

# Dynamic Modes as Time Representation for Spatiotemporal Forecasting

Menglin Kong, Vincent Zhihao Zheng, Xudong Wang and Lijun Sun\*

McGill University, Montreal, Canada

{menglin.kong, zhihao.zheng, xudong.wang2}@mail.mcgill.ca, lijun.sun@mcgill.ca

## Abstract

This paper introduces a data-driven time embedding method for modeling long-range seasonal dependencies in spatiotemporal forecasting tasks. The proposed approach employs Dynamic Mode Decomposition (DMD) to extract temporal modes directly from observed data, eliminating the need for explicit timestamps or hand-crafted time features. These temporal modes serve as time representations that can be seamlessly integrated into deep spatiotemporal forecasting models. Unlike conventional embeddings such as time-of-day indicators or sinusoidal functions, our method captures complex multi-scale periodicity through spectral analysis of spatiotemporal data. Extensive experiments on urban mobility, highway traffic, and climate datasets demonstrate that the DMD-based embedding consistently improves long-horizon forecasting accuracy, reduces residual correlation, and enhances temporal generalization. The method is lightweight, model-agnostic, and compatible with any architecture that incorporates time covariates.

Schmidhuber1997, Chung *et al.*2014], Temporal Convolutional Networks (TCNs) [Yu and Koltun2016], Graph Convolutional Networks (GCNs) [Defferrard *et al.*2016, Kipf and Welling2016], and Transformer-based models [Cai *et al.*2020, Grigsby *et al.*2021]. While these models effectively capture short-term dynamics, they often struggle to represent global temporal dependencies unless aided by explicit temporal encodings. A common workaround is to include hand-crafted time features (e.g., *time-of-day*, *day-of-week*) [Kazemi *et al.*2019], but such features are often domain-specific and insufficient to express complex multi-scale seasonality.

To address this limitation, we propose a data-driven approach to extract and encode periodic structure directly from observations, using **Dynamic Mode Decomposition (DMD)** [Schmid2022]. By leveraging Koopman operator theory, DMD decomposes spatiotemporal signals into interpretable oscillatory modes without requiring timestamp metadata. We construct time embeddings by extracting the real and imaginary components of the dominant DMD modes [Brunton *et al.*2016a, Jovanović *et al.*2014], resulting in a compact spectral representation of temporal dynamics. These embeddings can be seamlessly integrated as covariates into any forecasting model to enhance its ability to capture global periodic dependencies.

Our method offers three key advantages:

1. **Interpretable:** The learned embeddings reveal dominant periodic components, providing insight into temporal patterns in the data.
2. **Model-agnostic:** The embeddings can be incorporated into any spatiotemporal forecasting architecture that supports time covariates.
3. **Domain-agnostic:** The method is purely data-driven and does not rely on calendar time or hand-crafted features, making it broadly applicable across domains.

We evaluate our method on three real-world spatiotemporal datasets: GZ-METRO (urban metro ridership) [Li *et al.*2018], PEMS04 (highway traffic flow) [Song *et al.*2020], and Daymet (daily climate temperature) [Thornton *et al.*2022]. We integrate our DMD-based time embeddings into several deep learning models and demonstrate improved long-term forecasting accuracy, reduced residual correlations, and enhanced temporal generalization across diverse settings.

## 1 Introduction

Spatiotemporal forecasting plays a central role in diverse real-world applications, including traffic flow prediction [Vlahogianni *et al.*2014], climate modeling [Reichstein *et al.*2019], energy demand estimation [Hong *et al.*2016], and environmental monitoring [Pettorelli *et al.*2014]. These systems often exhibit strong periodic or seasonal patterns—such as daily commuting cycles, weekly temperature fluctuations, or yearly rainfall changes—that are essential for accurate long-term forecasting and planning. However, capturing such long-range periodic structures remains a persistent challenge for spatiotemporal forecasting [Vlahogianni *et al.*2014, Karlaftis and Vlahogianni2011].

Recent advances in deep learning have enabled powerful spatiotemporal forecasting models by combining neural sequence architectures and spatial representations, such as Recurrent Neural Networks (RNNs) [Hochreiter and

\*Corresponding author

82 The rest of the paper is organized as follows. Section 2 re-  
 83 views related work on spatiotemporal forecasting and time  
 84 representations. Section 3 presents our proposed method.  
 85 Section 4 describes the experiments and analysis. Section 5  
 86 concludes the paper.

## 87 2 Related Work

### 88 2.1 Spatiotemporal Forecasting Models

89 Existing spatiotemporal forecasting methods can be broadly  
 90 categorized into three families.

91 **Statistical and linear algebraic models.** Classical statis-  
 92 tical models analyze time series in either the time domain  
 93 (e.g., ARMA, ARIMA, SARIMA [Box *et al.* 2015]) or fre-  
 94 quency domain (e.g., Fourier and wavelet transforms [Cryer  
 95 and Chan2008]). Matrix factorization techniques have also  
 96 been applied to low-rank spatiotemporal forecasting and im-  
 97 putation tasks [Yu *et al.* 2016].

98 **Dynamic systems models.** DMD [Schmid2022, Avila and  
 99 Mezić2020] and Sparse Identification of Nonlinear Dynamics  
 100 (SINDy) [Brunton *et al.* 2016b, Champion *et al.* 2019a] aim to  
 101 recover governing structures from observations using data-  
 102 driven operator-based formulations.

103 **Machine learning models.** Early approaches include  
 104 tree ensembles and support vector machines [Bishop and  
 105 Nasrabadi2006], while deep learning (DL) models have be-  
 106 come the dominant paradigm for capturing complex spa-  
 107 tiotemporal dependencies [LeCun *et al.* 2015]. Our work con-  
 108 tributes to this direction by enhancing DL models with a  
 109 novel spectral time embedding.

### 110 2.2 Deep Learning for Spatiotemporal Forecasting

111 DL-based spatiotemporal models combine temporal encoders  
 112 and spatial graph modules. DCRNN [Li *et al.* 2018] integrates  
 113 diffusion convolution into GRUs to model both space and  
 114 time. STGCN [Yu *et al.* 2018] uses GCNs for spatial structure  
 115 and CNNs for temporal encoding. ASTGCN [Guo *et al.* 2019]  
 116 adds spatial-temporal attention to the STGCN backbone.

