Data Science and Analytics Thrust, Information Hub
The Hong Kong University of Science and Technology (Guangzhou)

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

ACKNOWLEDGEMENTS

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

I would like to express my sincere gratitude to the committee for your dedication and efforts in joining this event. Thank you to my supervisors and mentors for your guidance and help throughout my research journey. I appreciate it all.

RESEARCH OVERVIEW

Investigations on Deterministic and Probabilistic Forecasting

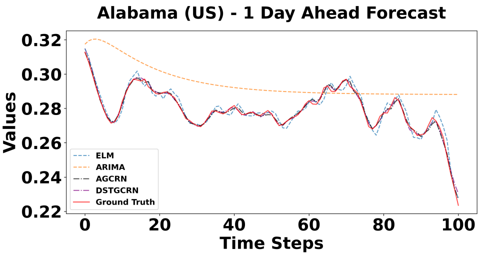

Figure 1: A glimpse of forecasts for deterministic forecasting

Environmental Science.

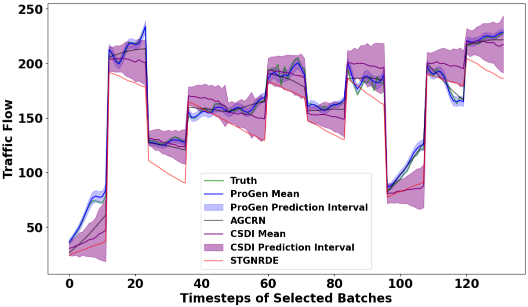

Figure 2: A glimpse of forecasts for probabilistic forecasting

Generative Modelling.

OUTLINE

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

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

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

Problem Definitions and Setup

- **Multisource Time Series Forecasting:** Forecast future values using data from multiple regions on features like temperature and AQI.

$$\text{Forecast } \boldsymbol{Y}_{t+1}, \ldots, \boldsymbol{Y}_{t+Q} \text{ using } f : \mathbb{R}^{N \times P \times C} \to \mathbb{R}^{N \times Q}$$

  Here, $\boldsymbol{Y}_t \in \mathbb{R}^N$ denotes the output vector at time t, P is the number of past time steps considered, and C represents the number of features per step.

- **Regional Carbon Emission Network:** Models interdependencies among regions through a graph structure to enhance predictive accuracy by integrating spatial dynamics.

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{A}), \quad \boldsymbol{A}_{ij} = \begin{cases} 1 & \text{if there is a direct connection between regions } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases}$$

  Here, $\mathcal{V}$ are nodes (regions), $\mathcal{E}$ are edges (connections), and $\boldsymbol{A}$ is the adjacency matrix showing regional connections.

METHODOLOGY

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

AGCRN Bai et al., 2020-Core Modeling Approach

» **Node Embedding and Graph Convolution:**

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

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

» **Integration with GRU:**

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

METHODOLOGY

Dynamic Spatial-Temporal Modeling

- **Dynamic Embeddings:** Generate node embeddings that evolve over time, reflecting changing regional interdependencies. The update mechanism for these embeddings is given by:

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

  where $\boldsymbol{\mathcal{X}}_t$ represents the input features across P past time steps, N regions, and C features.

- **Multihead Attention:** Applies multihead attention to capture distinct temporal patterns, enhancing the model's predictive accuracy. The mechanism is defined as:

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

  where $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ are the queries, keys, and values, respectively, transformed by the weight matrices $\boldsymbol{W}_Q, \boldsymbol{W}_K, \boldsymbol{W}_V$, $\boldsymbol{\mathcal{X}}_t''$ is the processed input, and d_e denotes the embedding dimension.

METHODOLOGY

Framework Visualization

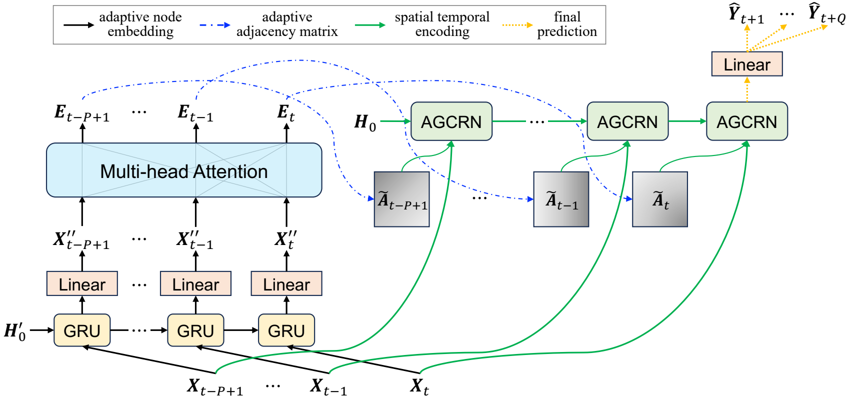

Figure 3: The architecture of the Dynamic Spatial-Temporal Graph Convolutional Recurrent Network (DSTGCRN).

