SDA-GRIN for Adaptive Spatial-Temporal Multivariate Time Series Imputation

Amir Eskandari, *Aman Anand, Drishti Sharma, Farhana Zulkernine*

School of Computing, Queen's University, Kingston, ON, Canada | Connected Minds CFREF

July 8, 2025

CONNECTED MINDS
Neural and Machine Systems
for a Healthy, Just Society

Queen's UNIVERSITY

Introduction & Problem

Multivariate Time Series (MTS) Data:

- Prevalent in diverse real-world domains:
  - Healthcare (patient vitals)
  - Geoscience (environmental sensors)
  - IoT (interconnected device data)
  - Traffic Management (sensor readings)

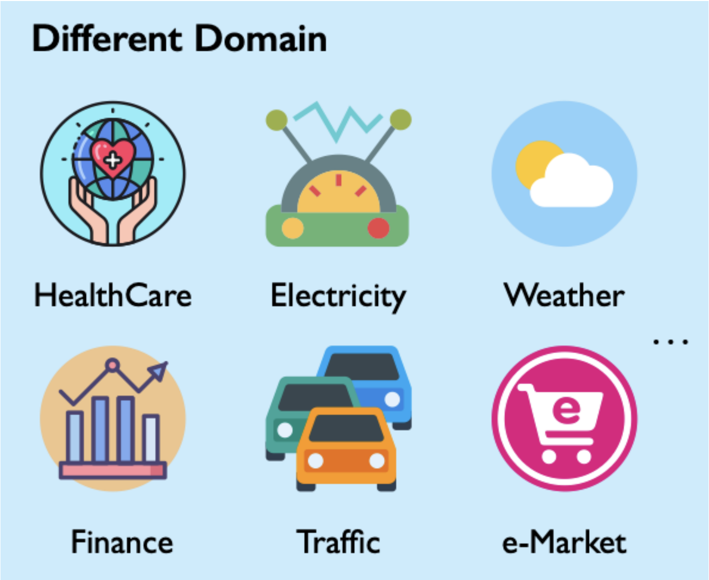


Figure source: Wang, Yuxuan, et al. "Deep time series models: A comprehensive survey and benchmark."

Missing Data is a Big Challenge:

- MTS data often suffers from missing values.
- Caused by various factors (e.g., sensor malfunctions, network issues).
- Significantly disrupts systems that rely on this data, leading to inaccuracies or failures.

Standard Architectures for MTS Imputation

Leveraging Dependencies:

- Established MTS imputation methods typically exploit both temporal and spatial dependencies within the data.

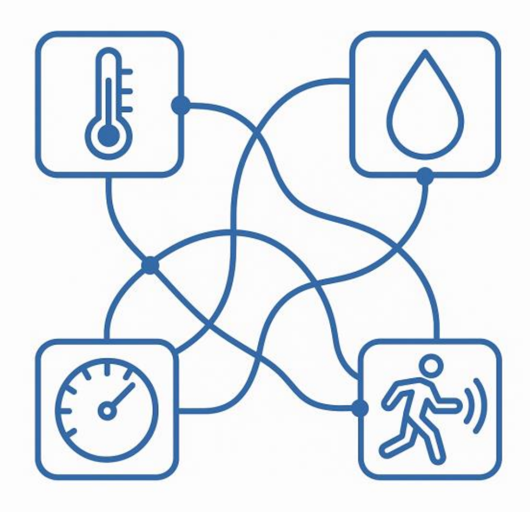


Figure source: Chatbots OpenAI

Temporal Dependency Learning:

- **Recurrent Neural Networks (RNNs):** Architectures like LSTMs and GRUs are widely used.

Spatial Relationship Learning:

- **Graph-based Models (e.g., GNNs):** Such as Graph Convolutional Networks (GCNs) or Message Passing Neural Networks (MPNNs).