117 Graph WaveNet [Wu *et al.* 2019] introduces an adaptive  
 118 adjacency matrix and dilated temporal convolutions, offer-  
 119 ing flexibility in learning the spatial graph structure. STS-  
 120 GCN [Song *et al.* 2020] models temporal adjacency by link-  
 121 ing graphs across consecutive time steps. GMAN [Zheng *et*  
 122 *al.* 2020] adopts dual attention over space and time with em-  
 123 bedding fusion. FC-GAGA [Oreshkin *et al.* 2021] extends N-  
 124 BEATS [Oreshkin *et al.* 2020] with a gated graph layer that  
 125 learns sparse, non-Markovian structures.

126 While spatial modules have become increasingly expres-  
 127 sive, temporal modeling in these architectures often remains  
 128 confined to short-range input windows (e.g., one hour), mak-  
 129 ing it difficult to capture long-term seasonal dynamics.

### 130 2.3 Time Embedding and Periodic Structure 131 Learning

132 Encoding long-term temporal dependency is critical in spa-  
 133 tiotemporal forecasting due to prevalent seasonal patterns.  
 134 The most common practice is to use hand-crafted features  
 135 such as *time-of-day* or *day-of-week*, which capture fixed daily

or weekly periodicities. However, these features are rigid and  
 136 may not reflect latent periodic structures in the data.  
 137

138 Time2Vec [Kazemi *et al.* 2019] extends sinusoidal posi-  
 139 tional encoding [Vaswani *et al.* 2017] by learning frequencies  
 140 and phase shifts, offering a flexible way to encode global  
 141 time. This method has been widely adopted in forecasting  
 142 tasks [Grigsby *et al.* 2021] and is related to Fourier feature  
 143 mappings [Tancik *et al.* 2020, Li *et al.* 2021]. However, re-  
 144 cent studies suggest that Fourier features often overfit high-  
 145 frequency fluctuations and struggle to encode low-frequency  
 146 seasonality [Tancik *et al.* 2020].

147 In contrast, we propose a spectral embedding approach  
 148 based on DMD. Our method extracts dominant oscillatory  
 149 modes from observed time series via Koopman operator anal-  
 150 ysis [Brunton *et al.* 2016a, Jovanović *et al.* 2014]. The real  
 151 and imaginary components of these modes form a compact  
 152 and interpretable time representation, which captures com-  
 153 plex and data-driven periodic patterns beyond hand-crafted or  
 154 fixed-frequency alternatives. These embeddings can be seam-  
 155 lessly integrated into any DL forecasting model to enhance  
 156 long-term temporal modeling.

## 157 3 Methodology

158 In this section, we introduce our method to obtain the spectral  
 159 properties of spatiotemporal data and the method to encode  
 160 periodic information in forecasting models.

### 161 3.1 Spatiotemporal Forecasting

162 We consider a general spatiotemporal system observed over  
 163 a network, which can be represented as a directed graph  
 164  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ , where  $\mathcal{V}$  is a set of spatial locations or sensing  
 165 nodes with cardinality  $|\mathcal{V}| = N$ ,  $\mathcal{E}$  is a set of edges represent-  
 166 ing spatial relationships, and  $\mathbf{A} \in \mathbb{R}^{N \times N}$  is the weighted  
 167 adjacency matrix encoding spatial proximity or correlation.

168 Let  $\mathbf{z}_t \in \mathbb{R}^N$  denote the signal observed at all  $N$  locations  
 169 at time  $t$ , and let  $\mathbf{c}_t \in \mathbb{R}^D$  represent auxiliary covariates (e.g.,  
 170 temporal features) associated with time  $t$ . The spatiotemporal  
 171 forecasting task aims to learn a function  $f(\cdot)$  that maps the  
 172 past  $P$  time steps of observations and covariates to the next  
 173  $Q$  steps of future signals:

$$[\mathbf{X}_{t-P+1:t}, \mathbf{C}_{t-P+1:t+Q}] \xrightarrow{f(\cdot)} [\mathbf{Y}_{t+1:t+Q}], \quad (1)$$

174 where  $\mathbf{X}_{t-P+1:t} = \{\mathbf{z}_{t-P+1}, \dots, \mathbf{z}_t\}$  is the histor-  
 175 ical sequence of observed signals,  $\mathbf{C}_{t-P+1:t+Q} =$   
 176  $\{\mathbf{c}_{t-P+1}, \dots, \mathbf{c}_{t+Q}\}$  is the sequence of covariates (possibly  
 177 including future ones), and  $\mathbf{Y}_{t+1:t+Q} = \{\mathbf{z}_{t+1}, \dots, \mathbf{z}_{t+Q}\}$  is  
 178 the target sequence to be predicted.

179 The function  $f(\cdot)$  typically takes the form of a multivariate  
 180 sequence-to-sequence (Seq2Seq) model and can be parame-  
 181 terized by either classical statistical models or modern deep  
 182 learning models. In practice, the forecasting horizon  $Q$  is of-  
 183 ten set equal to the historical window length  $P$ .

### 184 3.2 Spectral Analysis of Spatiotemporal Data with 185 DMD

186 To extract periodic structure from spatiotemporal data, we  
 187 propose a three-stage spectral analysis pipeline based on  
 188 DMD. The method consists of:

- 189 1. **Hankel embedding** to expand the effective spatial dimension of the data (since typically  $N \ll T$  in spatiotemporal forecasting) [Champion *et al.* 2019b, Kamb *et al.* 2020, Wang and Sun 2022];  
190 2. **Total DMD (TDMD)** for unbiased, noise-aware spectral decomposition [Hemati *et al.* 2017];  
191 3. **Sparsity-Promoting DMD (SPDMD)** to select dominant dynamic modes [Jovanović *et al.* 2014].

192 Our goal is to obtain a compact and interpretable set of oscillatory components whose temporal evolution will serve as covariates in downstream forecasting models.