RESULTS

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
	DSTGCRN*	**.012**	**.003**	**.002**	**1.54%**	**1.07%**	**0.78%**	**.021**	**.005**	**.005**	**2.41%**	**1.42%**	**1.11%**	**0.975**	**0.991**	**0.997**
	Improvement	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%
	STD	.001	.000	.000	.001	.001	.001	.001	.000	.001	.002	.001	.002	.004	.001	.001
Q=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
	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
	DSTGCRN*	**.009**	**.004**	**.003**	**1.19%**	**1.48%**	**1.01%**	**.016**	**.006**	**.006**	**1.93%**	**1.98%**	**1.50%**	**0.983**	**0.981**	**0.995**
	Improvement	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%
	STD	.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.

RESULTS

Long Term Predictions

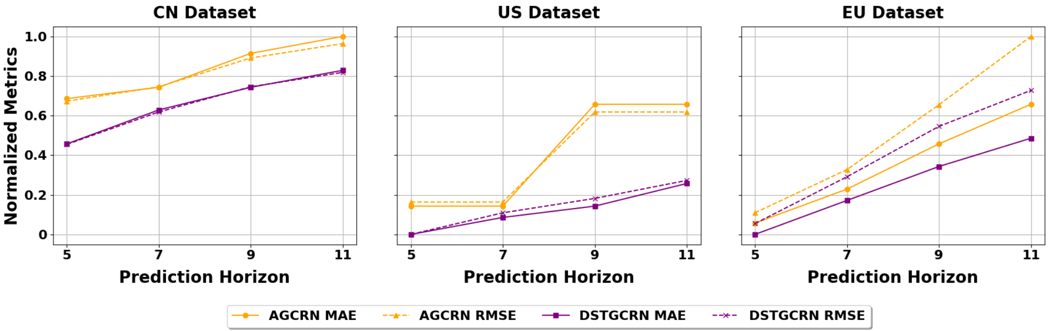

Figure 5: Performance comparison of AGCRN and DSTGCRN across geographies over longer horizons.

RESULTS

Factor Analysis

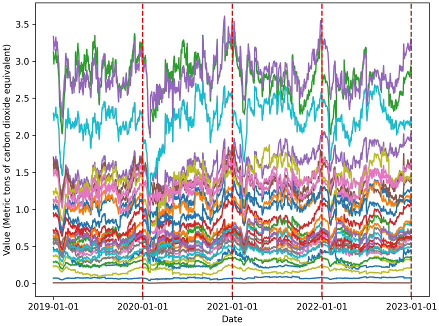

Figure 6: Carbon Emissions Trends from 2019 to 2022 in China.

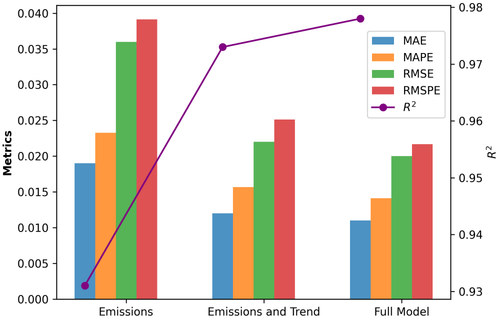

Figure 7: Performance metrics across three scenarios.

RESULTS

Relationship Evolution and Ablation Studies

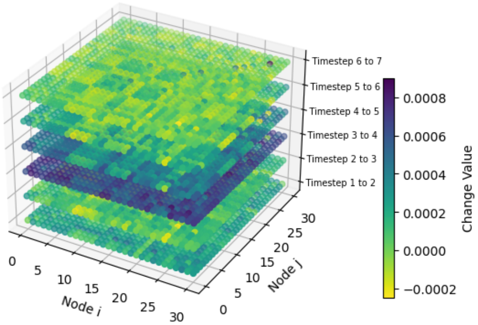

Figure 8: Differential Temporal Adjacency Matrix Evolution.