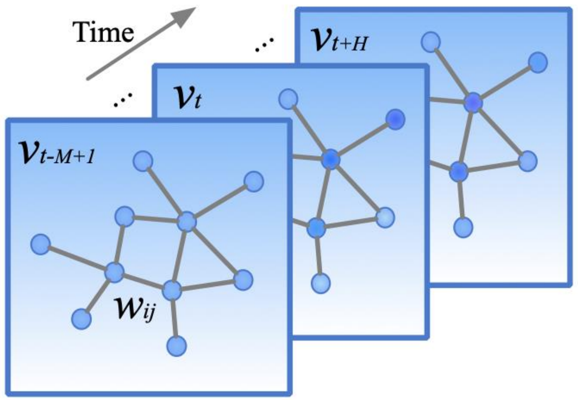


Figure source: Yu, Bing, Haoteng Yin, and Zhanxing Zhu. "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting."

The Core Challenge: Dynamic Spatial Dependencies

Static Approaches Have Limitations:

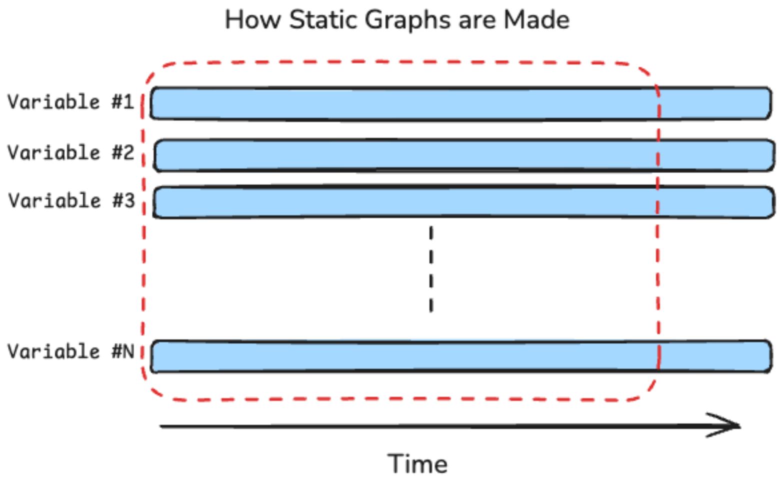
- Existing models assume fixed spatial relationships.
- Static graphs (top figure) are derived from the entire training data, leading to an averaged representation that cannot adapt to new or evolving relationships in test data, thus missing dynamic nuances.

Reality: Relationships are Dynamic:

- Real-world data (bottom figure) shows constantly evolving spatial relationships.
- These changes occur due to various reasons, like time-of-day or events.

The Mismatch Problem:

- A static graph cannot capture fluid dependencies.
- Leads to sub-optimal imputation performance.



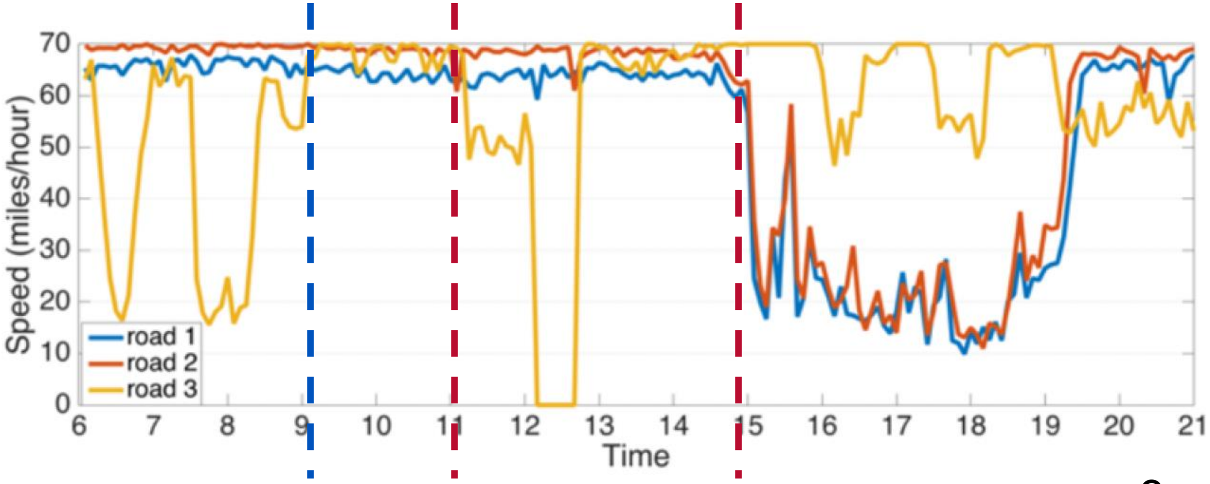


Figure source: Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting

Our Contributions

Introducing SDA-GRIN: A Spatial Dynamic Aware Graph Recurrent Imputation Network.