193 Let  $\mathbf{z}_t \in \mathbb{R}^N$  denote the observed system state at time step  $t$ , across  $N$  spatial nodes. The dynamics of the system can be modeled as a nonlinear discrete-time system:

$$\mathbf{z}_{t+1} = f(\mathbf{z}_t), \quad (2)$$

200 where  $f$  is an unknown nonlinear function. To linearize the system, we follow the Koopman operator framework [Schmid2022], which lifts the dynamics into a higher-dimensional observable space via a function  $\phi(\cdot)$  such that:

$$\phi(\mathbf{z}_{t+1}) = \mathcal{K}\phi(\mathbf{z}_t), \quad (3)$$

207 where  $\mathcal{K}$  is a linear (but infinite-dimensional) Koopman operator. Its spectral decomposition yields:

$$\mathcal{K}\Psi = \Psi\Lambda, \quad (4)$$

209 where  $\Psi \in \mathbb{C}^{d \times r}$  contains Koopman eigenfunctions and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$  is a diagonal matrix of corresponding eigenvalues. Assuming the initial observable satisfies  $\phi(\mathbf{z}_0) = \Psi\mathbf{a}$ , the time evolution becomes:

$$\phi(\mathbf{z}_t) = \mathcal{K}^t\phi(\mathbf{z}_0) = \Psi\Lambda^t\mathbf{a}. \quad (5)$$

213 Let  $\phi_t = \phi(\mathbf{z}_t)$  denote the observable at time  $t$ . We stack the raw input signals over time into the data matrix  $\mathbf{D} = [\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_{T-1}] \in \mathbb{R}^{N \times T}$ . Its Koopman-based approximation becomes:

$$\hat{\mathbf{D}} = \Psi \cdot \text{diag}(\mathbf{a}) \cdot \mathbf{C}, \quad (6)$$

217 where  $\mathbf{C} \in \mathbb{C}^{r \times T}$  is a Vandermonde matrix with:

$$\mathbf{C}_{ij} = \lambda_i^{j-1}, \quad i = 1, \dots, r, \quad j = 1, \dots, T.$$

218 To encourage interpretability, we apply SPDMD [Jovanović *et al.* 2014] to prune to the  $r$  dominant modes, yielding:

$$\hat{\mathbf{D}} \approx \bar{\Psi} \cdot \text{diag}(\bar{\mathbf{a}}) \cdot \bar{\mathbf{C}}, \quad (7)$$

221 where  $\bar{\Psi} \in \mathbb{C}^{N \times r}$  contains selected spatial modes,  $\bar{\mathbf{a}} \in \mathbb{C}^r$  are their amplitudes, and  $\bar{\mathbf{C}} \in \mathbb{C}^{r \times T}$  captures their temporal evolution.

224 To improve the performance of DMD on high-dimensional data, we first enhance the data using a circulant Hankel embedding [Schmid2022, Wang and Sun 2023], which converts  $\mathbf{D}$  into a tall matrix:

$$\mathbf{H} = \begin{bmatrix} \mathbf{z}_0 & \mathbf{z}_1 & \cdots & \mathbf{z}_{T-1} \\ \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}_{T-1} & \mathbf{z}_T & \cdots & \mathbf{z}_{T-2} \end{bmatrix} \in \mathbb{R}^{N\tau \times T}. \quad (8)$$

228 Since directly computing the SVD of  $\mathbf{HH}^\top \in \mathbb{R}^{N\tau \times N\tau}$  is  
229 computationally expensive for large  $\tau$ , we apply the *method*  
230 of *snapshots* [Brunton and Kutz2022], which leverages the  
231 symmetry of the inner product space to reduce the computation  
232 to a smaller Gram matrix. Specifically, we perform the  
233 truncated SVD on the smaller matrix  $\mathbf{H}^\top \mathbf{H} \in \mathbb{R}^{T \times T}$ :

$$\mathbf{H}^\top \mathbf{H} = \tilde{\mathbf{V}} \tilde{\Sigma}^2 \tilde{\mathbf{V}}^\top, \quad (9)$$

234 where  $\tilde{\mathbf{V}} \in \mathbb{R}^{T \times r}$  contains the top- $r$  right singular vectors;  
235  $\tilde{\Sigma} \in \mathbb{R}^{r \times r}$  is the diagonal matrix of singular values.  
236

237 Then, we recover the corresponding left singular vectors  
238 using:

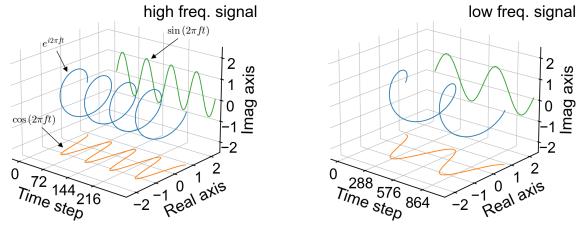
$$\tilde{\mathbf{U}} = \mathbf{H} \tilde{\mathbf{V}} \tilde{\Sigma}^{-1}, \quad (10)$$

239 with  $\tilde{\mathbf{U}} \in \mathbb{R}^{N\tau \times r}$  forming an orthonormal basis for the dom-  
240 inant subspace of  $\mathbf{H}$ .

241 Let  $\mathbf{H}' \in \mathbb{R}^{N\tau \times T}$  denote the time-shifted version of  $\mathbf{H}$ ,  
242 i.e., containing the next-step snapshots. Using  $\tilde{\mathbf{U}}$  and  $\tilde{\mathbf{V}}$ , we  
243 construct the reduced Koopman operator:

$$\mathbf{A}_{\text{DMD}} = \tilde{\mathbf{U}}^\top \mathbf{H}' \tilde{\mathbf{V}} \tilde{\Sigma}^{-1}, \quad (11)$$

244 which approximates the linear transition matrix in the low-  
245 rank temporal subspace. The eigenvalues  $\{\lambda_1, \dots, \lambda_r\}$   
246 of  $\mathbf{A}_{\text{DMD}}$  characterize the oscillatory behavior of the sys-  
247 tem (magnitude  $\rightarrow$  growth/decay, imaginary part  $\rightarrow$  fre-  
248 quency) [Schmid2022].



249 Figure 1: Visualization of temporal modes extracted via DMD for  
250 the GZ-METRO dataset. Each mode evolves as a complex expon-  
251 ential  $e^{i2\pi f t}$ , with real (cosine), imaginary (sine), and modulus com-  
252 ponents. Left: high-frequency mode; Right: low-frequency mode.  
253 The real and imaginary parts form the final time embedding.

