	**0.006**	**1.496%**	**0.995**
	Improvements	38.6%	38.5%	38.2%	35.8%	0.3%	21.7%	20.3%	20.8%	16.4%	0.3%
	STD	0.000	0.001	0.001	0.002	0.001	0.000	0.001	0.001	0.002	0.001

To more comprehensively demonstrate the superiority of DSTGCRN, we carried out additional experiments on the US and EU datasets. This comparative analysis, as part of the comprehensive comparison in Table 2, serves to highlight the distinctive advantages of DSTGCRN in different geographical and environmental contexts, thereby illustrating its robustness in forecasting carbon emissions.

For the US dataset, DSTGCRN exhibited a superior performance across all metrics, outshining AGCRN and other models in the benchmark. The improvements in RMSE and R^2 values are particularly noteworthy, showcasing the model's adeptness in capturing the underlying patterns in the data with high precision and reliability. Despite a higher STD for longer horizons, the MAPE and RMSPE remain lower compared to our backbone model, indicating a consistent and robust predictive accuracy even in scenarios with increased variability.

In the analysis of the EU dataset, DSTGCRN's dominance becomes more pronounced. It registered substantial improvements in key metrics such as MAE, MAPE, and RMSE, while achieving near-perfect R^2 values, indicative of an excellent fit to the data. The low STD values corroborate the model's robustness, underscoring its reliability in forecasting carbon emissions in the EU context.

The results unequivocally showcase the superior performance of DSTGCRN compared to the baseline models, presenting consistent enhancements across various metrics and forecasting scenarios in diverse contexts. Strikingly, DSTGCRN exhibits marked improvement over AGCRN, our foundational model, especially in the datasets for China and the EU. This underscores the efficiency of the innovative methods integrated into DSTGCRN, such as the development of dynamic embeddings and the establishment of an adaptive adjacency matrix at each time step, enabling DSTGCRN to identify evolving inter-region correlations. The disparate percentage gains over AGCRN at different horizons in the EU and China datasets suggest a individual adaptability of DSTGCRN to the individual characteristics of each dataset. In particular, the notable improvements in the EU dataset at horizon = 1 and in the China dataset at horizon = 3 suggest that DSTGCRN might be uniquely attuned to diverse features of these datasets.

5.2. Feature engineering and factor analysis

We performed a sensitivity analysis, depicted in Figure 3, utilizing the DSTGCRN model on the China dataset, as detailed in Table 2, to evaluate the influence of different

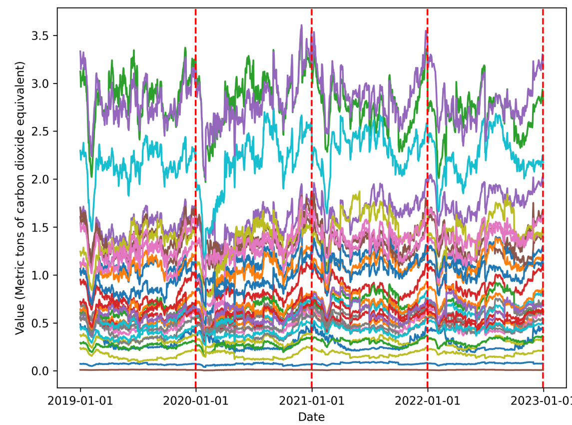


Figure 2: Provincial Carbon Emissions Trends from 2019 to 2022. Each color indicates the emission levels of individual provinces in China. Vertical dashed lines mark the commencement of each year. (Color in print).

feature sets on forecasting accuracy. This enabled us to thoroughly examine how these elements contribute to predicting emissions across various provinces.