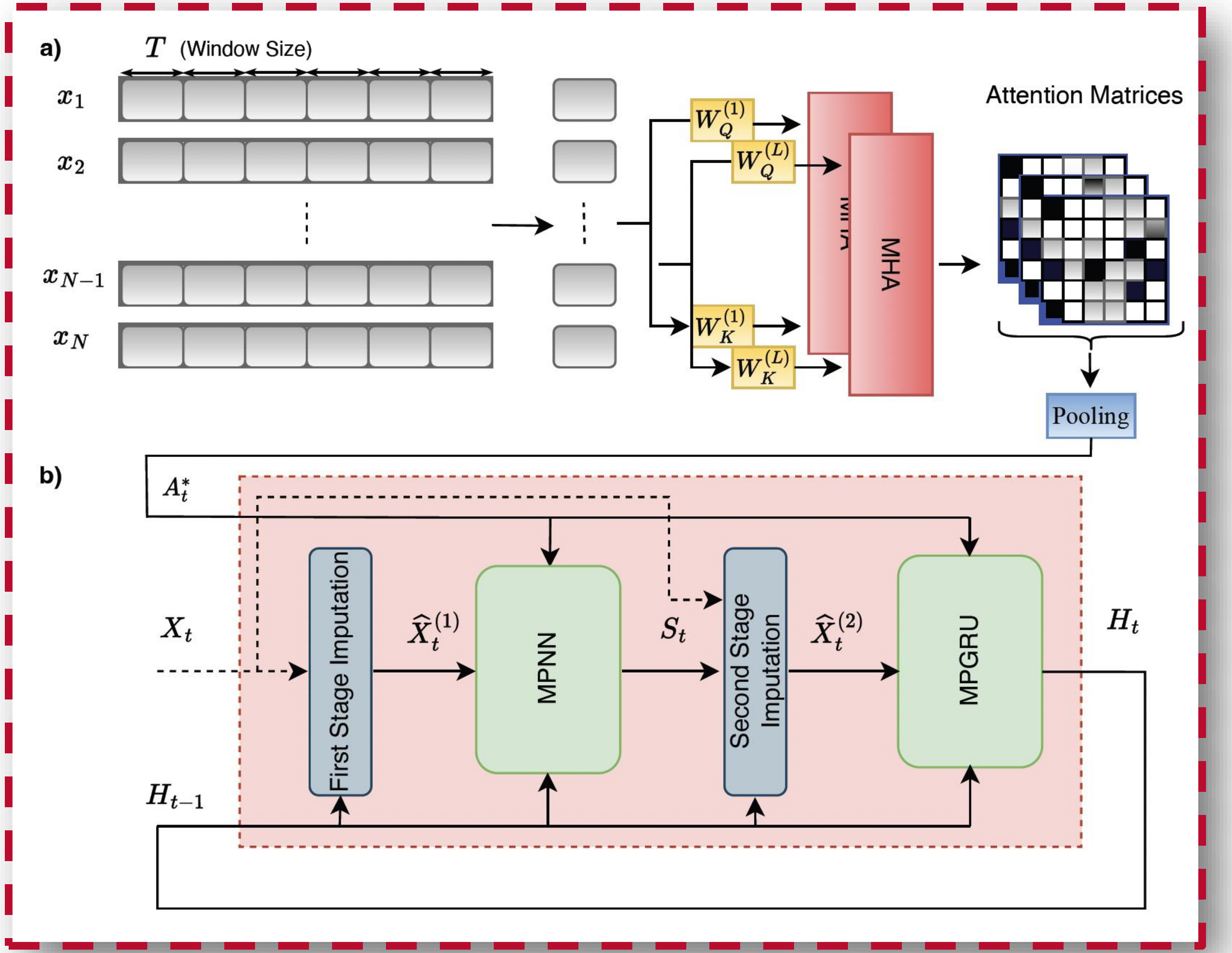
Key Innovation: Enhances imputation by adaptively capturing dynamic spatial dependency changes over time.