254 From these eigenvalues, we construct the temporal dynam-  
255 ics matrix  $\bar{\mathbf{C}} \in \mathbb{C}^{r \times T}$  as a Vandermonde matrix:  
256

$$\bar{\mathbf{C}} = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{T-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{T-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_r & \lambda_r^2 & \cdots & \lambda_r^{T-1} \end{bmatrix}. \quad (12)$$

257 Each row in  $\bar{\mathbf{C}}$  defines the temporal profile of a dominant  
258 DMD mode. These complex-valued signals form the basis of  
259 our time embedding, as illustrated in Figure 1.

260 **Final Output: Time Embedding for Forecasting.** At each  
261 time step  $t$ , we extract the real and imaginary parts of  $\bar{\mathbf{C}}[:, t]$   
262 to construct a  $2r$ -dimensional time covariate:

$$\mathbf{c}_t^{(\text{DMD})} = \begin{bmatrix} \text{Re}(\bar{\mathbf{C}}[:, t]) \\ \text{Im}(\bar{\mathbf{C}}[:, t]) \end{bmatrix} \in \mathbb{R}^{2r}. \quad (13)$$

256 This embedding can be appended to any spatiotemporal  
 257 forecasting model. If the original model input has shape  
 258  $(N, T, m)$ , representing spatial dimension, temporal steps,  
 259 and feature channels respectively, then our method expands it  
 260 to shape  $(N, T, m + 2r)$  by concatenating DMD-based time  
 261 embeddings at each time step.

## 262 4 Experiments

263 To evaluate the effectiveness of the proposed method, we carried  
 264 out experiments on three spatiotemporal datasets: GZ-  
 265 METRO, PEMS04, and Daymet. GZ-METRO records 15-  
 266 min ridership based on smart card tap-in data at 159 metro  
 267 stations for 3 months in Guangzhou, China. PEMS04 is a  
 268 highway traffic flow datasets used in [Song *et al.* 2020]. The  
 269 dataset records 5-min aggregated traffic flow from the Cal-  
 270 trans Performance Measurement System (PeMS) [Chen *et*  
 271 *al.* 2001]. Daymet is a dataset that provides the daily max-  
 272 imum temperature recordings for Continental North Amer-  
 273 ica from 2010 to 2021 [Thornton *et al.* 2022]. The summary  
 274 of datasets can be found in Table 1. For GZ-METRO and  
 275 Daymet, 70% of the data were used for training, 20% for  
 276 testing, and the remaining 10% for validation. For PEMS04,  
 277 the split ratio is 6:2:2 as used in the original paper [Song  
 278 *et al.* 2020]. All datasets were applied z-score normalization  
 279 with statistics obtained from the training set. We selected sev-  
 280 eral spatiotemporal forecasting frameworks as the base mod-  
 281 els:

- 282 • FC-LSTM [Sutskever *et al.* 2014]: A Sequence-to-  
 283 Sequence LSTM model with fully-connected LSTM  
 284 layers in both encoder and decoder.
- 285 • DCRNN [Li *et al.* 2018]: Diffusion convolution recur-  
 286 rent neural network, which wraps diffusion convolution  
 287 operation into recurrent neural networks to achieve spa-  
 288 tiotemporal forecasting.
- 289 • Graph WaveNet [Wu *et al.* 2019]: A spatiotemporal fore-  
 290 casting model that combines dilated 1D convolution for  
 291 modeling temporal dynamics and graph convolution for  
 292 modeling spatial dynamics.
- 293 • AGCRN [Bai *et al.* 2020]: Similar to Graph WaveNet  
 294 which uses an adaptive adjacency matrix, the model uses  
 295 GRU to model temporal dependency.

Table 1: Dataset Description

Datasets	Number of nodes	Time range
GZ-METRO	159	7/1/2017 - 9/29/2017
PEMS04	307	1/1/2018 - 2/28/2018
Daymet	628	1/1/2010 - 12/31/2021

296 We implemented these models using the original source  
 297 code (or their PyTorch version). All models use 12 steps  
 298 of historical observations ( $P = 12$ ) to predict 12 steps of  
 299 future values ( $Q = 12$ ). The evaluation metrics are mean  
 300 absolute error (MAE) and root mean squared error (RMSE),

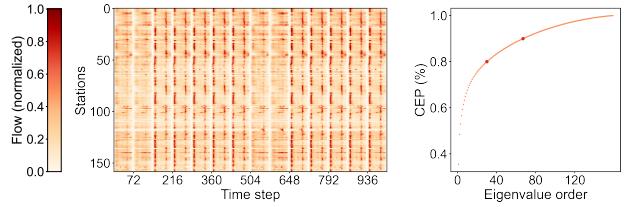


Figure 2: Left: Normalized ridership at all stations of the  
 301 Guangzhou Metro. Right: Cumulative eigenvalue percentage (CEP)  
 302 calculated by SVD.

where missing values are excluded:

$$303 \text{MAE} = \frac{1}{N} \sum_{i=1}^N |z_i - \hat{z}_i|, \quad (14)$$

$$304 \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_i - \hat{z}_i)^2}. \quad (15)$$

305 Early stopping was applied to prevent over-fitting when the  
 306 validation loss keeps growing for more than 30 epochs. The  
 307 learning rate was set to 0.001 with a decay rate of 0.0001.  
 308 The Adam algorithm was used for optimization. Our exper-  
 309 iments were conducted under a computer environment with  
 310 one Intel(R) Xeon(R) CPU E5-2698 v4 @ 2.20GHz and four  
 311 NVIDIA Tesla V100 GPU. All results were computed based  
 312 on the average of three runs of training.

### 313 4.1 Quantitative Study

314 We begin our quantitative analysis by examining the temporal  
 315 and spectral characteristics of the GZ-METRO dataset. Fig-  
 316 ure 2 (left) visualizes the ridership flow over two weeks for all  
 317 metro stations. The repeated daily and weekly patterns sug-  
 318 gest the presence of strong periodic signals in both time and  
 319 space. To further quantify this structure, we perform sin-  
 320 gular value decomposition (SVD) and compute the cumula-  
 321 tive eigenvalue percentage (CEP), as shown in Figure 2 (right).  
 322 We observe that more than 90% of the data variance can be  
 323 captured using only a small number of singular components,  
 324 confirming the low-rank nature of the data and motivating the  
 325 use of spectral methods such as DMD to extract dominant  
 326 temporal dynamics.

