



## SPATIO-TEMPORAL FORECASTING

---

Deterministic Graph Neural Networks for Carbon Emissions and Generative  
Probabilistic Stochastic Differential Equation-based Diffusion for Traffic Flow

Prime Supervisor: Prof. Lei CHEN

Co-Supervisor: Prof. Jia LI

Project Mentor: Dr. Jenny, Beijinni LI

Presented by Mingze Gong on August 12, 2024

# OUTLINE

---

## 1 Introduction

Research Overview

## 2 Deterministic Carbon Emissions Modeling

Introduction

Literature Review

Methodology

Results

## 3 Probabilistic Traffic Flow Forecasting

Introduction to ProGen

Literature Review

Preliminaries

Methodology

Results

# RESEARCH OVERVIEW

## Investigations on Deterministic and Probabilistic Forecasting

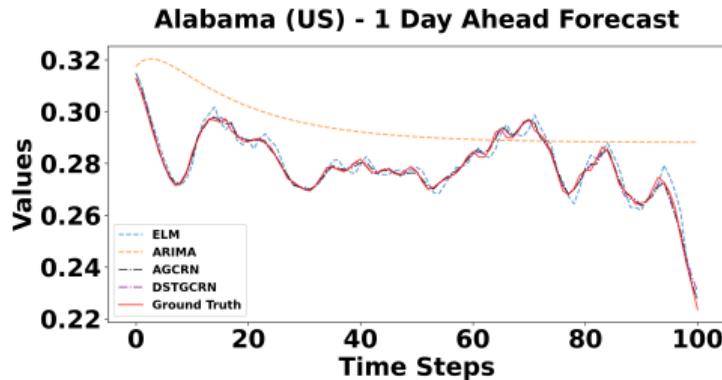


Figure 1: A glimpse of forecasts comparison for deterministic carbon emissions forecasting

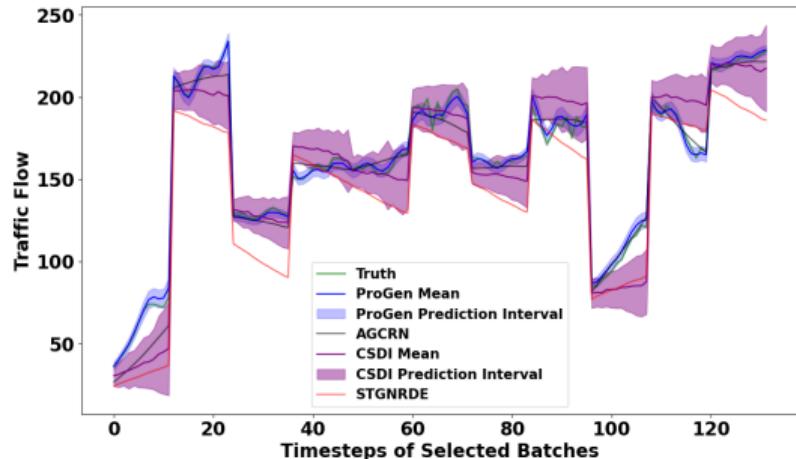


Figure 2: A glimpse of forecasts comparison for probabilistic traffic flow forecasting

# OUTLINE

---

## 1 Introduction

Research Overview

## 2 Deterministic Carbon Emissions Modeling

Introduction

Literature Review

Methodology

Results

## 3 Probabilistic Traffic Flow Forecasting

Introduction to ProGen

Literature Review

Preliminaries

Methodology

Results

# INTRODUCTION

## DSTGCRN: Background and Motivation

### Background:

- » Increase in carbon emissions due to fossil fuels and land degradation.
- » Accurate forecasting is vital to:
  - Inform sustainable policies.
  - Meet reduction targets:
    - China: Peak by 2030.
    - US: Reduce 50-52
    - EU: Cut 55

### Challenges:

- » Models often overlook non-linear trends and regional dependencies.
- » Forecasting difficulties:
  - Variability within regions impacts total emissions.
  - Dynamics between regions significantly influence local emissions.

# INTRODUCTION

## DSTGCRN: Motivations and Contributions

### Motivations:

- » Overcome limitations of traditional methods in handling complex, dynamic environmental data.
- » Leverage insights from the success of Graph Neural Networks (GNNs) in sectors such as traffic and energy to enhance spatial-temporal analysis.

### Contributions:

- » Combines Graph Convolutional Networks (GCN) and Recurrent Neural Networks (RNN) to model evolving inter-regional relationships and spatial-temporal dynamics.
- » Boosts predictive accuracy and offers comprehensive insights to guide environmental policy.
- » Facilitates informed, real-time policy decisions adapted to specific regional contexts.

# LITERATURE REVIEW

## Statistical and Machine Learning Approaches

### » Statistical Methods:

- **ARIMA Models:** Employed for time series forecasting; adjusts for trends and seasonality.
- **Grey Forecasting Models (GM):** Effective under conditions of limited or incomplete data, applicable in emerging markets.
- **Hybrid Models:** Combining GM and ARIMA to address non-linear and non-stationary data, enhancing forecast accuracy.

### » Machine Learning Methods:

- **Deep Learning:** Excels in learning complex data patterns, significantly improving prediction capabilities.
- **Hybrid Approaches:** Integration of neural networks with statistical methods boosts accuracy and reliability.
- **Regional Variability:** Challenges include accommodating diverse environmental conditions, impacting scalability and model performance.

# LITERATURE REVIEW

## Advancements in Spatial-Temporal Predictions

### » **Spatial-Temporal Graph Neural Networks (STGNNs):**

- At the cutting edge for modeling dynamic interdependencies across locations and times, crucial for precise environmental forecasts.
- Adaptive graph structures in these models allow responsiveness to temporal changes, enhancing long-term prediction reliability.

### » **Attention Mechanisms:**

- Prioritize crucial features and time steps, improving focus on significant data and reducing irrelevant noise.
- Demonstrates enhanced detail and accuracy in environmental data analysis, effectively managing spatial and temporal dimensions.

# METHODOLOGY OVERVIEW

## Problem Definitions and Setup

- » **Multisource Time Series Forecasting:** Predict future values from multiple regional sources using observed data on various features like temperature and AQI.

Forecast  $\mathbf{Y}_{t+1}, \dots, \mathbf{Y}_{t+Q}$  using  $f : \mathbb{R}^{N \times P \times C} \rightarrow \mathbb{R}^{N \times Q}$

where  $\mathbf{Y}_t \in \mathbb{R}^N$  is the output vector at time  $t$ ,  $P$  is the number of past time steps, and  $C$  is the number of features.

- » **Regional Carbon Emission Network:** Uses a graph structure to model interdependencies among regions, enhancing prediction accuracy by incorporating spatial dynamics.

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A}), \quad \mathbf{A}_{ij} = \begin{cases} 1 & \text{if regions } i \text{ and } j \text{ are connected,} \\ 0 & \text{otherwise.} \end{cases}$$

where  $\mathcal{V}$  represents the set of nodes (regions),  $\mathcal{E}$  represents the edges (connections between regions), and  $\mathbf{A}$  is the adjacency matrix.

# ADAPTIVE GRAPH CONVOLUTIONAL RECURRENT NETWORK (AGCRN)



## Core Modeling Approach

### » Node Embedding and Graph Convolution:

$$\mathbf{X}'_t = \left( \mathbf{I}_N + \text{softmax} \left( \text{ReLU} \left( \mathbf{E} \cdot \mathbf{E}^\top \right) \right) \right) \mathbf{X}_t \Theta$$

where  $\mathbf{I}_N$  is the identity matrix,  $\Theta$  is the weight matrix.