- This directly tackles the static graph limitation.

Demonstrated Superior Performance:

- Significant improvements on four real-world datasets (AQI, AQI-36, PEMS-BAY, METR-LA).
- Outperforms 9 previous state-of-the-art methods (MEAN, KNN, MICE, MF, VAR, BRITS, rGAIN, MPGRU, GRIN).

In-depth Analysis: Ablation study provides insights into the impact of missing data rate and window size on model performance.



SDA-GRIN: Spatial Dependency Awareness (MHA)

Dynamic Graph Adaptation with MHA:

- SDA-GRIN's core: learn and adapt relationships among variables in real-time.

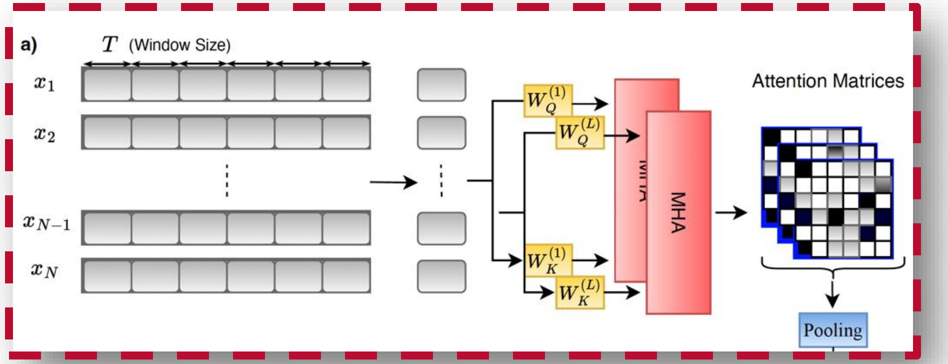
  1. MHA computes query (Q) and key (K) matrices from input X_t (as shown in the figure):

$$Q_t^{(l)} = X_t W_Q^{(l)} \quad K_t^{(l)} = X_t W_K^{(l)}$$

  ($W_Q^{(l)}$, $W_K^{(l)}$ are learnable weights for each head l.)

  2. Attention weights ($A_t^{(l)}$) between variables are calculated:

$$\hat{A}_t^{(l)} = \mathrm{softmax}\left(\frac{Q_t^{(l)} K_t^{(l)T}}{\sqrt{d}}\right)$$



  3. Pooled attention then adapts the static graph to a time-varying A_t^*.

Benefit: Provides highly relevant spatial context for improved imputation.

SDA-GRIN: Imputation Framework

SDA-GRIN uses a multi-stage recurrent process to impute missing values by integrating temporal context with dynamic spatial features.

Step 1: First-Stage Imputation: An initial imputation is performed using the hidden state (H_{t-1}) from the previous time step.

$$\hat{X}_t^{(1)} = \Phi(H_{t-1}V_h + b_h)$$

Step 2: Spatial Feature Extraction: Spatial features (S_t) are extracted via a Message-Passing Neural Network (MPNN) that operates on the dynamic graph (A_t^*) to model inter-variable dependencies.

$$S_t = MPNN(\hat{X}_t^{(1)}, M_t, H_{t-1}, A_t^*)$$

Step 3: Second-Stage Imputation: The imputation is refined by combining the new spatial features (S_t) with the previous hidden state (H_{t-1}) for a richer context.

$$\hat{X}_t^{(2)} = \Phi([S_t||H_{t-1}]V_s + b_s)$$

Step 4: Recurrent State Update: The hidden state is updated to (H_t) using a Message-Passing GRU (MPGRU), preparing the model for the next time step.

$$H_t = MPGRU(\hat{X}_t^{(2)}, M_t, H_{t-1}, A_t^*)$$

SDA-GRIN: Experimental Setup

To validate the performance of SDA-GRIN, a comprehensive experimental design was established using public benchmarks, standard evaluation metrics, and a suite of baseline models.