We considered three scenarios: (1) Emissions-Only Model, (2) Emissions and Trend Model, and (3) Full Model. During the initial phase of the analysis, feature engineering techniques were employed. We introduced the day of the year and the season as additional features to capture the temporal patterns within the emissions data. This allowed us to account for fluctuations in emissions that might be influenced by the time of year as observed in Figure 2 where we observed a noticeable trend in the emissions for each province in China, as illustrated in Figure 2. By decomposing the emission data into three components (residual, trend, and seasonal), we were able to separate the long-term emission trend from the short-term seasonal variations. This feature engineering step allowed us to gain a better understanding of the underlying patterns in the emissions data. In the subsequent phase, we also included the average temperature and the average AQI for each province as additional predictors for our prediction model. This allowed us to examine the impact of these factors on emissions and further enhance the prediction accuracy.

The sensitivity analysis provides compelling evidence for the beneficial impact of incorporating additional trend-based and environmental features. Notably, the introduction of trend-related features led to a significant improvement in the model's predictive performance. This finding highlights the importance of capturing temporal patterns within emissions data.

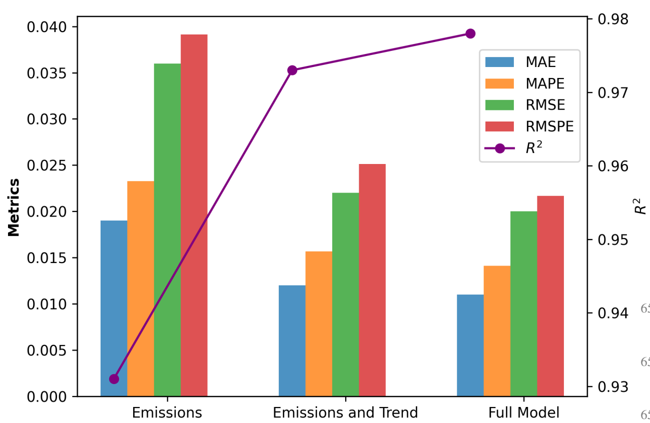


Figure 3: Comparison of three models. (1) Emissions-Only Model: uses only emissions data for predictions. (2) Emissions and Trend Model: incorporates emissions data and additional trend-based features such as the day of the year, the season, and the decomposition of emissions into residual, trend, and seasonal components. (3) Full Model: includes all features from the previous models and adds average temperature and average AQI as additional predictors. (Color in print).

5.3. Temporal dynamics of the adjacency matrix

In this section, our aim is to delve into the evolving inter-provincial relationships concerning carbon emissions as captured in Figure 4. These changes in the adjacency matrix signify alterations in the carbon emissions-related relationships among provinces in China.

The color gradient in the plot provides a visual measure of the magnitude of changes in the adjacency matrix between consecutive time steps. The middle of the plot, particularly between timestep 3 and 4, presents the deepest color. This suggests a significant change in the adjacency matrix in this interval, meaning that the relationships between provinces regarding carbon emissions experienced considerable alterations in this period.

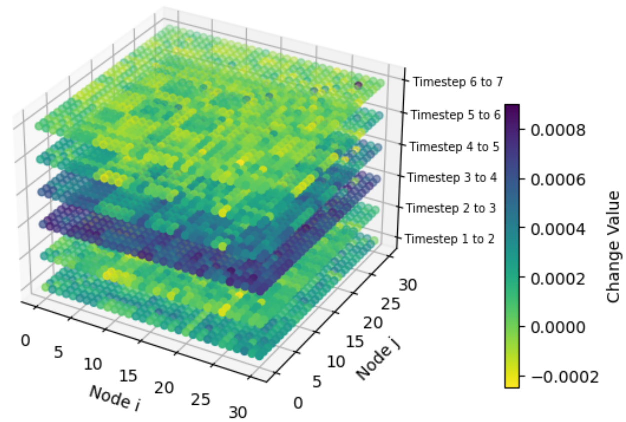


Figure 4: Differential Temporal Adjacency Matrix Evolution. This 3D visualization represents the average temporal evolution of the adjacency matrix for China's test dataset over a period of 7 days. Each point in the 3D space corresponds to the difference in the adjacency matrix between two consecutive days, illustrating the minor changes in relationships among nodes over time. The plot spans six unique spaces, revealing intricate patterns of change within our network's structure. (Color in print).