### » Integration with GRU:

$$\tilde{\mathbf{A}} = \text{softmax}(\text{ReLU}(\mathbf{E} \cdot \mathbf{E}^\top))$$

$$\mathbf{R}_t = \sigma(\tilde{\mathbf{A}}[\mathbf{X}_t, \mathbf{H}_{t-1}] \mathbf{E} \mathbf{W}_r + \mathbf{E} \mathbf{b}_r)$$

$$\mathbf{H}_t = \mathbf{R}_t \odot \mathbf{H}_{t-1} + (1 - \mathbf{R}_t) \odot \hat{\mathbf{H}}_t$$

### » Purpose:

Captures both spatial and temporal dependencies efficiently, integrating graph-based and sequence-based modeling.

# METHODOLOGY DETAIL

## Dynamic Spatial-Temporal Modeling

- » **Dynamic Embeddings:** Generates dynamic node embeddings that adapt over time to capture evolving regional interdependencies. The embedding update mechanism is represented as:

$$\mathbf{E}_t = \text{DynamicEmbedding}(\mathcal{X}_t), \quad \mathcal{X}_t \in \mathbb{R}^{P \times N \times C}$$

where  $\mathcal{X}_t$  denotes the input features over  $P$  past time steps for  $N$  regions with  $C$  features each.

- » **Multihead Attention:** Enhances the model's capability to discern distinct temporal patterns, significantly improving predictive accuracy. The multihead attention mechanism can be defined as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left( \frac{(\mathcal{X}_t'' \mathbf{W}_Q)(\mathcal{X}_t'' \mathbf{W}_K)^\top}{\sqrt{d_e}} \right) \mathcal{X}_t'' \mathbf{W}_V$$

- $\mathbf{Q}, \mathbf{K}, \mathbf{V}$  represent the queries, keys, and values, respectively, each transformed by distinct weight matrices  $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$ .
- $\mathcal{X}_t''$  is the input to the attention layer, and  $d_e$  is the dimension of the embedding space.

# DYNAMIC SPATIAL-TEMPORAL MODELING

## Framework Visualization

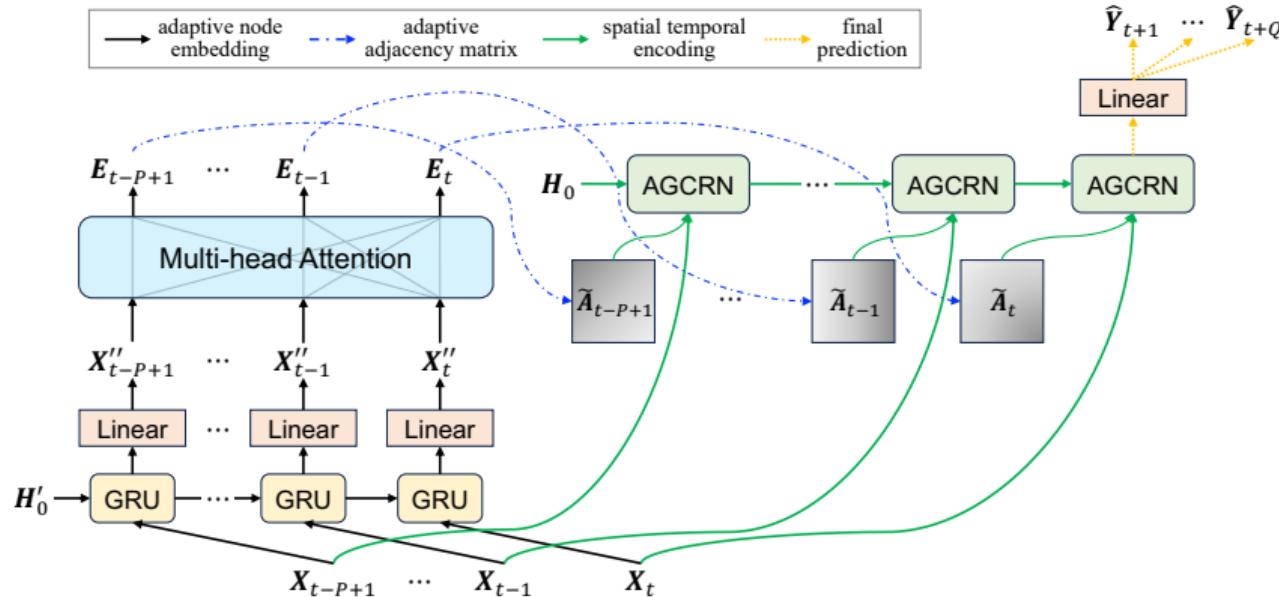


Figure 3: The architecture of the Dynamic Spatial-Temporal Graph Convolutional Recurrent Network (DSTGCRN).

# COMPARISON OF DSTGCRN PERFORMANCE



香港科技大学(广州)  
THE HONG KONG  
UNIVERSITY OF SCIENCE AND  
TECHNOLOGY (GUANGZHOU)

## Performance Across Datasets

Horizon	Model	MAE			MAPE			RMSE			RMSPE			$R^2$		
		China	US	EU	China	US	EU	China	US	EU	China	US	EU	China	US	EU
Q=1	AR/VAR*	.113	.038	.049	12.80%	16.02%	23.71%	.133	.045	.061	15.17%	18.90%	31.63%	-2.434	-3.123	-0.725
	SVR	.036	.011	.021	4.57%	5.22%	5.89%	.044	.013	.028	5.53%	6.22%	7.40%	0.363	0.313	0.743
	ARIMA	.083	.027	.059	9.83%	11.34%	15.29%	.101	.032	.076	11.63%	13.35%	18.68%	-1.656	-1.434	-0.908
	MLP	.045	.018	.032	5.80%	7.02%	11.11%	.070	.030	.069	8.33%	9.41%	15.93%	0.715	0.535	0.453
	FC-LSTM	.075	.023	.031	9.97%	9.50%	11.99%	.112	.040	.064	13.36%	12.17%	17.08%	0.263	0.312	0.426
	ELM	.025	.005	.005	2.36%	1.71%	1.41%	.032	.006	.007	3.19%	2.19%	1.86%	0.772	0.907	0.985
	AGCRN*	.014	.003	.003	1.79%	1.10%	1.08%	.025	.005	.006	2.89%	1.49%	1.51%	0.964	0.989	0.994
	GMAN*	.069	.020	.036	9.76%	10.10%	12.97%	.094	.028	.046	14.08%	14.56%	18.90%	0.211	0.273	0.186
	MISTAGCN*	.062	.018	.027	8.28%	8.74%	9.06%	.092	.029	.058	11.21%	12.41%	13.49%	-0.958	-0.173	0.469
	MTGNN*	.016	.004	.007	4.56%	3.62%	3.71%	.027	.008	.018	16.98%	9.83%	7.16%	-0.411	0.741	0.765
Improvement	DSTGCRN*	<b>.012</b>	<b>.003</b>	<b>.002</b>	<b>1.54%</b>	<b>1.07%</b>	<b>0.78%</b>	<b>.021</b>	<b>.005</b>	<b>.005</b>	<b>2.41%</b>	<b>1.42%</b>	<b>1.11%</b>	<b>0.975</b>	<b>0.991</b>	<b>0.997</b>
	STD	16.1%	7.8%	38.6%	14.0%	3.1%	38.5%	16.7%	13.4%	38.2%	16.7%	4.9%	35.8%	1.2%	0.2%	0.3%
	AR/VAR*	.122	.039	.051	13.82%	15.90%	23.92%	.144	.046	.064	16.40%	18.67%	32.01%	-3.035	-3.251	-0.823
	SVR	.040	.012	.025	5.01%	5.76%	6.96%	.048	.015	.032	6.08%	6.82%	8.67%	0.249	0.176	0.662
Q=3	ARIMA	.085	.026	.056	10.00%	10.68%	14.84%	.103	.030	.072	11.77%	12.57%	18.02%	-1.797	-2.119	-0.745
	MLP	.045	.017	.032	5.85%	6.99%	11.04%	.070	.030	.069	8.37%	9.37%	15.81%	0.708	0.529	0.458
	FC-LSTM	.079	.024	.032	10.39%	9.90%	12.27%	.116	.040	.068	13.99%	12.90%	17.18%	0.223	0.297	0.383
	ELM	.025	.007	.022	2.62%	2.83%	4.84%	.033	.009	.027	3.66%	3.66%	6.17%	0.770	0.794	0.844
	AGCRN*	.016	.004	.003	1.97%	1.58%	1.21%	.027	.008	.007	3.17%	2.19%	1.74%	0.956	0.975	0.992
	GMAN*	.071	.021	.033	9.84%	10.46%	12.29%	.112	.039	.076	14.55%	15.76%	19.50%	0.204	0.247	0.261
	MISTAGCN*	.065	.018	.042	8.76%	8.92%	11.53%	.095	.028	.092	11.96%	12.56%	15.53%	-1.186	-0.141	0.050
	MTGNN*	.022	.007	.011	3.76%	4.84%	8.29%	.038	.015	.023	8.14%	12.42%	17.19%	0.610	0.688	-1.087
Improvement	DSTGCRN*	<b>.009</b>	<b>.004</b>	<b>.003</b>	<b>1.19%</b>	<b>1.48%</b>	<b>1.01%</b>	<b>.016</b>	<b>.006</b>	<b>.006</b>	<b>1.93%</b>	<b>1.98%</b>	<b>1.50%</b>	<b>0.983</b>	<b>0.981</b>	<b>0.995</b>
	STD	40.6%	12.5%	21.7%	39.6%	7.0%	20.3%	40.3%	24.6%	20.8%	39.0%	11.0%	16.4%	2.8%	0.6%	0.3%
	AR/VAR*	.002	.001	.000	.002	.003	.001	.003	.001	.001	.003	.003	.002	.005	.006	.001