Model	MAE	MAPE	RMSE	RMSPE	R^2
Static	0.014	1.791%	0.025	2.881%	0.964
Time 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

Figure 9: Ablation experiments on DSTGCRN.

OUTLINE

INTRODUCTION

Background and Challenges

Background:

- Spatial-temporal data exhibits complex spatial dependencies and dynamic temporal evolution.
- Conventional forecasting models often provide deterministic outputs, which do not account for data uncertainties.

Challenges:

- Modeling spatial-temporal interactions is complex due to their intricate structures.
- Traditional deterministic approaches fail to handle the probabilistic nature of real-world data, impairing decision-making.

INTRODUCTION

Motivation and Contributions

Motivation:

- Advances in generative AI, especially diffusion models, offer new avenues for probabilistic forecasting that embrace data uncertainty.
- ProGen leverages stochastic differential equations to model data's continuous-time evolution, improving forecast precision.

Contributions:

- **Conceptual:** Pioneering a generative modeling framework for continuous-time spatial-temporal forecasting.
- **Technical:** Develops a unique denoising approach and custom SDE for improved spatiotemporal correlation handling.
- **Empirical:** Proven enhancements over traditional models validated through rigorous real-world data testing.

INTRODUCTION

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

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.

Diffusion Models:

- Originated by Ho et al. (2020) and Song et al. (2021), using discrete and continuous SDEs to generate data from noise.
- Enhanced for conditional generation, aligning outputs with specific attributes.

Probabilistic Time Series Forecasting:

- Diffusion models adapted for time series, exemplified by TimeGrad and ScoreGrad, though challenges persist with speed and precision.
- ProGen employs a continuous approach, enhancing the capture of spatial and temporal correlations.

LITERATURE REVIEW

Spatio-Temporal Forecasting Methods

Spatio-Temporal Forecasting:

- Techniques such as AGCRN and DSTAGNN use dynamic graphs; STG-NRDE employs neural rough differential equations.

Continuity in Time Series:

- Discrete diffusion models explored probabilistic forecasting but struggled with continuity; ProGen introduces a novel continuous-time method, offering a distinct solution.

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

Problem Setup

Our objective is to forecast future values of a spatio-temporal series based on historical data:

- Define $\mathcal{D} = \{\mathbf{X_t}\}_{t=1}^{T}$ where $\mathbf{X_t} \in \mathbb{R}^{N \times D}$ denotes observations at time t across N locations, each with D features.
- Spatial dependencies are modeled through a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, comprising nodes $\mathcal{V}$, edges $\mathcal{E}$, and an adjacency matrix A.

Probabilistic Prediction Task

Aim to predict the distribution:

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

where L and H denote the length of the historical window and the forecasting horizon, respectively.

METHODOLOGY

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

Stochastic Differential Equations

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

$$d\mathbf{X} = f(\mathbf{X}, t)dt + g(\mathbf{X}, t)dW, \tag{2}$$

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.

METHODOLOGY

Reverse Stochastic Differential Equations

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

$$d\mathbf{X} = \left[f(\mathbf{X}, t) - g^2(\mathbf{X}, t)\nabla_X \log p_t(\mathbf{X})\right] dt + g(\mathbf{X}, t)d\bar{W}, \tag{3}$$

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

METHODOLOGY

Denoising Score Matching (DSM)

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

$$\mathcal{L}(\theta) = \mathbb{E}_{t\sim\text{Uniform}(0,K)}\mathbb{E}_{X\sim p_{\text{data}}}[||\nabla_X \log q_\theta(\mathbf{X^t}|t) - \nabla_X \log p_{\text{data}}(\mathbf{X^t}|t)||^2] \tag{4}$$

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.
- θ denotes the model parameters optimized during training.

Overview of ProGen Framework

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.

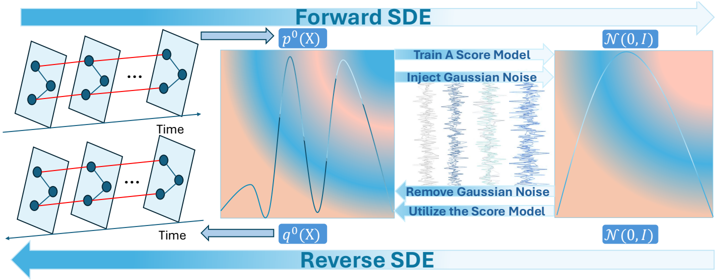

Figure 10: Overview of the two primary processes in ProGen.