On the contrary, the lighter color at the two ends of the plot indicates that the changes in the adjacency matrix are relatively stable at the beginning and end of the week period. However, it is noteworthy that the color at the lower end is deeper than at the upper end. This subtle difference in color depth implies that the changes in the adjacency matrix, and hence the dynamics of the carbon emissions inter-provincial relationships, were slightly more pronounced at the start of the period under observation.

This differential temporal evolution of the adjacency matrix provides vital insights into the changes in the carbon emissions patterns among the provinces. The changes in the matrix reflect the proficiency of our model in capturing the temporal dynamics of carbon emissions across different

Table 3
Ablation experiments on DSTGCRN.

Model	MAE	MAPE	RMSE	RMSPE	R^2
Static	0.014	1.791%	0.025	2.881%	0.964
Time Sepcific	0.016	2.017%	0.029	3.275%	0.950
w/o GRU	0.019	2.559%	0.033	4.149%	0.930
w/o Attention	0.012	1.491%	0.021	2.467%	0.974
w/o GRU/ATT	0.017	2.230%	0.030	3.508%	0.947
DSTGCRN	**0.011**	**1.410%**	**0.020**	**2.169%**	0.978

provinces. This subsequently supports accurate prediction of future carbon emissions based on the observed patterns of change.

5.4. Ablation study

To evaluate the impact of each component in our proposed model, DSTGCRN, we conducted an ablation study to dissect the individual contributions of its core components in Table 3. We examine four variants of the model: (1) Static, which replaces the dynamic embedding with a static one; (2) Time Sepcific, the naive solution mentioned in Section 3.3 where a time-specific node embedding matrix $\boldsymbol{E}_t \in \mathbb{R}^{N\times d_e}$ is introduced; (3) w/o GRU, where the GRU layer is removed; (4) w/o Attention, where the attention layer is removed; and (5) w/o GRU/ATT, which excludes both the GRU and attention layers.

Static led to a performance decrease, emphasizing the dynamic embedding's role in capturing evolving node features. Time Specific exhibited a performance drop, illustrating the issue of overfitting. w/o GRU saw a significant performance degradation, highlighting the GRU's role in temporal dependency capture. w/o Attention showed a slight performance dip, indicating the attention layer's contributory, yet less significant, role compared to the dynamic embedding and GRU layer. The w/o GRU/ATT variant demonstrated a substantial performance drop. This underlines these layers' crucial roles and indicates that their collective contribution is greater than the sum of their individual effects.

In contrast, the complete DSTGCRN model, with all components integrated, outperformed all variants across metrics, testifying to the synergistic effect of the components.

6. Conclusions and Future Work

In this research, we introduced DSTGCRN, a spatial-temporal model designed to predict carbon emissions leveraging dynamic adjacency graph convolution and recurrent networks. Tested rigorously across datasets from China, the US, and the EU, DSTGCRN has demonstrated its applicability in diverse geographical contexts, excelling over traditional statistical and machine learning models across all evaluation metrics. This superior performance is attributed to its ability to model dynamic correlations among different regions and track carbon emission dependencies within each region over time. The incorporation of environmental variables such as temperature and AQI data has substantiated its efficacy in enhancing the precision of carbon emissions forecasting in China. Distinguished by its skill in identifying subtle changes in inter-provincial relationships in China over time series, DSTGCRN utilizes the differential temporal evolution of the adjacency matrix to reveal intricate patterns of transformation within the network's structure. The ablation study has been instrumental in evaluating the individual contributions of the core components of DSTGCRN, offering a deeper understanding of the model's strengths and pinpointing avenues for further optimization. The proposed DSTGCRN marks a significant leap in carbon emission prediction, offering potential for further research and practical applications. Looking forward, we envisage the inclusion of additional influential features to enhance the predictive accuracy of the model. Moreover, there is a substantial scope to broaden the geographical reach of this model, thereby augmenting its utility and relevance in the global context. Furthermore, future iterations of this research could delve into the quantification of uncertainties inherent in carbon emission predictions, fostering a more robust and reliable predictive framework. The integration of DSTGCRN with decision-making tools also emerges as a promising frontier, potentially facilitating more informed and effective strategies in the mitigation of carbon emissions globally.