Figure 4: Comparison of DSTGCRN with baselines after 5 runs on datasets from China, the US, and the EU. Models marked with an asterisk (\*) also considered temperature and AQI data for China. Experiments used a 7-day lag, predicting 1 and 3 days ahead.

# MODEL FORECAST COMPARISONS

## Long Term Predictions

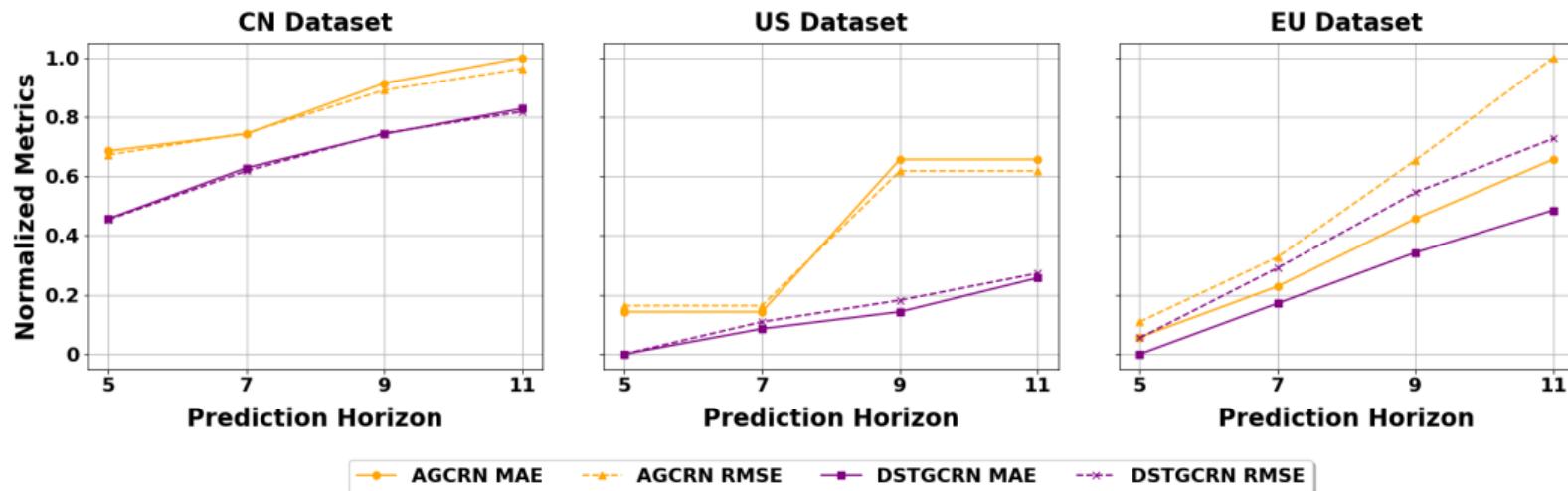


Figure 5: Performance comparison of AGCRN and DSTGCRN across geographies over longer horizons. This figure displays normalized MAE and RMSE metrics, enabling direct comparison on the same scale.

# EMISSION TRENDS AND MODEL COMPARISON

## Visual Analysis of Data and Performance

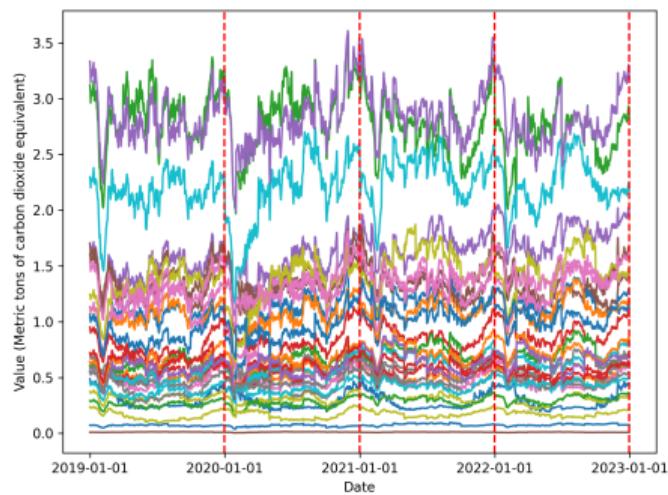


Figure 6: Provincial Carbon Emissions Trends from 2019 to 2022. Each color indicates the emission levels of individual provinces in China. Vertical dashed lines mark the start of each new forecast year.

August 12, 2024

Mingze Gong, DSA Thrust, Information Hub, HKUSTGZ

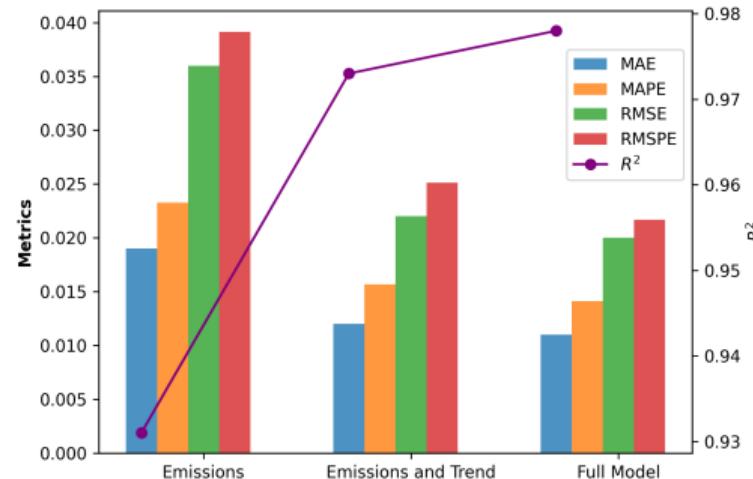


Figure 7: Comparison of model performance metrics across three predictive models.