327 Table 2 summarizes the performance of baseline forecast-  
 328 ing models on the three datasets, with and without our DMD-  
 329 based time embedding. Across all datasets and forecasting  
 330 horizons (3, 6, and 12 steps), we observe consistent improve-  
 331 ments when incorporating the proposed time covariates.

332 On GZ-METRO, the gain is particularly significant at  
 333 longer horizons (e.g., 12-step ahead), where capturing long-  
 334 term periodic patterns becomes more important. For instance,  
 335 Graph WaveNet improves from 93.62 (RMSE) to 87.79 with  
 336 our embedding, while DCRNN drops from 105.48 to 91.98.  
 337 This highlights the benefit of DMD-based features in model-  
 338 ing structured temporal dependencies in metro ridership data,  
 339 which has high regularity and low noise due to 15-minute ag-  
 340 gregation.

Table 2: Performance comparison of models with/without time embedding

Data	Model	3-step		6-step		12-step	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
GZ-METRO	FC-LSTM	w/o 48.05 w/ <b>47.4</b>	92.33 <b>90.3</b>	50.44 <b>49.02</b>	97.43 <b>94.13</b>	53.38 <b>51.18</b>	103.57 <b>99.66</b>
	DCRNN	w/o 43.39 w/ <b>42.32</b>	83.75 <b>80.29</b>	49.08 <b>45.46</b>	92.35 <b>84.71</b>	55.43 <b>50.07</b>	105.48 <b>91.98</b>
	AGCRN	w/o 42.66 w/ <b>40.7</b>	81.5 <b>77.61</b>	47.12 <b>44.35</b>	91.14 <b>84.08</b>	51.13 <b>48.77</b>	97.44 <b>91.46</b>
	Graph WaveNet	w/o 41.26 w/ <b>39.47</b>	76.33 <b>73.21</b>	44.66 <b>42.87</b>	82.82 <b>79.99</b>	50.05 <b>47.53</b>	93.62 <b>87.79</b>
	FC-LSTM	w/o 22.36 w/ <b>22.11</b>	36.44 <b>36.21</b>	22.52 <b>22.17</b>	36.57 <b>36.28</b>	22.96 <b>22.41</b>	36.99 <b>36.51</b>
	DCRNN	w/o 19.89 w/ <b>19.01</b>	31.29 <b>30.32</b>	21.95 <b>20.2</b>	34.46 <b>32.34</b>	26.1 <b>22.14</b>	40.81 <b>35.42</b>
PEMS04	AGCRN	w/o 18.65 w/ <b>18.48</b>	<b>29.83</b> 29.88	19.56 <b>19.24</b>	31.41 <b>31.34</b>	21.05 <b>20.44</b>	33.5 <b>33.16</b>
	Graph WaveNet	w/o 18.2 w/ <b>18.08</b>	29.1 <b>29.03</b>	19.35 <b>18.89</b>	30.74 <b>30.45</b>	21.32 <b>20.23</b>	33.28 <b>32.43</b>
	FC-LSTM	w/o <b>3.89</b> w/ 3.9	<b>5.11</b> 5.13	4.05 <b>4.03</b>	5.3 <b>5.29</b>	4.15 <b>4.09</b>	5.45 <b>5.39</b>
	AGCRN	w/o 3.71 w/ <b>3.61</b>	4.85 <b>4.75</b>	3.95 <b>3.86</b>	5.17 <b>5.07</b>	4.09 <b>3.92</b>	5.37 <b>5.16</b>
	Graph WaveNet	w/o 3.79 w/ <b>3.76</b>	4.96 <b>4.92</b>	4.07 <b>3.95</b>	5.33 <b>5.17</b>	4.31 <b>4.02</b>	5.7 <b>5.27</b>

On PEMS04, which has higher temporal resolution (5-minute) and shorter-range patterns, the performance improvement is less pronounced but still present. For example, AGCRN’s 12-step RMSE improves from 33.5 to 33.16, and Graph WaveNet improves from 33.28 to 32.43. The smaller margin may be attributed to the noisier nature of traffic data and the relatively short forecasting horizon (1 hour), which limits the benefit of global periodic signals.

For the Daymet dataset, which tracks long-term daily temperature patterns, our method also leads to consistent performance improvements. Graph WaveNet’s 12-step RMSE is reduced from 5.7 to 5.27, and AGCRN improves from 5.37 to 5.16. These results support that our time embedding is generalizable across domains and not restricted to transportation data.

Overall, the results suggest that our DMD-based embedding offers a lightweight and effective way to inject long-term periodic structure into spatiotemporal forecasting models, improving both short-term accuracy and long-term generalization. The extracted complex dynamics capture multiple time scales, including daily and weekly cycles, which are otherwise hard to model with conventional hand-crafted or learnable time features alone.

## 4.2 Qualitative Study

We conduct qualitative analysis of our method using Graph WaveNet as the backbone model and GZ-METRO as the evaluation dataset.

### Effectiveness of DMD-based time embedding

To evaluate the effectiveness of our DMD-based time embedding, we compare it with several alternatives for encoding periodic patterns: (i) time of day (D), (ii) time of day, day of week (DW), and (iii) the learnable sinusoidal embedding method Time2Vec [Kazemi *et al.* 2019].