CRediT authorship contribution statement

Mingze Gong: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing-Original draft preparation, Writing- Reviewing and Editing. **Yongqi Zhang:** Conceptualization, Methodology, Visualization, Writing-Original draft preparation, Writing-Reviewing and Editing. **Jia Li:** Resources, Data curation, Investigation, Writing-Reviewing and Editing. **Lei Chen:** Conceptualization, Project administration, Supervision, Writing-Reviewing and Editing.

References

[1] Abdul Latif, S.N., Chiong, M.S., Rajoo, S., Takada, A., Chun, Y.Y., Tahara, K., Ikegami, Y., 2021. The Trend and Status of Energy Resources and Greenhouse Gas Emissions in the Malaysia Power Generation Mix. Energies 14, 2200. doi:10.3390/en14082200.

[2] Aksu, İ.Ö., Demirdelen, T., 2022. The New Prediction Methodology for CO2 Emission to Ensure Energy Sustainability with the Hybrid Artificial Neural Network Approach. Sustainability 14, 15595. doi:10.3390/su142315595.

[3] Akyol, M., Uçar, E., 2021. Carbon footprint forecasting using time series data mining methods: The case of Turkey. Environmental Science and Pollution Research 28, 38552–38562. doi:10.1007/s11356-021-13431-6.

[4] Bai, L., Yao, L., Li, C., Wang, X., Wang, C., 2020. Adaptive graph convolutional recurrent network for traffic forecasting. Advances in neural information processing systems 33, 17804–17815.

[5] Basak, D., Pal, S., Patranabis, D., 2007. Support Vector Regression. Neural Information Processing – Letters and Reviews 11.

[6] Bergstra, J., Yamins, D., Cox, D., 2013. Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms, in: Python in Science Conference, Austin, Texas. pp. 13–19. doi:10.25080/Majora-8b375195-003.

[7] Bontempi, G., Ben Taieb, S., Le Borgne, Y.A., 2013. Machine Learning Strategies for Time Series Forecasting, in: Lecture Notes in Business Information Processing. volume 138. doi:10.1007/978-3-642-36318-4_3.

[8] Chen, Z., Wu, H., O'Connor, N.E., Liu, M., 2021. A Comparative Study of Using Spatial-Temporal Graph Convolutional Networks for Predicting Availability in Bike Sharing Schemes, in: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), pp. 1299–1305. doi:10.1109/ITSC48978.2021.9564831.

[9] Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. doi:10.48550/arXiv.1412.3555, arXiv:1412.3555.

[10] DeConto, R.M., Pollard, D., Alley, R.B., Velicogna, I., Gasson, E., Gomez, N., Sadai, S., Condron, A., Gilford, D.M., Ashe, E.L., Kopp, R.E., Li, D., Dutton, A., 2021. The Paris Climate Agreement and future sea-level rise from Antarctica. Nature 593, 83–89. doi:10.1038/s41586-021-03427-0.

[11] Gao, M., Yang, H., Xiao, Q., Goh, M., 2021. A novel fractional grey Riccati model for carbon emission prediction. Journal of Cleaner Production 282, 124471. doi:10.1016/j.jclepro.2020.124471.

[12] Guo, S., Lin, Y., Feng, N., Song, C., Wan, H., 2019. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence 33, 922–929. doi:10.1609/aaai.v33i01.3301922.

[13] Heydari, A., Garcia, D.A., Keynia, F., Bisegna, F., Santoli, L.D., 2019. Renewable Energies Generation and Carbon Dioxide Emission Forecasting in Microgrids and National Grids using GRNN-GWO Methodology. Energy Procedia 159, 154–159. doi:10.1016/j.egypro.2018.12.044.

[14] Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735.

[15] Jena, P.R., Managi, S., Majhi, B., 2021. Forecasting the CO2 Emissions at the Global Level: A Multilayer Artificial Neural Network Modelling. Energies 14, 6336. doi:10.3390/en14196336.