# ANALYSIS OF DSTGCRN

## Matrix Evolution and Ablation Studies

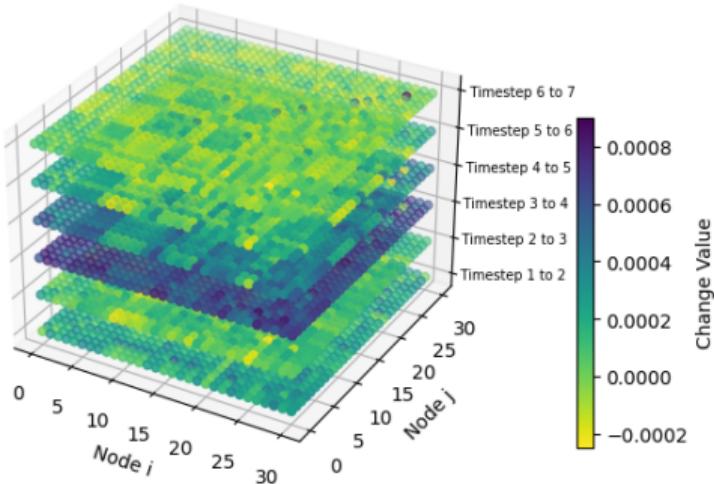


Figure 8: Differential Temporal Adjacency Matrix Evolution.

Model	MAE	MAPE	RMSE	RMSPE	$R^2$
Static	0.014	1.791%	0.025	2.881%	0.964
Time Specific	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
<b>DSTGCRN</b>	<b>0.011</b>	<b>1.410%</b>	<b>0.020</b>	<b>2.169%</b>	0.978

Figure 9: Ablation experiments on DSTGCRN.

# OUTLINE

---

## 1 Introduction

Research Overview

## 2 Deterministic Carbon Emissions Modeling

Introduction

Literature Review

Methodology

Results

## 3 Probabilistic Traffic Flow Forecasting

Introduction to ProGen

Literature Review

Preliminaries

Methodology

Results

# PROGEN FRAMEWORK

## Background and Challenges

### Background:

- » Spatial-temporal data features complex spatial dependencies and dynamic temporal evolution.
- » Traditional forecasting models often provide deterministic predictions that fail to capture inherent data uncertainties.

### Challenges:

- » Difficulty in modeling the complex structures of spatial-temporal interactions.
- » Existing deterministic models do not account for the probabilistic nature of real-world data, limiting their utility for decision-making.

# PROGEN FRAMEWORK

## Motivation and Contributions

### Motivation:

- » The rise of generative AI, particularly diffusion models, provides new tools for handling data uncertainty through probabilistic forecasts.
- » ProGen utilizes stochastic differential equations (SDEs) to model the continuous-time evolution of data, enhancing forecasting accuracy.

### Contributions:

- » **Conceptual:** A novel framework for continuous-time generative modeling for spatial-temporal forecasting.
- » **Technical:** An innovative denoising network and tailored SDE that enhance handling of spatiotemporal correlations.
- » **Empirical:** Demonstrated superiority over existing models through extensive validation on real-world datasets.

# STATE OF THE ART IN PROBABILISTIC AND SPATIO-TEMPORAL FORECASTING

## Diffusion Models, Probabilistic Forecasting, and Spatio-Temporal Methods

### Diffusion Models:

- » Pioneered by Ho et al. (2020) and Song et al. (2021), focusing on generating data from unconditioned noise through discrete and continuous SDEs.
- » Extended to conditional generation for targeted outcomes, incorporating attributes to steer outputs (Nichol and Dhariwal 2021).

### Spatio-Temporal Forecasting:

- » Existing models like AGCRN and DSTAGNN use dynamic graphs for forecasting, while STG-NRDE utilizes neural rough differential equations.
- » Discrete diffusion models have ventured into probabilistic forecasting but lack continuity in time series data handling, contrasting with ProGen's continuous-time approach (Wen et al. 2023).

### Probabilistic Time Series Forecasting:

- » Adaptation of diffusion models for time series forecasting, like TimeGrad and ScoreGrad, faces challenges with

# PROBABILISTIC SPATIO-TEMPORAL FORECASTING

## Problem Setup

We aim to predict future values of a spatio-temporal series using historical data:

- » Let  $\mathcal{D} = \{\mathbf{X}_t\}_{t=1}^T$  be a dataset where  $\mathbf{X}_t \in \mathbb{R}^{N \times D}$  represents observations at time  $t$ , across  $N$  locations and  $D$  features.
- » Spatial dependencies are encoded by a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ , with nodes  $\mathcal{V}$ , edges  $\mathcal{E}$ , and adjacency matrix  $A$ .

## Probabilistic Prediction Task

Estimate the distribution:

$$q_X(\mathbf{X}_{T+1:T+H} \mid \mathbf{X}_{T-L+1:T}, \mathcal{G}, \mathcal{C})$$

Here,  $L$  and  $H$  represent the historical window and forecasting horizon, respectively.

# STOCHASTIC DIFFERENTIAL EQUATIONS

## Mathematical Foundation

SDEs provide a framework for modeling continuous-time stochastic processes:

$$d\mathbf{X} = f(\mathbf{X}, t)dt + g(\mathbf{X}, t)dW, \quad (1)$$

where:

- »  $\mathbf{X} \in \mathbb{R}^d$  represents the state at time  $t$ .
- »  $f(\mathbf{X}, t)$  is the drift function, dictating deterministic dynamics.
- »  $g(\mathbf{X}, t)$  is the diffusion function, modeling stochastic effects.
- »  $dW$  denotes differential Brownian motion.

# REVERSE STOCHASTIC DIFFERENTIAL EQUATIONS



## Retrieving Data from Noisy States

Reverse SDEs describe how to denoise data back to its original distribution:

$$d\mathbf{X} = [f(\mathbf{X}, t) - g^2(\mathbf{X}, t)\nabla_{\mathbf{X}} \log p_t(\mathbf{X})] dt + g(\mathbf{X}, t)d\bar{W}, \quad (2)$$

- » The score function  $\nabla_{\mathbf{X}} \log p_t(\mathbf{X})$  guides the denoising process.
- »  $\bar{W}$  is the reverse Wiener process, introducing reverse dynamics.

# DENOISING SCORE MATCHING (DSM)

## Score-based Generative Modeling

DSM optimizes the match between the gradients of the log probabilities (scores) of the model and data distributions through the diffusion process:

$$\mathcal{L}(\theta) = \mathbb{E}_{t \sim \text{Uniform}(0, K)} \mathbb{E}_{X \sim p_{\text{data}}} [\|\nabla_X \log q_\theta(\mathbf{X}^t | t) - \nabla_X \log p_{\text{data}}(\mathbf{X}^t | t)\|^2] \quad (3)$$

where:

- »  $q_\theta(\mathbf{X}^t | t)$  and  $p_{\text{data}}(\mathbf{X}^t | t)$  are the model and data distributions at diffusion timestep  $t$ , respectively.
- »  $\nabla_X \log$  represents the gradient of the log probability.
- »  $\theta$  denotes the model parameters optimized during training.

# APPLICATION OF DSM AND SDES IN PROGEI

## Practical Implementation and Forecasting Implications

ProGen's implementation of DSM and SDEs offers several key advantages for spatio-temporal forecasting:

- » **Improved Forecast Accuracy:** By accounting for uncertainty and enabling the model to explore a range of possible futures.
- » **Robustness to Noise:** The use of SDEs helps handle the inherent noise and variability in spatio-temporal data effectively.
- » **Flexibility in Model Application:** Suitable for various types of spatio-temporal data beyond just traffic or weather, including economic and biological datasets.

Additionally, the continuous-time approach of ProGen allows for finer temporal resolution in predictions, crucial for dynamic systems monitoring and decision-making.