Table 3: Performance comparison of different time embedding methods using Graph WaveNet

Data	Method	3-step		6-step		12-step	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
GZ-METRO	D	41.48	77.23	44.89	83.64	48.72	90.62
	DW	43.58	83.75	47.02	91.46	50.62	96.00
	Time2Vec	40.17	74.68	44.05	81.83	48.40	91.09
	Ours	<b>39.47</b>	<b>73.21</b>	<b>42.87</b>	<b>79.99</b>	<b>47.53</b>	<b>87.79</b>
PEMS04	D	18.32	29.28	19.53	31.03	21.42	33.62
	DW	18.13	29.11	19.16	30.64	20.73	32.86
	Time2Vec	18.26	29.21	19.48	30.93	21.47	33.62
	Ours	<b>18.08</b>	<b>29.03</b>	<b>18.89</b>	<b>30.45</b>	<b>20.23</b>	<b>32.43</b>
Daymet	Time2Vec	<b>3.75</b>	<b>4.91</b>	4.02	5.26	4.23	5.58
	Ours	3.76	4.92	<b>3.95</b>	<b>5.17</b>	<b>4.02</b>	<b>5.27</b>

Time of day is encoded as a continuous real-valued feature, whereas day of week is represented using one-hot encoding. To ensure fairness, we align the embedding dimensions of our method and Time2Vec with that of the DW method. For example, if DW uses an embedding dimension of 8, we set the number of modes in our method to  $r = 4$ . As shown in Table 3, our method consistently outperforms all baselines across multiple datasets (GZ-METRO, PEMS04, Daymet) and prediction horizons (3-step, 6-step, 12-step). Notably, in GZ-METRO and PEMS04, our method achieves the lowest RMSE and MAE values in all settings. For example, at 12-step forecasting on GZ-METRO, our method achieves an RMSE of 87.79 compared to 91.09 from Time2Vec and 96.00 from discrete encoding.

Our observations align with the findings of [Tancik *et al.* 2020], which showed that learned Fourier-based time features tend to concentrate on high-frequency components and struggle to capture long-range periodicity. We further inspected the learned frequencies of Time2Vec and found that they mostly correspond to short-term fluctuations with periods between 5 and 15 time steps (i.e., 25–75 minutes), indicating a failure to model daily or weekly cycles. By contrast, our DMD-based embedding captures multiple frequencies of varying scales, including long-term modes, as visualized in Figure 1.

### Residual correlation and temporal dependency analysis

To evaluate whether our method effectively reduces systematic errors and better captures long-term temporal dependencies, we conduct two types of residual analysis on the GZ-METRO dataset using Graph WaveNet with and without our DMD-based time embedding.

Figure 3a and Figure 3b show the residual correlation matrices computed from the 12-step-ahead prediction errors at three time lags:  $S = 0, 72$ , and  $504$  steps, corresponding to contemporaneous, daily, and weekly intervals. Each panel visualizes the correlations between residual vectors  $\eta_t$  and  $\eta_{t-S}$  over all spatiotemporal positions. Without our method, we observe strong overall correlations at  $S = 0$  (0.138), and noticeable dependencies at  $S = 72$  (0.081) and  $S = 504$  (0.096). In contrast, our method reduces these values to 0.104, 0.063, and 0.079, respectively. These reductions indicate that our model effectively captures structured periodic patterns and reduces error autocorrelation across periods.

To further validate this observation, we analyze the auto-

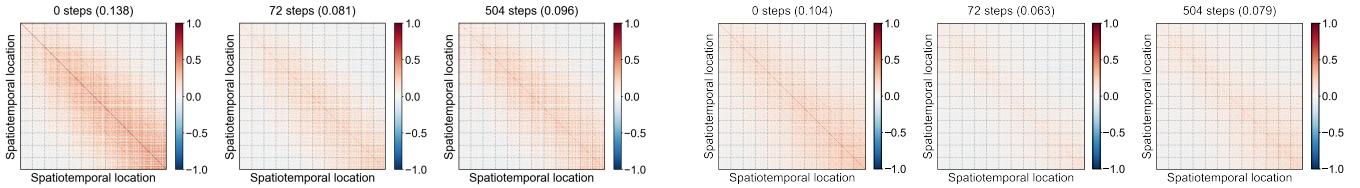


Figure 3: Residual correlation matrices of GZ-METRO at time lags  $S = 0, 72$ , and  $504$ . Each panel shows absolute residual correlation values between spatiotemporal locations. Values in parentheses indicate mean absolute correlation.

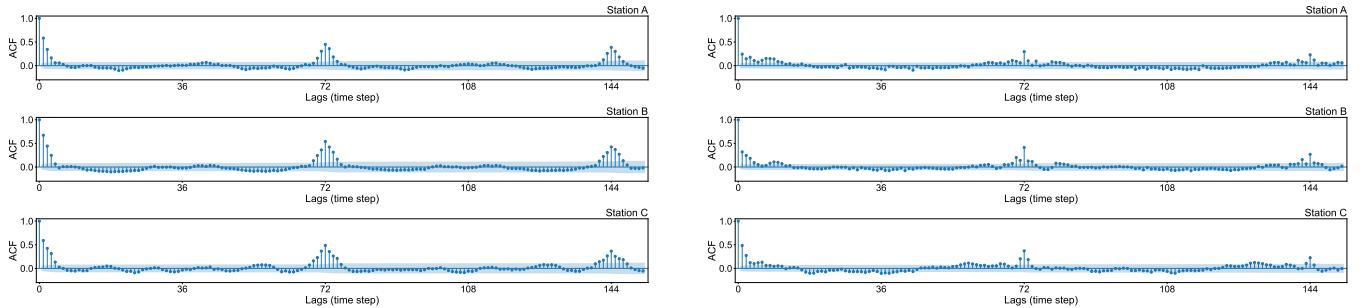


Figure 4: Autocorrelation functions (ACF) of 12-step prediction residuals from three representative sensors in GZ-METRO. Peaks around lag 72 indicate periodic errors; these are reduced by our time embedding.

correlation function (ACF) of the prediction residuals from three representative stations (A, B, and C). Figure 4a shows the ACF plots of the original Graph WaveNet model. We observe clear peaks around lag 72 and 144, corresponding to daily and bi-daily periodicity, suggesting that the model fails to fully capture these regular patterns, which manifest as cyclic structure in the residuals. In comparison, Figure 4b illustrates the ACF of the residuals using Graph WaveNet augmented with our time embedding. The periodic peaks at lag 72 and 144 are substantially diminished across all three stations. This indicates that the DMD-based time features help the model internalize the dominant periodic components in the data, leading to less temporally correlated residuals and improved forecasting generalization.

## 5 Conclusion

We propose a data-driven time embedding method based on DMD to enhance spatiotemporal forecasting. By extracting dominant temporal modes from the data and representing them through complex-valued dynamic patterns, our method captures long-term periodic structures and injects them as interpretable time covariates into forecasting models. Empirical results across multiple datasets show that our approach consistently improves long-horizon prediction accuracy and reduces residual autocorrelation, especially in datasets with strong seasonal patterns like GZ-METRO. The method is model-agnostic and can be easily integrated into existing architectures with time feature input.