[16] Jin, H., 2021. Prediction of direct carbon emissions of Chinese provinces using artificial neural networks. PLOS ONE 16, e0236685. doi:10.1371/journal.pone.0236685.

[17] Kim, J.H., An, C.H., 2021. A Study on Estimation and Prediction of Vector Time Series Model Using Financial Big Data (Interest Rates). Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, 309–316. doi:10.17762/turcomat.v12i5.951.

[18] Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

[19] Kipf, T.N., Welling, M., 2017. Semi-Supervised Classification with Graph Convolutional Networks. URL: http://arxiv.org/abs/1609.02907, arXiv:1609.02907.

[20] Kumari, S., Singh, S.K., 2022a. Machine learning-based time series models for effective CO2 emission prediction in India. Environmental Science and Pollution Research doi:10.1007/s11356-022-21723-8.

[21] Kumari, S., Singh, S.K., 2022b. Machine learning-based time series models for effective CO2 emission prediction in India. Environmental Science and Pollution Research doi:10.1007/s11356-022-21723-8.

[22] Li, M., Wang, W., De, G., Ji, X., Tan, Z., 2018a. Forecasting Carbon Emissions Related to Energy Consumption in Beijing-Tianjin-Hebei Region Based on Grey Prediction Theory and Extreme Learning Machine Optimized by Support Vector Machine Algorithm. Energies 11, 2475. doi:10.3390/en11092475.

[23] Li, Y., Yu, R., Shahabi, C., Liu, Y., 2018b. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. URL: http://arxiv.org/abs/1707.01926, arXiv:1707.01926.

[24] Liu, J.p., Zhang, X.b., Song, X.h., 2018. Regional carbon emission evolution mechanism and its prediction approach driven by carbon trading – A case study of Beijing. Journal of Cleaner Production 172, 2793–2810. doi:10.1016/j.jclepro.2017.11.133.

[25] Liu, Z., Ciais, P., Deng, Z., Davis, S.J., Zheng, B., Wang, Y., Cui, D., Zhu, B., Dou, X., Ke, P., Sun, T., Guo, R., Zhong, H., Boucher, O., Bréon, F.M., Lu, C., Guo, R., Xue, J., Boucher, E., Tanaka, K., Chevallier, F., 2020a. Carbon Monitor, a near-real-time daily dataset of global CO2 emission from fossil fuel and cement production. Scientific Data 7, 392. doi:10.1038/s41597-020-00708-7.

[26] Liu, Z., Ciais, P., Deng, Z., Lei, R., Davis, S.J., Feng, S., Zheng, B., Cui, D., Dou, X., Zhu, B., Guo, R., Ke, P., Sun, T., Lu, C., He, P., Wang, Y., Yue, X., Wang, Y., Lei, Y., Zhou, H., Cai, Z., Wu, Y., Guo, R., Han, T., Xue, J., Boucher, O., Boucher, E., Chevallier, F., Tanaka, K., Wei, Y., Zhong, H., Kang, C., Zhang, N., Chen, B., Xi, F., Liu, M., Bréon, F.M., Lu, Y., Zhang, Q., Guan, D., Gong, P., Kammen, D.M., He, K., Schellnhuber, H.J., 2020b. Near-real-time monitoring of global CO2 emissions reveals the effects of the COVID-19 pandemic. Nature Communications 11, 5172. doi:10.1038/s41467-020-18922-7.

[27] Liu, Z., Deng, Z., Zhu, B., Ciais, P., Davis, S.J., Tan, J., Andrew, R.M., Boucher, O., Arous, S.B., Canadell, J.G., Dou, X., Friedlingstein, P., Gentine, P., Guo, R., Hong, C., Jackson, R.B., Kammen, D.M., Ke, P., Le Quéré, C., Monica, C., Janssens-Maenhout, G., Peters, G.P., Tanaka, K., Wang, Y., Zheng, B., Zhong, H., Sun, T., Schellnhuber, H.J., 2022. Global patterns of daily CO2 emissions reductions in the first year of COVID-19. Nature Geoscience 15, 615–620. doi:10.1038/s41561-022-00965-8.