- **Datasets:** Experiments were conducted on four challenging real-world datasets from two different domains:
  - **Air Quality (AQ):**
    - **AQI:** Data from 437 monitoring stations across 43 cities in China.
    - **AQI-36:** A smaller version of AQI with 36 sensors.
  - **Traffic:**
    - **PEMS-BAY:** Data from 325 traffic sensors in the San Francisco Bay Area.
    - **METR-LA:** Data from 207 sensors on Los Angeles County highways.

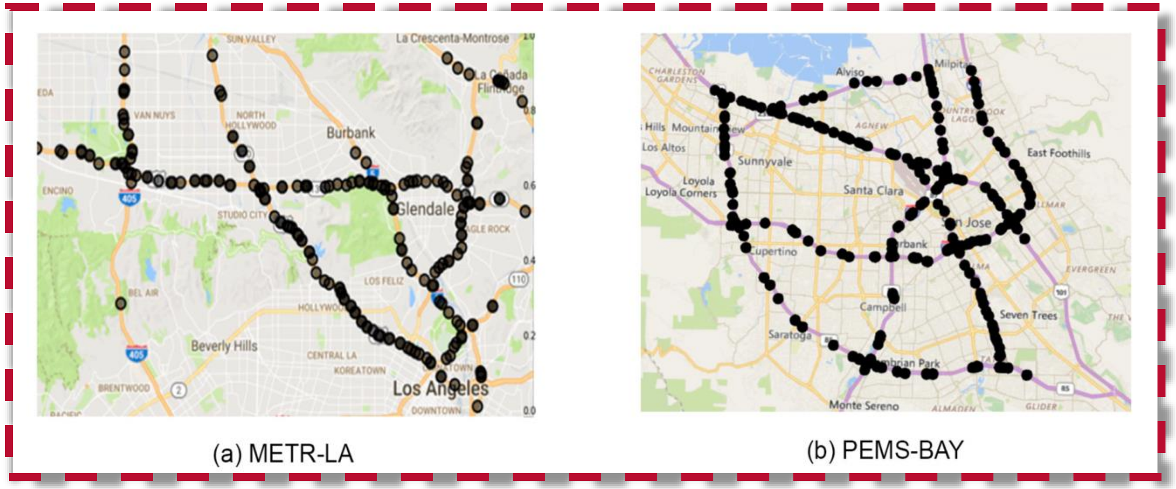


Figure source: Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting.

- **Baselines:** SDA-GRIN's performance was compared against nine established baselines:
  - **Traditional Methods:** MEAN, K-Nearest Neighbors (KNN), MICE, and Matrix Factorization (MF), Vector Auto-Regressive (VAR),
  - **Deep Learning Methods**: BRITS, rGAIN, MPGRU, and GRIN