METHODOLOGY

Forward Diffusion Process

The forward diffusion process incrementally transforms the data points into Gaussian noise:

$$\tilde{\mathbf{X}}_{\mathbf{F}}^{\mathbf{t}} = \mu(\mathbf{X}_{\mathbf{F}}, t, \beta_0, \beta_1) + \sigma(\mathbf{X}_{\mathbf{F}}, t, \beta_0, \beta_1) \times Z, \tag{5}$$

where μ and σ are functions that determine the mean and standard deviation at each time step t.

This method trains the model to simulate transitions from actual data distributions to a noise state, aiding in understanding the underlying data dynamics.

METHODOLOGY

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

Training the Denoising Score Model

Training is focused on minimizing the discrepancy between the estimated and true gradients of the data:

$$\mathcal{L}(\theta) = \mathbb{E}_t \Big\{ \lambda(t) \mathbb{E}_{\mathbf{X}_F^0, \mathbf{X}_\mathbf{H}} \Big[\mathbb{E}_{\tilde{\mathbf{X}}_F^t | \mathbf{X}_F^0, \mathbf{X}_\mathbf{H}} \left\| s_\theta(\tilde{\mathbf{X}}_F^t, \mathbf{X}_\mathbf{H}, \mathcal{G}, t, \mathbf{P}_\mathbf{H}) - \nabla_{\tilde{\mathbf{X}}_F^t} \log p(\tilde{\mathbf{X}}_F^t | \mathbf{X}_F^0, \mathbf{X}_\mathbf{H}) \right\|_2^2 \Big] \Big\} \tag{6}$$

This loss function aims to align the model's score estimates with actual distribution changes, thereby enhancing prediction accuracy.

METHODOLOGY

Model Architecture

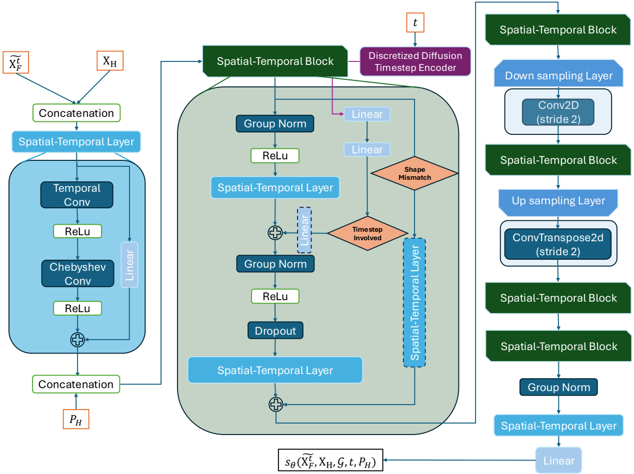

Figure 11: Architecture of the Denoising Score Matching Model in ProGen.

Adaptive Reverse Prediction Process

A new Spatial-Temporal SDE (ST SDE) is introduced to guide the reverse diffusion process:

$$d\mathbf{X} = -\frac{1}{2}\beta(t)\left(\mathbf{X} - \alpha A\mathbf{X}\right)dt + \sqrt{\beta(t)(1 - e^{-2\int_0^t \beta(s)\,ds})}\,dw, \tag{7}$$

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

$$\begin{aligned}\tilde{\mathbf{X}}_{\mathbf{F}}^{t-\frac{1}{K}} = \tilde{\mathbf{X}}_{\mathbf{F}}^{t} - \Bigg[& -\frac{1}{2}\beta(t)\left(\tilde{\mathbf{X}}_{\mathbf{F}}^{t} - \alpha A\tilde{\mathbf{X}}_{\mathbf{F}}^{t}\right)dt \\ & - \left(\sqrt{\beta(t)(1 - e^{-2\int_0^t \beta(s)\,ds})}\right)^2 \times s_\theta\Bigg] \\ & + \sqrt{\beta(t)(1 - e^{-2\int_0^t \beta(s)\,ds})} \cdot dW,\end{aligned} \tag{8}$$

RESULTS

Overview of ProGen Performance on Full Test Runs

Datasets	Metric	AGCRN	STSGCN	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	**15.07**
	RMSE	28.25	29.21	27.21	27.09	27.06	47.59	35.11	28.56	29.56	**25.09**
PEMS08	MAE	15.95	17.13	15.67	15.45	15.32	31.09	23.09	15.71	18.25	**14.99**
	RMSE	25.22	26.80	24.77	24.81	24.72	44.32	35.17	24.62	28.06	**24.00**

Figure 12: Deterministic performance on full PEMS03 and PEMS08 datasets.