However, on noisy datasets with weak or irregular periodicity—such as those with short aggregation intervals—it

benefits are less pronounced. Future work will focus on adapting the method to handle such non-stationary settings and extending it to architectures with different input paradigms, such as Transformers [Vaswani *et al.* 2017] where time embeddings interact with token-level representations.

## Ethical Statement

There are no ethical issues.

## References

- [Avila and Mezić, 2020] Allan M Avila and I Mezić. Data-driven analysis and forecasting of highway traffic dynamics. *Nature communications*, 11(1):1–16, 2020.
- [Bai *et al.*, 2020] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in Neural Information Processing Systems*, 33:17804–17815, 2020.
- [Bishop and Nasrabadi, 2006] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [Box *et al.*, 2015] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [Brunton and Kutz, 2022] Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.

- 470 [Brunton *et al.*, 2016a] Bingni W Brunton, Lise A Johnson,  
 471 Jeffrey G Ojemann, and J Nathan Kutz. Extracting spatial–  
 472 temporal coherent patterns in large-scale neural recordings  
 473 using dynamic mode decomposition. *Journal of neuro-  
 474 science methods*, 258:1–15, 2016.
- 475 [Brunton *et al.*, 2016b] Steven L Brunton, Joshua L Proctor,  
 476 and J Nathan Kutz. Discovering governing equations from  
 477 data by sparse identification of nonlinear dynamical sys-  
 478 tems. *Proceedings of the national academy of sciences*,  
 479 113(15):3932–3937, 2016.
- 480 [Cai *et al.*, 2020] Ling Cai, Krzysztof Janowicz, Gengchen  
 481 Mai, Bo Yan, and Rui Zhu. Traffic transformer: Captur-  
 482 ing the continuity and periodicity of time series for traffic  
 483 forecasting. *Transactions in GIS*, 24(3):736–755, 2020.
- 484 [Champion *et al.*, 2019a] Kathleen Champion, Bethany  
 485 Lusch, J Nathan Kutz, and Steven L Brunton. Data-  
 486 driven discovery of coordinates and governing equations.  
 487 *Proceedings of the National Academy of Sciences*,  
 488 116(45):22445–22451, 2019.
- 489 [Champion *et al.*, 2019b] Kathleen P Champion, Steven L  
 490 Brunton, and J Nathan Kutz. Discovery of nonlinear multi-  
 491 scale systems: Sampling strategies and embeddings. *SIAM  
 492 Journal on Applied Dynamical Systems*, 18(1):312–333,  
 493 2019.
- 494 [Chen *et al.*, 2001] Chao Chen, Karl Petty, Alexander Sk-  
 495 abardonis, Pravin Varaiya, and Zhanfeng Jia. Freeway per-  
 496 formance measurement system: mining loop detector data.  
 497 *Transportation Research Record*, 1748(1):96–102, 2001.
- 498 [Chung *et al.*, 2014] Junyoung Chung, Caglar Gulcehre,  
 499 KyungHyun Cho, and Yoshua Bengio. Empirical Evalu-  
 500 ation of Gated Recurrent Neural Networks on Sequence  
 501 Modeling. *arXiv*, 2014.
- 502 [Cryer and Chan, 2008] Jonathan D Cryer and Kung-Sik  
 503 Chan. *Time series analysis: with applications in R*, vol-  
 504 ume 2. Springer, 2008.
- 505 [Defferrard *et al.*, 2016] Michaël Defferrard, Xavier Bres-  
 506 son, and Pierre Vandergheynst. Convolutional neural  
 507 networks on graphs with fast localized spectral filtering.  
 508 *Advances in neural information processing systems*, 29,  
 509 2016.
- 510 [Grigsby *et al.*, 2021] Jake Grigsby, Zhe Wang, and Yanjun  
 511 Qi. Long-range transformers for dynamic spatiotemporal  
 512 forecasting. *arXiv preprint arXiv:2109.12218*, 2021.
- 513 [Guo *et al.*, 2019] Shengnan Guo, Youfang Lin, Ning Feng,  
 514 Chao Song, and Huaiyu Wan. Attention based spatial-  
 515 temporal graph convolutional networks for traffic flow  
 516 forecasting. In *Proceedings of the AAAI conference on  
 517 artificial intelligence*, volume 33, pages 922–929, 2019.
- 518 [Hemati *et al.*, 2017] Maziar S Hemati, Clarence W Rowley,  
 519 Eric A Deem, and Louis N Cattafesta. De-biasing the dy-  
 520 namic mode decomposition for applied koopman spectral  
 521 analysis of noisy datasets. *Theoretical and Computational  
 522 Fluid Dynamics*, 31(4):349–368, 2017.
- 523 [Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and  
 524 Jrgen Schmidhuber. Long Short-Term Memory. *Neural  
 525 Computation*, 9(8):1735–1780, 1997.
- 526 [Hong *et al.*, 2016] Tao Hong, Pierre Pinson, and Shu Fan.  
 527 Probabilistic energy forecasting: Global energy forecast-  
 528 ing competition 2014 and beyond. *International Journal  
 529 of Forecasting*, 32(3):896–913, 2016.
- 530 [Jovanović *et al.*, 2014] Mihailo R Jovanović, Peter J  
 531 Schmid, and Joseph W Nichols. Sparsity-promoting  
 532 dynamic mode decomposition. *Physics of Fluids*,  
 533 26(2):024103, 2014.
- 534 [Kamb *et al.*, 2020] Mason Kamb, Eurika Kaiser, Steven L  
 535 Brunton, and J Nathan Kutz. Time-delay observables for  
 536 koopman: Theory and applications. *SIAM Journal on Ap-  
 537 plied Dynamical Systems*, 19(2):886–917, 2020.
- 538 [Karlaftis and Vlahogianni, 2011] Matthew G Karlaftis and  
 539 Eleni I Vlahogianni. Statistical methods versus neural  
 540 networks in transportation research: Differences, similar-  
 541 ities and some insights. *Transportation Research Part C:  
 542 Emerging Technologies*, 19(3):387–399, 2011.
- 543 [Kazemi *et al.*, 2019] Seyed Mehran Kazemi, Rishab Goel,  
 544 Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, San-  
 545 jay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and  
 546 Marcus Brubaker. Time2vec: Learning a vector represen-  
 547 tation of time. *arXiv preprint arXiv:1907.05321*, 2019.
- 548 [Kipf and Welling, 2016] Thomas N Kipf and Max Welling.  
 549 Semi-supervised classification with graph convolutional  
 550 networks. In *International conference on learning repre-  
 551 sentations*, 2016.
- 552 [LeCun *et al.*, 2015] Yann LeCun, Yoshua Bengio, and Ge-  
 553 offrey Hinton. Deep learning. *nature*, 521(7553):436–444,  
 554 2015.
- 555 [Li *et al.*, 2018] Yaguang Li, Rose Yu, Cyrus Shahabi, and  
 556 Yan Liu. Diffusion convolutional recurrent neural net-  
 557 work: Data-driven traffic forecasting. In *International  
 558 conference on learning representations*, 2018.
- 559 [Li *et al.*, 2021] Yang Li, Si Si, Gang Li, Cho-Jui Hsieh,  
 560 and Samy Bengio. Learnable fourier features for multi-  
 561 dimensional spatial positional encoding. *Advances in  
 562 Neural Information Processing Systems*, 34:15816–15829,  
 563 2021.
- 564 [Oreshkin *et al.*, 2020] Boris N Oreshkin, Dmitri Carpow,  
 565 Nicolas Chapados, and Yoshua Bengio. N-beats: Neu-  
 566 ral basis expansion analysis for interpretable time series  
 567 forecasting. In *International conference on learning rep-  
 568 resentations*, 2020.
- 569 [Oreshkin *et al.*, 2021] Boris N Oreshkin, Arezou Amini,  
 570 Lucy Coyle, and Mark Coates. Fc-gaga: Fully connected  
 571 gated graph architecture for spatio-temporal traffic fore-  
 572 casting. In *Proceedings of the AAAI Conference on Artifi-  
 573 cial Intelligence*, volume 35, pages 9233–9241, 2021.
- 574 [Pettorelli *et al.*, 2014] Nathalie Pettorelli, William F Lau-  
 575 rance, Timothy G O’Brien, Martin Wegmann, Harini Na-  
 576 gendra, and Woody Turner. Satellite remote sensing for