# OVERVIEW OF PROGEN FRAMEWORK

## Operational Processes

ProGen combines a forward diffusion process with a reverse prediction process:

- » **Forward Diffusion:** Transforms training data into a Gaussian state while training a score model.
  - » **Reverse Prediction:** Iteratively denoises to generate predictions, guided by the score model.

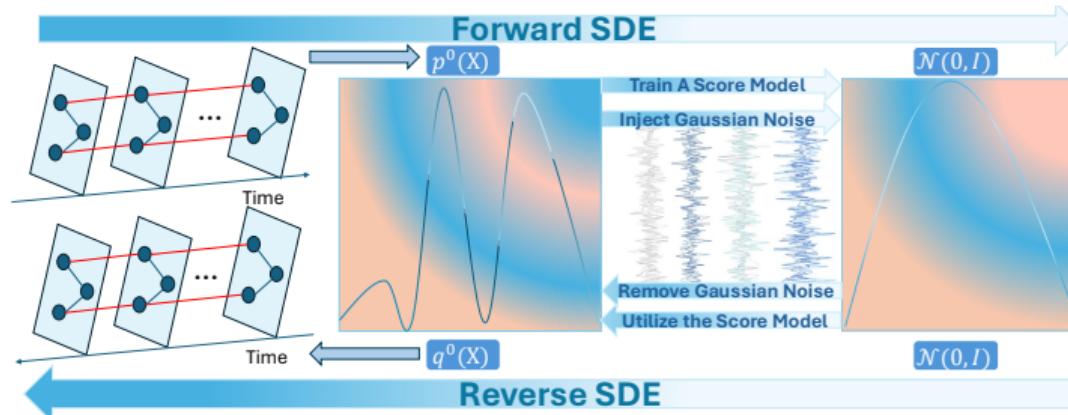


Figure 10: Overview of the two primary processes in ProGen.

# FORWARD DIFFUSION PROCESS

## Transforming Data into Gaussian State

The forward process perturbs the data point by point into Gaussian noise:

$$\tilde{\mathbf{X}}_{\mathbf{F}}^t = \mu(\mathbf{X}_{\mathbf{F}}, t) + \sigma(\mathbf{X}_{\mathbf{F}}, t) \times Z, \quad Z \sim \mathcal{N}(0, I) \quad (4)$$

where  $\mu$  and  $\sigma$  control the mean and standard deviation across discretized time steps.

This process trains the model to understand and simulate the transition from real data distributions to noise.

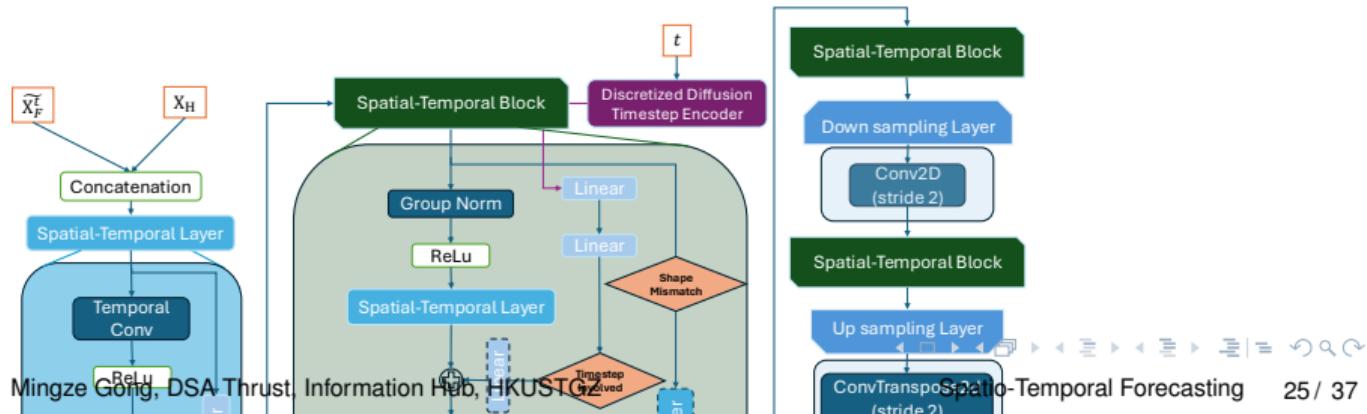
# TRAINING THE DENOISING SCORE MODEL

## Optimizing the Score Estimation

Training focuses on minimizing the discrepancy between the estimated and true data gradients:

$$\mathcal{L}(\theta) = \mathbb{E}_t \left\{ \mathbb{E}_{\mathbf{X}_F, \mathbf{X}_H} \left[ \|\nabla \log p(\tilde{\mathbf{X}}_F^t | \mathbf{X}_F) - s_\theta(\tilde{\mathbf{X}}_F^t, \mathbf{X}_H)\|^2 \right] \right\} \quad (5)$$

This loss function aligns the model's score estimates with the true distribution changes, enhancing prediction accuracy.



# ADAPTIVE REVERSE PREDICTION PROCESS

## Restoring Original Data Distribution

The reverse SDE effectively restores data from its noisy state using learned scores:

$$d\mathbf{X} = (f(\mathbf{X}, t) - g(\mathbf{X}, t)^2 \nabla \log p(\mathbf{X}|t))dt + g(\mathbf{X}, t)d\bar{W} \quad (6)$$

This equation integrates insights from the score model to reverse the diffusion, guiding the data back to its original state.

ProGen's approach modifies traditional methods by adapting to data dynamics more effectively and efficiently.

# OVERVIEW OF PROGEN PERFORMANCE

## Full Test Run

Datasets	Metric	AGCRN	STGCN	DSTAGNN	STGNCDE	STGNRDE	ARIMA	FCLSTM	MTGNN	ASTGCN(r)	ProGen
PEMS03	MAE	15.98	17.48	15.57	15.57	15.50	35.41	21.33	16.46	17.34	<b>15.07</b>
	RMSE	28.25	29.21	27.21	27.09	27.06	47.59	35.11	28.56	29.56	<b>25.09</b>
PEMS08	MAE	15.95	17.13	15.67	15.45	15.32	31.09	23.09	15.71	18.25	<b>14.99</b>
	RMSE	25.22	26.80	24.77	24.81	24.72	44.32	35.17	24.62	28.06	<b>24.00</b>

Figure 12: Distribution and mean values of predictions vs. actual truths across initial batch iterations in PEMS04.

Method	MAE	RMSE	CRPS
Latent ODE	26.05	39.50	0.11
DeepAR	21.56	33.37	0.07
CSDI	32.11	47.40	0.11
TimeGrad	24.46	38.06	0.09
MC Dropout	19.01	29.35	0.07
DiffSTG	18.60	28.20	0.06
ProGen	<b>14.99</b>	<b>24.00</b>	<b>0.05</b>

Figure 13: Training loss curves for different spatiotemporal layers in PEMS08 over 100 epochs.