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	**14.99**	**24.00**	**0.05**

Figure 13: Probabilistic performance on full PEMS08 dataset.

RESULTS

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

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	**14.14**
	RMSE	29.26	31.05	37.93	37.82	69.62	38.93	**22.71**
	CRPS	\	\	\	0.11	0.24	0.11	**0.07**
	MIS	\	\	\	205.78	466.97	245.86	**106.50**
PEMS04	MAE	21.60	21.25	30.67	29.59	32.55	27.83	**18.87**
	RMSE	34.32	32.82	44.29	46.01	46.89	42.71	**30.01**
	CRPS	\	\	\	0.10	0.11	0.10	**0.07**
	MIS	\	\	\	237.75	209.85	223.77	**138.88**
PEMS07	MAE	**21.50**	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	**35.11**
	CRPS	\	\	\	0.07	0.09	0.07	**0.05**
	MIS	\	\	\	242.61	293.45	262.15	**186.78**
PEMS08	MAE	16.90	15.78	25.10	23.15	44.47	19.00	**15.46**
	RMSE	26.47	**24.36**	36.32	35.92	60.72	28.99	24.71
	CRPS	\	\	\	0.08	0.15	0.07	**0.05**
	MIS	\	\	\	191.20	284.64	146.06	**120.53**

Figure 14: Overall performance on all datasets.

RESULTS

Generating Process Visualization

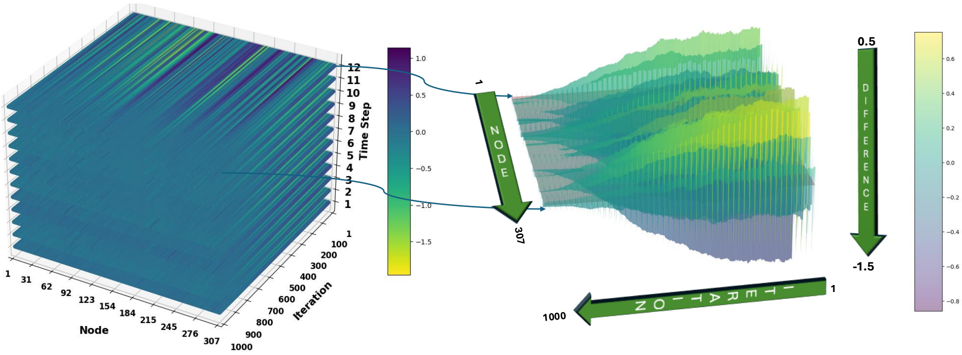

Figure 15: Visualization of the difference between the mean predictions and the actual values for the average test data in PEMS04.

RESULTS

Generating Process Visualization

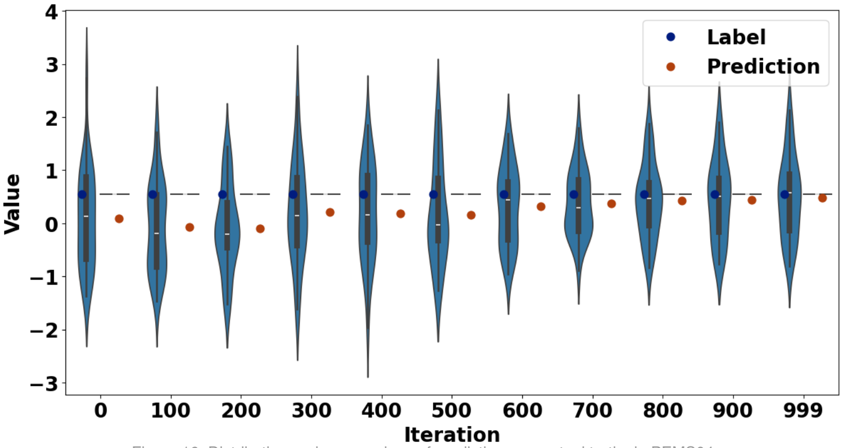

Figure 16: Distribution and mean values of predictions vs. actual truths in PEMS04.