[28] Ma, X., 2019. A brief introduction to the Grey Machine Learning. doi:10.48550/arXiv.1805.01745, arXiv:1805.01745.

[29] Mansouri, K., Alti, A., Roose, P., Laborie, S., 2018. Dynamic semantic-based green bio-inspired approach for optimizing energy and cloud services qualities: Dynamic semantic-based green bio-inspired approach for optimizing energy and cloud services qualities. Transactions on Emerging Telecommunications Technologies 29, e3305. doi:10.1002/ett.3305.

[30] Mentes, M., 2023. Sustainable development economy and the development of green economy in the European Union. Energy, Sustainability and Society 13, 32. doi:10.1186/s13705-023-00410-7.

[31] Ning, L., Pei, L., Li, F., 2021. Forecast of China's Carbon Emissions Based on ARIMA Method. Discrete Dynamics in Nature and Society 2021, 1–12. doi:10.1155/2021/1441942.

[32] Panch, T., Szolovits, P., Atun, R., . Artificial intelligence, machine learning and health systems. Journal of Global Health 8, 020303. doi:10.7189/jogh.08.020303.

[33] Parmezan, A.R.S., Souza, V.M., Batista, G.E., 2019. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. Information Sciences 484, 302–337. doi:10.1016/j.ins.2019.01.076.

[34] Qi, Y., Li, Q., Karimian, H., Liu, D., . A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory 664, 1–10. doi:10.1016/j.scitotenv.2019.01.333.

[35] Ren, F., Long, D., 2021. Carbon emission forecasting and scenario analysis in Guangdong Province based on optimized Fast Learning Network. Journal of Cleaner Production 317, 128408. doi:10.1016/j.

jclepro.2021.128408.

[36] Sahili, Z.A., Awad, M., 2023. Spatio-Temporal Graph Neural Networks: A Survey. URL: http://arxiv.org/abs/2301.10569, arXiv:2301.10569.

[37] Shao, Z., Zhang, Z., Wang, F., Xu, Y., 2022. Pre-training Enhanced Spatial-temporal Graph Neural Network for Multivariate Time Series Forecasting, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. pp. 1567–1577. doi:10.1145/3534678.3539396.

[38] Sun, W., Sun, J., 2017. Prediction of carbon dioxide emissions based on principal component analysis with regularized extreme learning machine: The case of China. Environmental Engineering Research 22, 302–311. doi:10.4491/eer.2016.153.

[39] Tao, S., Zhang, H., Yang, F., Wu, Y., Li, C., 2023. Multiple Information Spatial–Temporal Attention based Graph Convolution Network for traffic prediction. Applied Soft Computing 136, 110052. doi:10.1016/j.asoc.2023.110052.

[40] The White House, . National climate task force. The White House. URL: https://www.whitehouse.gov/climate/. accessed: 2023-09-18.

[41] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

[42] Wan, J., Zhang, H., Zhang, Q., Li, M.Z., Xu, Y., . Deep learning framework for forecasting en route airspace emissions considering temporal-spatial correlation 905, 166986. doi:10.1016/j.scitotenv.2023.166986.

[43] Wang, H., Zhang, L., Zhao, H., Wu, R., Sun, X., Cen, Y., Zhang, L., . Feature multi-level attention spatio-temporal graph residual network: A novel approach to ammonia nitrogen concentration prediction in water bodies by integrating external influences and spatio-temporal correlations 906, 167591. doi:10.1016/j.scitotenv.2023.167591.

[44] Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., Jia, C., Yu, J., 2020. Traffic Flow Prediction via Spatial Temporal Graph Neural Network, in: Proceedings of The Web Conference 2020, Association for Computing Machinery, New York, NY, USA. pp. 1082–1092. doi:10.1145/3366423.3380186.