# OVERVIEW OF PROGEN PERFORMANCE

## Random Test Runs

Datasets	Metric	AGCRN*	DSTAGNN*	STGNRDE*	DeepAR	DiffSTG	CSDI	ProGen
PEMS03	MAE	16.79	17.03	25.40	23.76	53.52	24.52	<b>14.14</b>
	RMSE	29.26	31.05	37.93	37.82	69.62	38.93	<b>22.71</b>
	CRPS	/	/	/	0.11	0.24	0.11	<b>0.07</b>
	MIS	/	/	/	205.78	466.97	245.86	<b>106.50</b>
PEMS04	MAE	21.60	21.25	30.67	29.59	32.55	27.83	<b>18.87</b>
	RMSE	34.32	32.82	44.29	46.01	46.89	42.71	<b>30.01</b>
	CRPS	/	/	/	0.10	0.11	0.10	<b>0.07</b>
	MIS	/	/	/	237.75	209.85	223.77	<b>138.88</b>
PEMS07	MAE	<b>21.50</b>	24.06	33.48	28.77	37.80	30.57	21.91
	RMSE	36.08	39.02	46.49	44.80	50.31	45.92	<b>35.11</b>
	CRPS	/	/	/	0.07	0.09	0.07	<b>0.05</b>
	MIS	/	/	/	242.61	293.45	262.15	<b>186.78</b>
PEMS08	MAE	16.90	15.78	25.10	23.15	44.47	19.00	<b>15.46</b>
	RMSE	26.47	<b>24.36</b>	36.32	35.92	60.72	28.99	24.71
	CRPS	/	/	/	0.08	0.15	0.07	<b>0.05</b>
	MIS	/	/	/	191.20	284.64	146.06	<b>120.53</b>

Figure 14: Distribution and mean values of predictions vs. actual truths across initial batch iterations in PEMS04.

# OVERVIEW OF PROGEN PERFORMANCE

## Random Test Runs

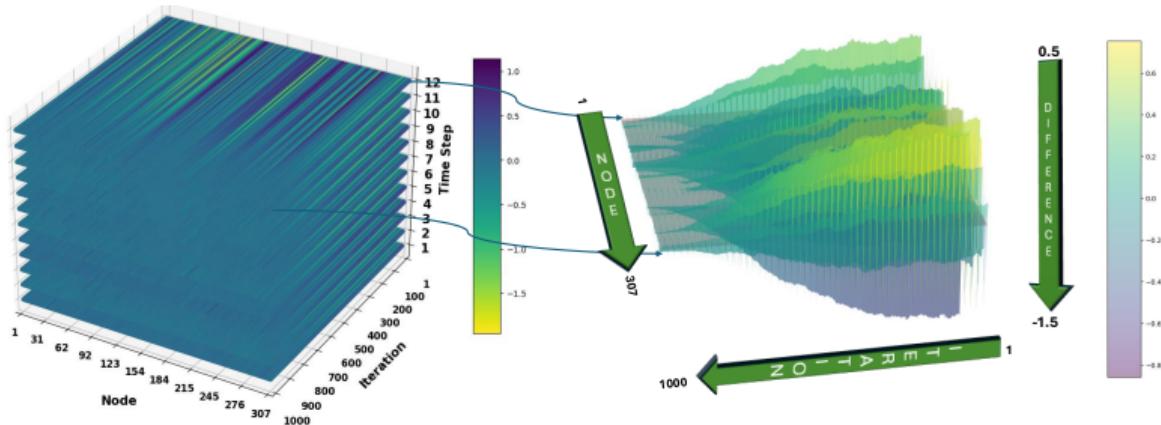


Figure 15: Distribution and mean values of predictions vs. actual truths across initial batch iterations in PEMS04.

# OVERVIEW OF PROGEN PERFORMANCE

Random Test Runs

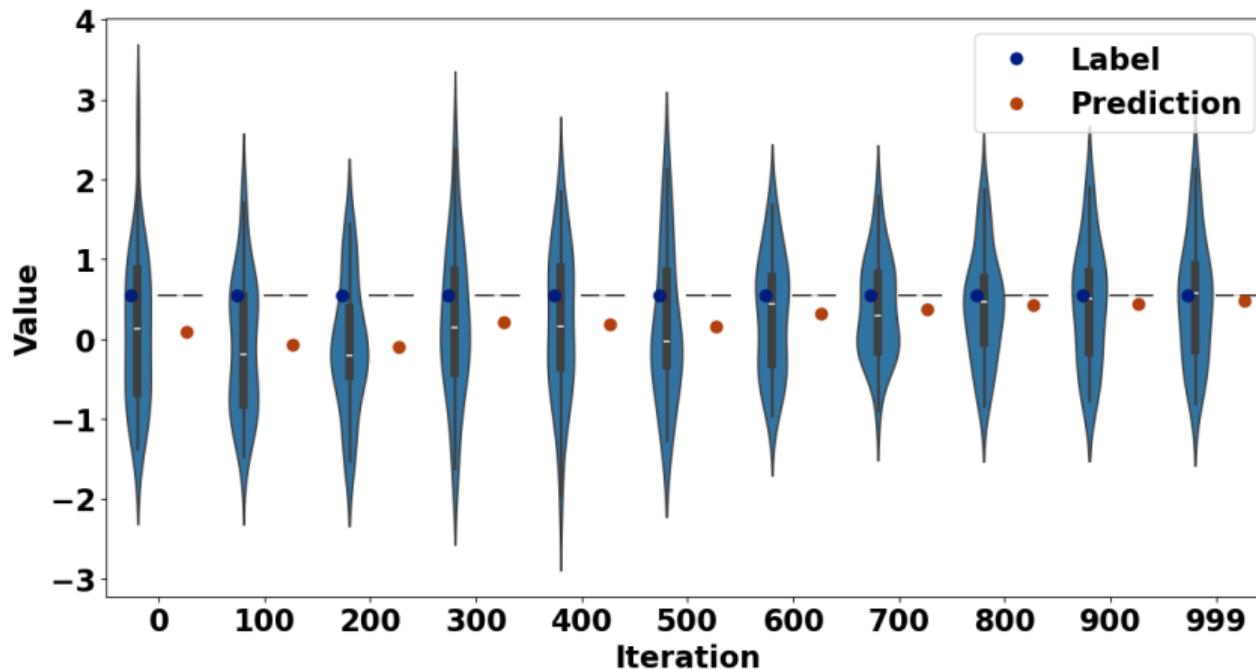


Figure 16: Distribution and mean values of predictions vs. actual truths across initial batch iterations in PEMS04.  
Mingze Gong, DSA Thrust, Information Hub, HKUSTGZ