RESULTS

Ablation Studies

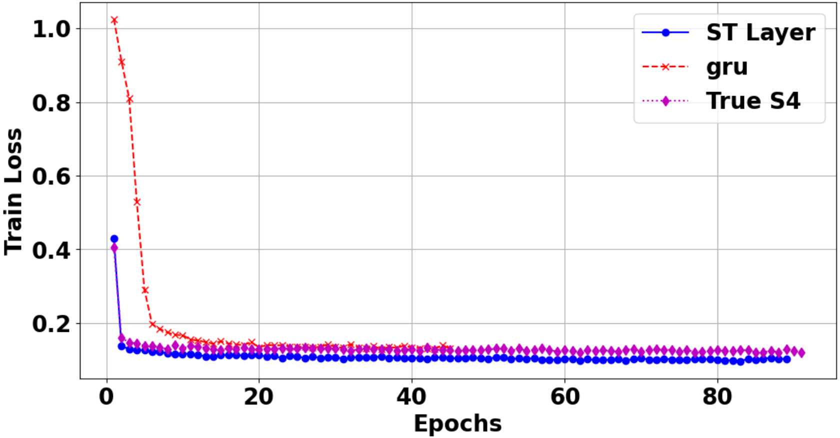

Figure 17: Training loss curves for different spatiotemporal layers in PEMS08.

RESULTS

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

Ablation Studies

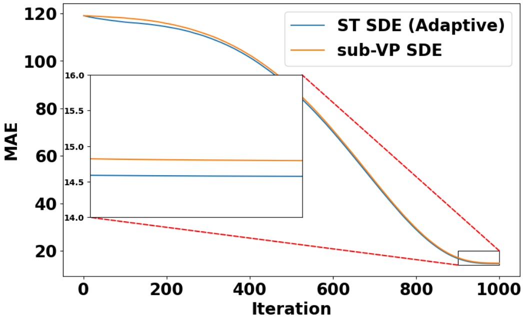

Figure 18: MAE comparison between the tailored spatial-temporal SDE and the subVP-SDE in PEMS08.

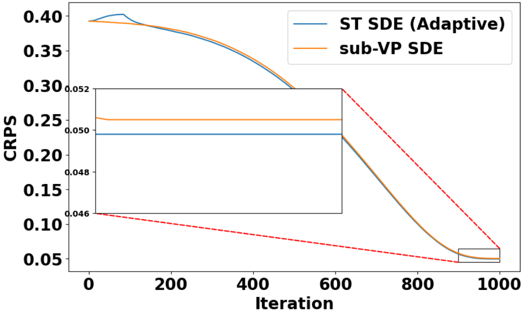

Figure 19: CRPS comparison between the tailored spatial-temporal SDE and the subVP-SDE in PEMS08.

RESULTS

Ablation Studies

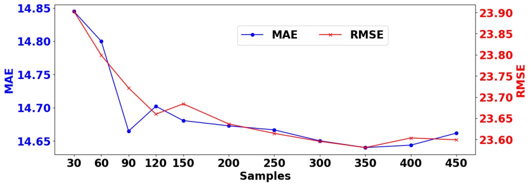

Figure 20: MAE and RMSE across initial batch iterations in PEMS04.

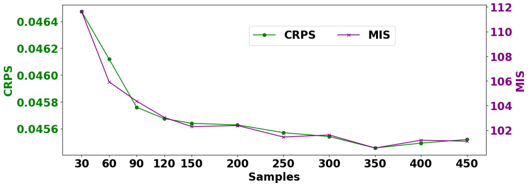

Figure 21: CRPS and MIS across initial batch iterations in PEMS04.

RESULTS

Ablation Studies

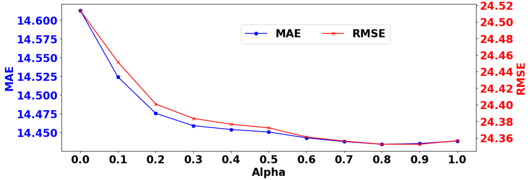

Figure 22: MAE and RMSE across initial batch iterations in PEMS04.

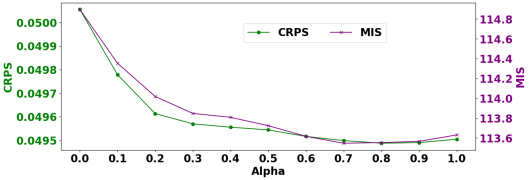

Figure 23: CRPS and Miss Rate across initial batch iterations in PEMS04.

References

Bai, L., Yao, L., Li, C., Wang, X., & Wang, C. (2020). Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting. Retrieved July 29, 2024, from http://arxiv.org/abs/2007.02842

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *NeurIPS*. Retrieved April 19, 2024, from http://arxiv.org/abs/2006.11239

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2021). Score-Based Generative Modeling through Stochastic Differential Equations. *ICLR*.

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

THANK YOU!