SDA-GRIN: Main Results

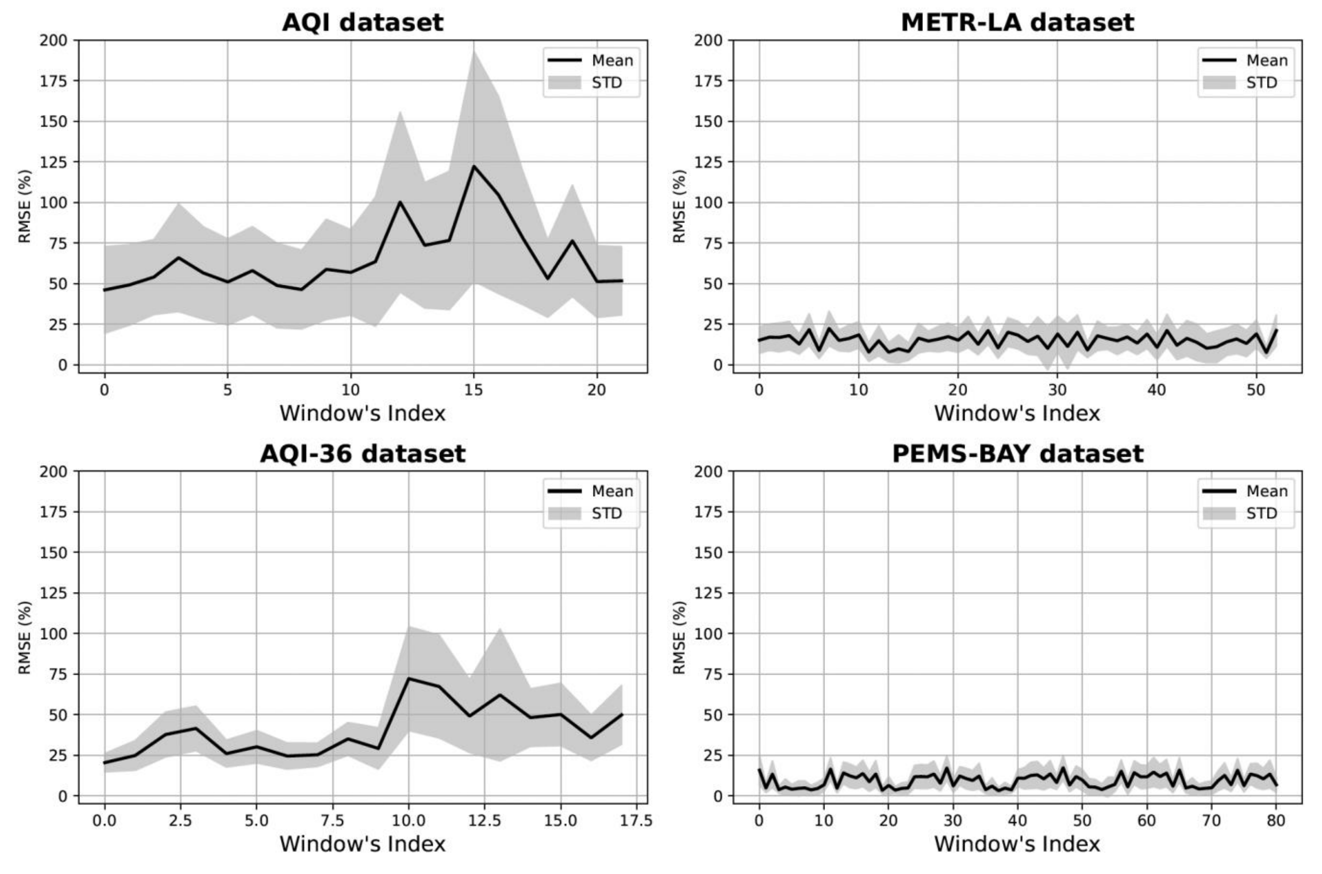
SDA-GRIN significantly outperforms nine baselines, especially on datasets with dynamic spatial patterns. It reduced MSE by 9.51% (AQI), 9.40% (AQI-36), and 1.94% (PEMS-BAY), while achieving competitive performance on METR-LA, matching GRIN.