# OVERVIEW OF PROGEN PERFORMANCE

Random Test Runs

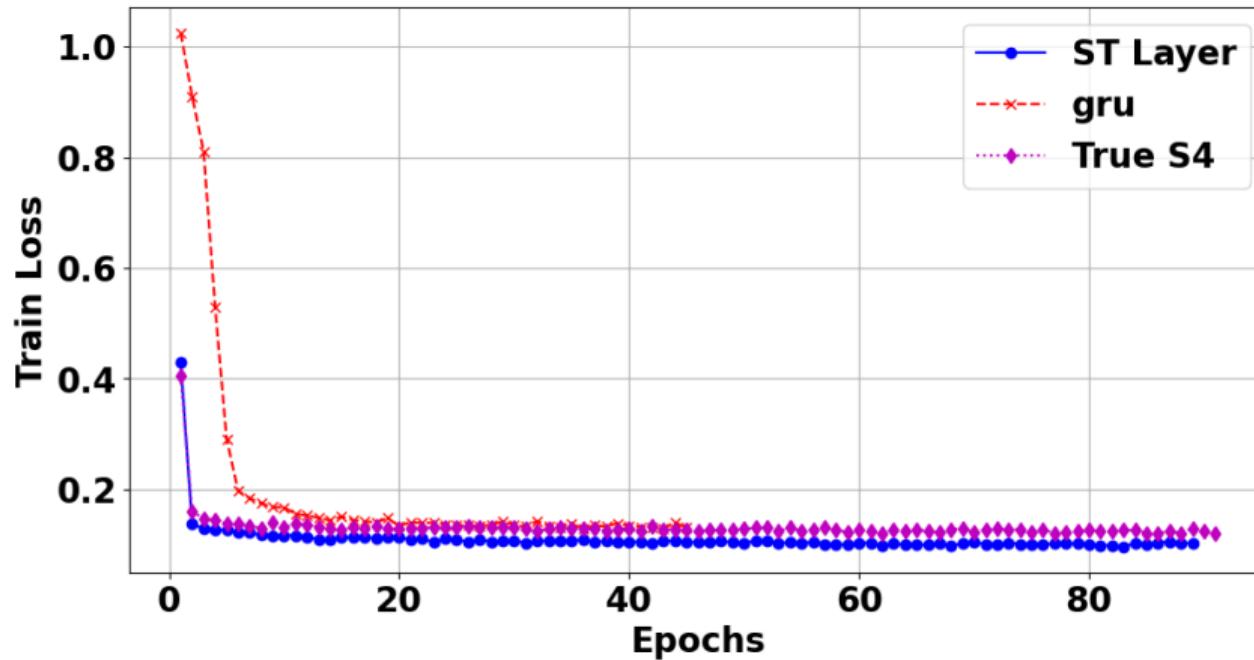
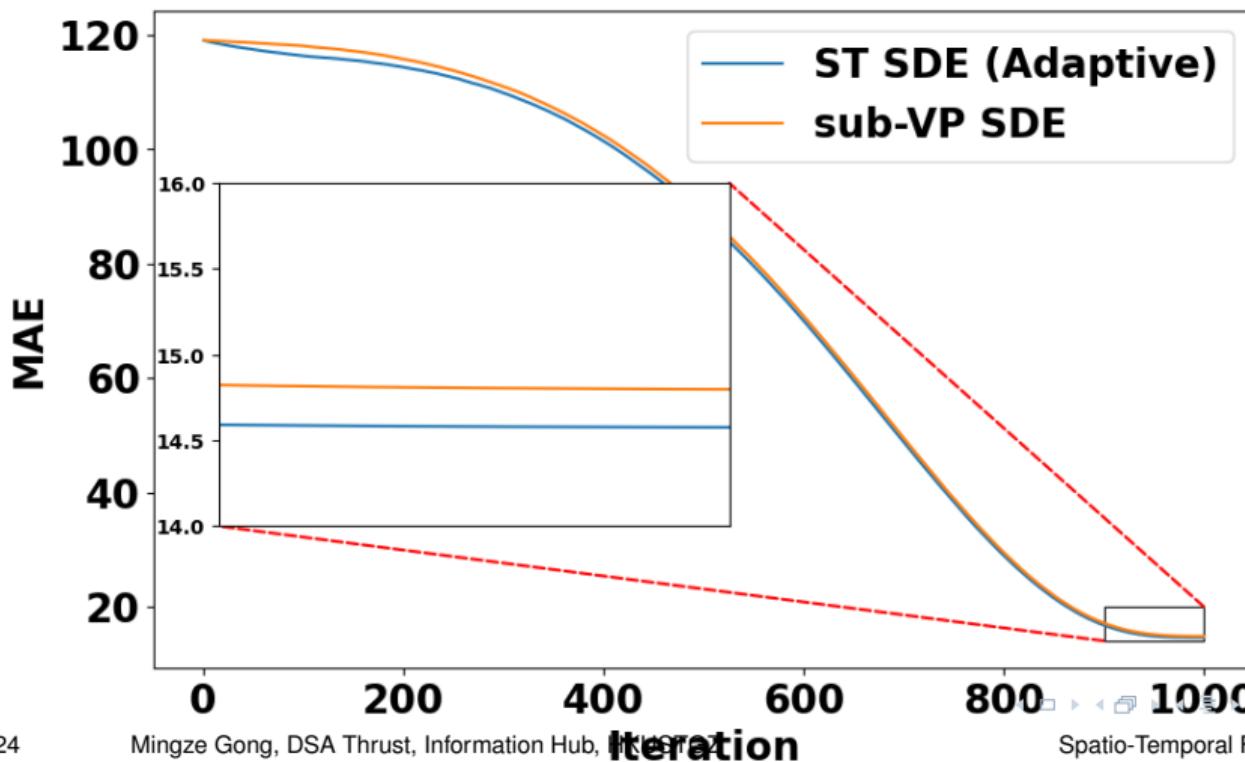


Figure 17: Distribution and mean values of predictions vs. actual truths across initial batch iterations in PEMSO4.