- 577 applied ecologists: opportunities and challenges. *Journal*  
578 *of Applied Ecology*, 51(4):839–848, 2014.
- 579 [Reichstein *et al.*, 2019] Markus Reichstein, Gustau Camps-  
580 Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno  
581 Carvalhais, and Prabhat. Deep learning and process un-  
582 derstanding for data-driven earth system science. *Nature*,  
583 566(7743):195–204, 2019.
- 584 [Schmid, 2022] Peter J Schmid. Dynamic mode decomposi-  
585 tion and its variants. *Annual Review of Fluid Mechanics*,  
586 54:225–254, 2022.
- 587 [Song *et al.*, 2020] Chao Song, Youfang Lin, Shengnan Guo,  
588 and Huaiyu Wan. Spatial-temporal synchronous graph  
589 convolutional networks: A new framework for spatial-  
590 temporal network data forecasting. In *Proceedings of the*  
591 *AAAI Conference on Artificial Intelligence*, volume 34,  
592 pages 914–921, 2020.
- 593 [Sutskever *et al.*, 2014] Ilya Sutskever, Oriol Vinyals, and  
594 Quoc V Le. Sequence to sequence learning with neural  
595 networks. *Advances in neural information processing sys-*  
596 *tems*, 27, 2014.
- 597 [Tancik *et al.*, 2020] Matthew Tancik, Pratul Srinivasan,  
598 Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan,  
599 Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron,  
600 and Ren Ng. Fourier features let networks learn high fre-  
601 quency functions in low dimensional domains. *Advances*  
602 *in Neural Information Processing Systems*, 33:7537–7547,  
603 2020.
- 604 [Thornton *et al.*, 2022] MM Thornton, R Shrestha, Y Wei,  
605 PE Thornton, S Kao, and BE Wilson. Daymet: Monthly  
606 climate summaries on a 1-km grid for north america, ver-  
607 sion 4 r1. ornl daac, oak ridge, tennessee, usa, 2022.
- 608 [Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki  
609 Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez,  
610 Łukasz Kaiser, and Illia Polosukhin. Attention is all you  
611 need. *Advances in neural information processing systems*,  
612 30, 2017.
- 613 [Vlahogianni *et al.*, 2014] Eleni I Vlahogianni, Matthew G  
614 Karlaftis, and John C Golias. Short-term traffic forecast-  
615 ing: Where we are and where we’re going. *Transportation*  
616 *Research Part C: Emerging Technologies*, 43:3–19, 2014.
- 617 [Wang and Sun, 2022] Xudong Wang and Lijun Sun. Ex-  
618 tracting dynamic mobility patterns by hankel dynamic  
619 modes decomposition. In *The 11th Triennial Symposium*  
620 *on Transportation Analysis, Mauritius Island*, 2022.
- 621 [Wang and Sun, 2023] Xudong Wang and Lijun Sun. Anti-  
622 circulant dynamic mode decomposition with sparsity-  
623 promoting for highway traffic dynamics analysis. *Trans-*  
624 *portation Research Part C: Emerging Technologies*,  
625 153:104178, 2023.
- 626 [Wu *et al.*, 2019] Zonghan Wu, Shirui Pan, Guodong Long,  
627 Jing Jiang, and Chengqi Zhang. Graph wavenet for deep  
628 spatial-temporal graph modeling. page 1907–1913, 2019.
- 629 [Yu and Koltun, 2016] Fisher Yu and Vladlen Koltun. Multi-  
630 scale context aggregation by dilated convolutions. In *In-*  
631 *ternational conference on learning representations*, 2016.
- [Yu *et al.*, 2016] Hsiang-Fu Yu, Nikhil Rao, and Inderjit S Dhillon. Temporal regularized matrix factorization for high-dimensional time series prediction. *Advances in neural information processing systems*, 29, 2016. 632  
[Yu *et al.*, 2018] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, page 3634–3640, 2018. 633  
[Zheng *et al.*, 2020] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1234–1241, 2020. 634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645