Dataset	Method	MAE ↓	MSE ↓	MRE (%) ↓
PEMS-BAY	Mean	5.42 ± 0.00	86.59 ± 0.00	8.67 ± 0.00
	KNN	4.30 ± 0.00	49.80 ± 0.00	6.88 ± 0.00
	MF	3.29 ± 0.01	51.39 ± 0.64	5.27 ± 0.02
	MICE	3.09 ± 0.02	31.43 ± 0.41	4.95 ± 0.02
	VAR	1.30 ± 0.00	6.52 ± 0.01	2.07 ± 0.01
	rGAIN	1.88 ± 0.02	10.37 ± 0.20	3.01 ± 0.04
	BRITS	1.47 ± 0.00	7.94 ± 0.03	2.36 ± 0.00
	MPGRU	1.11 ± 0.00	7.59 ± 0.02	1.77 ± 0.00
	GRIN	<u>0.67</u> ± 0.00	<u>1.55</u> ± 0.01	<u>1.08</u> ± 0.00
	SDA-GRIN	**0.66** ± 0.00	**1.52** ± 0.01	**1.07** ± 0.00
METR-LA	Mean	7.56 ± 0.00	142.22 ± 0.00	13.10 ± 0.00
	KNN	7.88 ± 0.00	129.29 ± 0.00	13.65 ± 0.00
	MF	5.56 ± 0.03	113.46 ± 1.08	9.62 ± 0.05
	VAR	2.69 ± 0.00	21.10 ± 0.02	4.66 ± 0.00
	MICE	4.42 ± 0.07	55.07 ± 1.46	7.65 ± 0.12
	BRITS	2.34 ± 0.00	16.46 ± 0.05	4.05 ± 0.00
	rGAIN	2.83 ± 0.01	20.03 ± 0.09	4.91 ± 0.01
	MPGRU	2.44 ± 0.00	22.17 ± 0.03	4.22 ± 0.00
	GRIN	**1.91** ± 0.00	**10.41** ± 0.03	**3.30** ± 0.00
	SDA-GRIN	**1.91** ± 0.00	<u>10.48</u> ± 0.04	**3.30** ± 0.01

Dataset	Method	MAE ↓	MSE ↓	MRE (%) ↓
AQI	Mean	39.60 ± 0.00	3231.04 ± 0.00	59.25 ± 0.00
	KNN	34.10 ± 0.00	3471.14 ± 0.00	51.02 ± 0.00
	VAR	22.95 ± 0.30	1402.84 ± 52.63	33.99 ± 0.44
	BRITS	20.21 ± 0.22	1157.89 ± 25.66	29.94 ± 0.33
	rGAIN	21.78 ± 0.50	1274.93 ± 60.28	32.26 ± 0.75
	MPGRU	18.76 ± 0.11	1194.35 ± 15.23	27.79 ± 0.16
	GRIN	<u>14.73</u> ± 0.15	<u>775.91</u> ± 28.49	<u>21.82</u> ± 0.23
	SDA-GRIN	**14.43** ± 0.28	**702.12** ± 22.82	**21.59** ± 0.42
AQI-36	Mean	53.48 ± 0.00	4578.08 ± 0.00	76.77 ± 0.00
	KNN	30.21 ± 0.00	2892.31 ± 0.00	43.36 ± 0.00
	BRITS	14.50 ± 0.35	662.36 ± 65.16	20.41 ± 0.50
	rGAIN	15.37 ± 0.26	641.92 ± 33.89	21.63 ± 0.36
	VAR	15.64 ± 0.08	833.46 ± 13.85	22.02 ± 0.11
	MPGRU	16.79 ± 0.52	1103.04 ± 106.83	23.63 ± 0.73
	GRIN	<u>12.08</u> ± 0.47	<u>523.14</u> ± 57.17	**17.00** ± 0.67
	SDA-GRIN	**12.05** ± 0.33	**473.94** ± 34.65	<u>17.31</u> ± 0.47

SDA-GRIN: Main Results

SDA-GRIN significantly outperforms nine baselines, especially on datasets with dynamic spatial patterns. It reduced MSE by 9.51% (AQI), 9.40% (AQI-36), and 1.94% (PEMS-BAY), while achieving competitive performance on METR-LA, matching GRIN.



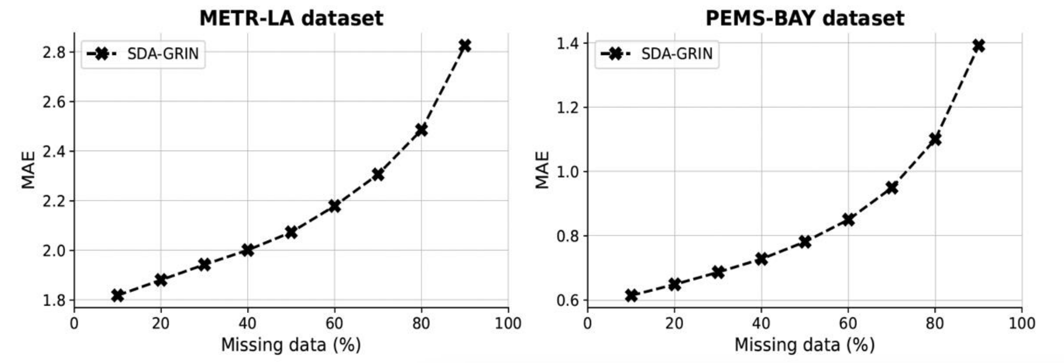
Dataset	Original (K)	Added (K)	Increase (%)
AQI	607	16	2.7
AQI-36	214	2	1.1
METR-LA	674	33	4.9
PEMS-BAY	690	33	4.8