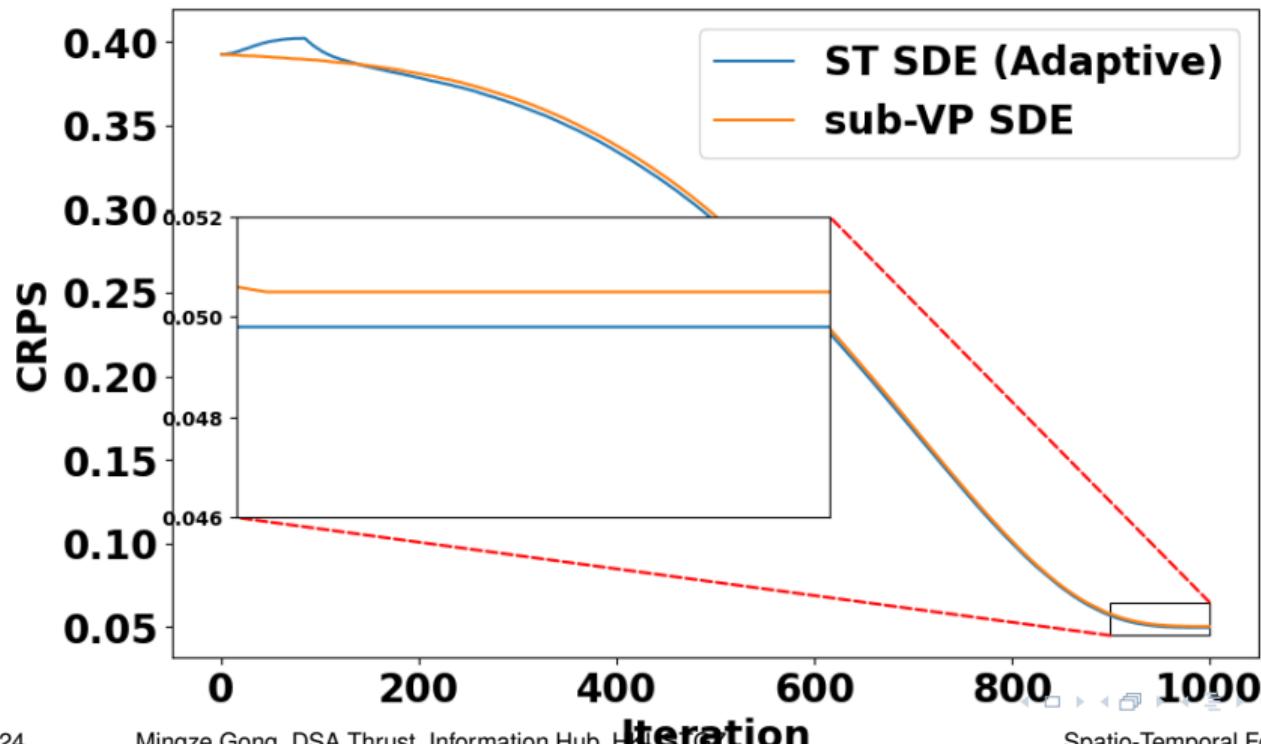
# OVERVIEW OF PROGEN PERFORMANCE

Random Test Runs



# OVERVIEW OF PROGEN PERFORMANCE

Random Test Runs



# OVERVIEW OF PROGEN PERFORMANCE

## Random Test Runs

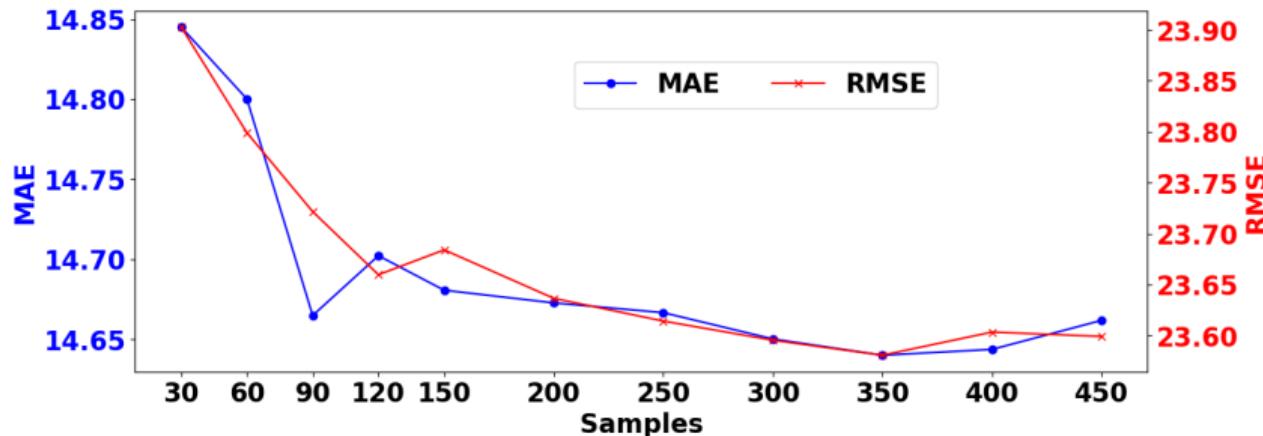


Figure 20: Distribution and mean values of predictions vs. actual truths across initial batch iterations in PEMS04.

# OVERVIEW OF PROGEN PERFORMANCE

## Random Test Runs

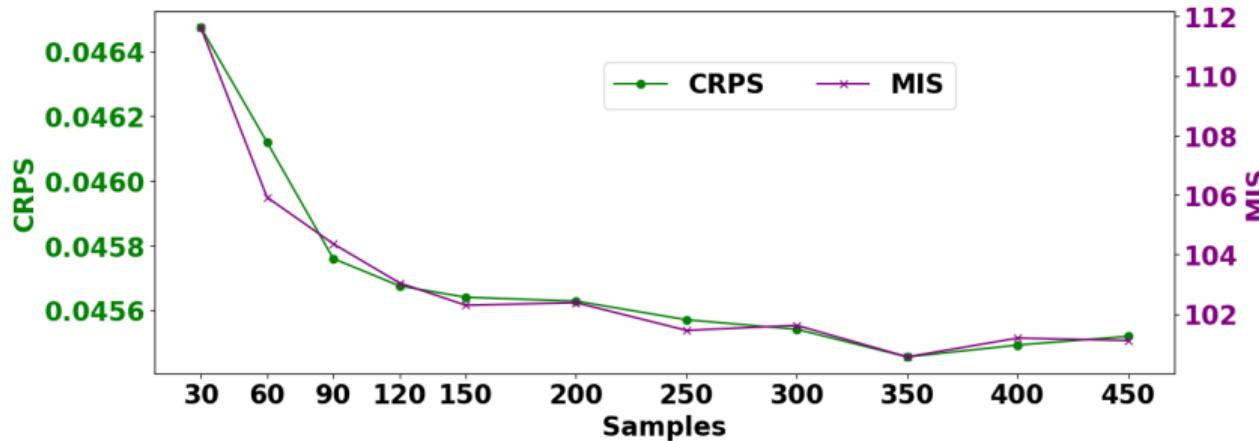


Figure 21: Distribution and mean values of predictions vs. actual truths across initial batch iterations in PEMS04.

# OVERVIEW OF PROGEN PERFORMANCE

## Random Test Runs

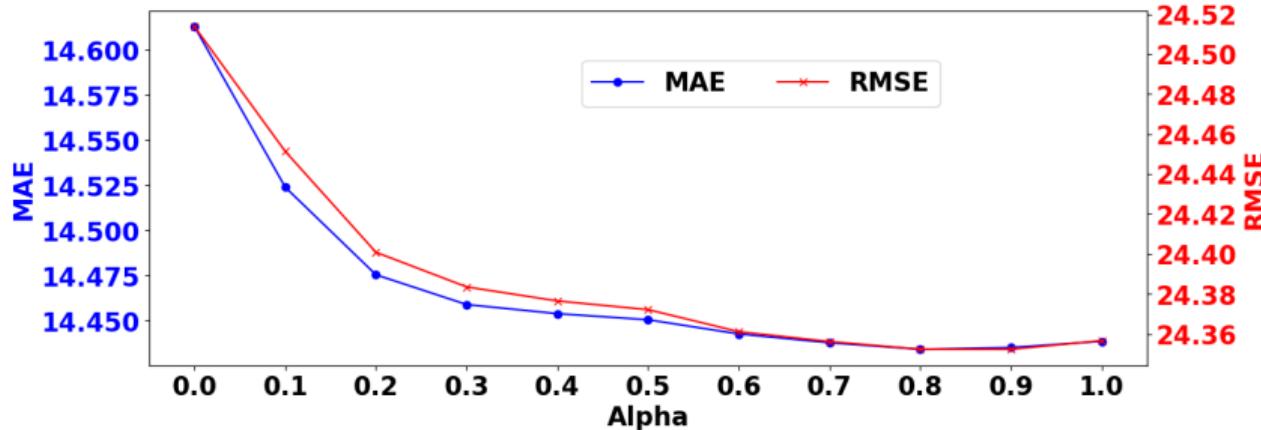


Figure 22: Distribution and mean values of predictions vs. actual truths across initial batch iterations in PEMS04.

# OVERVIEW OF PROGEN PERFORMANCE

## Random Test Runs

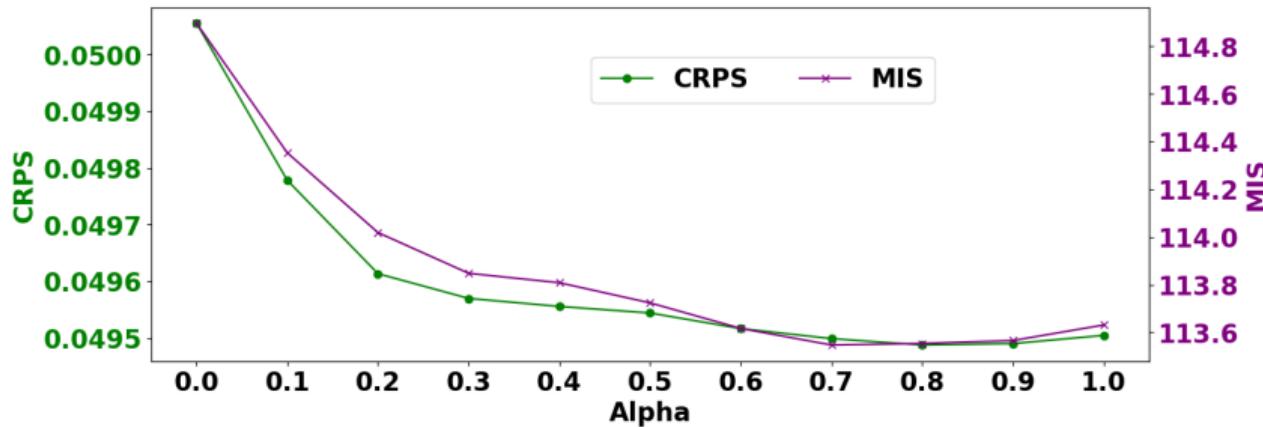


Figure 23: Distribution and mean values of predictions vs. actual truths across initial batch iterations in PEMS04.

## References



香港科技大学(广州)  
THE HONG KONG  
UNIVERSITY OF SCIENCE AND  
TECHNOLOGY (GUANGZHOU)

THANK YOU!