[45] Wen, C., Liu, S., Yao, X., Peng, L., Li, X., Hu, Y., Chi, T., . A novel spatiotemporal convolutional long short-term neural network for air pollution prediction 654, 1091–1099. doi:10.1016/j.scitotenv.2018.11.086.

[46] Weng, W., Fan, J., Wu, H., Hu, Y., Tian, H., Zhu, F., Wu, J., 2023. A Decomposition Dynamic graph convolutional recurrent network for traffic forecasting. Pattern Recognition 142, 109670. doi:10.1016/j.patcog.2023.109670.

[47] Wu, C.B., Huang, G.H., Xin, B.G., Chen, J.K., 2018. Scenario analysis of carbon emissions' anti-driving effect on Qingdao's energy structure adjustment with an optimization model, Part i: Carbon emissions peak value prediction. Journal of Cleaner Production 172, 466–474. doi:10.1016/j.jclepro.2017.10.216.

[48] Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., Zhang, C., 2020. Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. URL: http://arxiv.org/abs/2005.11650, arXiv:2005.11650.

[49] Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C., 2019. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. URL: http://arxiv.org/abs/1906.00121, arXiv:1906.00121.

[50] Xu, X., Liao, M., 2022. Prediction of Carbon Emissions in China's Power Industry Based on the Mixed-Data Sampling (MIDAS) Regression Model. Atmosphere 13, 423. doi:10.3390/atmos13030423.

[51] Ye, L., Yang, D., Dang, Y., Wang, J., 2022a. An enhanced multivariable dynamic time-delay discrete grey forecasting model for predicting China's carbon emissions. Energy 249, 123681. doi:10.1016/j.energy.2022.123681.

[52] Ye, L., Yang, D., Dang, Y., Wang, J., 2022b. An enhanced multivariable dynamic time-delay discrete grey forecasting model for predicting China's carbon emissions. Energy 249, 123681. doi:10.1016/j.energy.2022.123681.

[53] Yu, H., Yang, Y., Li, B., Liu, B., Guo, Y., Wang, Y., Guo, Z., Meng, R., 2023. Research on the community electric carbon emission prediction considering the dynamic emission coefficient of power system. Scientific Reports 13, 5568. doi:10.1038/s41598-023-31022-y.

[54] Yu, M., Masrur, A., Blaszczak-Boxe, C., . Predicting hourly PM2.5 concentrations in wildfire-prone areas using a SpatioTemporal Transformer model 860, 160446. doi:10.1016/j.scitotenv.2022.160446.

[55] Yu, Z., Zhang, Y., Zhang, J., Zhang, W., 2022. Analysis and prediction of the temporal and spatial evolution of carbon emissions in China's eight economic regions. PLOS ONE 17, e0277906. doi:10.1371/journal.pone.0277906.

[56] Zhao, Y., Shen, L., 2011. Application of time series auto regressive model in price forecast, in: 2011 International Conference on Business Management and Electronic Information, pp. 768–771. doi:10.1109/

ICBMEI.2011.5921078.

[57] Zheng, C., Fan, X., Wang, C., Qi, J., 2020. GMAN: A Graph Multi-Attention Network for Traffic Prediction. Proceedings of the AAAI Conference on Artificial Intelligence 34, 1234–1241. doi:10.1609/aaai.v34i01.5477.

[58] Zhou, J., Xu, X., Li, W., Guang, F., Yu, X., Jin, B., 2019. Forecasting CO2 Emissions in China's Construction Industry Based on the Weighted Adaboost-ENN Model and Scenario Analysis. Journal of Energy 2019, e8275491. doi:10.1155/2019/8275491.

[59] Zhou, Y., Zhang, J., Hu, S., 2021. Regression analysis and driving force model building of CO2 emissions in China. Scientific Reports 11, 6715. doi:10.1038/s41598-021-86183-5.

[60] Zhu, J., Deng, F., Zhao, J., Zheng, H., . Attention-based parallel networks (APNet) for PM2.5 spatiotemporal prediction 769, 145082. doi:10.1016/j.scitotenv.2021.145082.