- *Why SDA-GRIN perform better for AQI datasets compared to traffic data?* AQI datasets have more volatile spatial dependencies.
- *High Efficiency with Minimal Overhead:* The framework adds only 1.1% - 4.9% more parameters to the baseline GRIN model

SDA-GRIN: Ablation Study

An ablation study was conducted to analyze the effect of three key parameters on model performance: **window size**, **missing data rate**, and the number of **attention heads**.

Effect of Missing Data Rate: Performance declines as missing rates increase , as high data sparsity provides fewer samples for the MHA mechanism to attend to.



Effect of Window Size (Context Length): Larger window sizes improve performance by providing more context for attention, except for the low-variable AQI-36 dataset.

Dataset	Window Size	MAE ↓	MSE ↓	MRE (%) ↓
AQI	32	14.35 ± 0.09	701.19 ± 6.74	21.48 ± 0.14
	64	14.44 ± 0.28	701.08 ± 22.82	21.60 ± 0.42
	128	**14.29 ± 0.07**	**684.83 ± 5.93**	**21.38 ± 0.11**
	256*	—	—	—
AQI-36	32	**12.06 ± 0.33**	473.94 ± 34.65	**17.31 ± 0.47**
	64	12.06 ± 0.20	**460.42 ± 9.17**	17.32 ± 0.29
	128	12.66 ± 0.25	476.53 ± 30.19	18.17 ± 0.36
	256	13.27 ± 0.30	520.02 ± 35.68	19.05 ± 0.42
METR-LA	32	1.97 ± 0.00	10.98 ± 0.01	3.42 ± 0.00
	64	1.96 ± 0.00	10.86 ± 0.03	3.39 ± 0.00
	128	1.93 ± 0.00	10.60 ± 0.03	3.35 ± 0.00
	256	**1.91 ± 0.00**	**10.49 ± 0.04**	**3.30 ± 0.01**
PEMS-BAY	32	0.68 ± 0.00	1.54 ± 0.01	1.09 ± 0.00
	64	0.68 ± 0.00	1.52 ± 0.01	1.08 ± 0.00
	128	0.67 ± 0.00	**1.51 ± 0.01**	1.07 ± 0.00
	256	**0.67 ± 0.00**	1.53 ± 0.01	**1.07 ± 0.00**

Effect of Number of Attention Heads: A small number of attention heads (1-2) is most effective. Performance on AQI and AQI-36 datasets degraded with more heads.

Dataset	Heads	MAE ↓	MSE ↓	MAPE ↓
AQI36	1	**12.19**	**461.80**	0.4247
	2	12.28	483.42	**0.4097**
	3	12.37	464.73	0.4340
	4	14.09	519.04	0.6130
AQI	1	14.77	748.67	**0.3576**
	2	**14.41**	**708.47**	0.3640
	4	14.78	721.88	0.3889

Conclusion

- **Problem:** Existing methods for time series imputation often use static graphs, failing to capture that relationships between sensors can change over time.
- **Our Solution (SDA-GRIN):** We proposed SDA-GRIN, a framework that uses a Multi-Head Attention (MHA) mechanism to learn and adapt to these dynamic spatial dependencies in real-time.
- **Key Findings:**
  - SDA-GRIN significantly improves imputation accuracy, reducing MSE by up to 9.51% on the AQI dataset.
  - The approach is most effective on datasets with highly volatile spatial relationships, validating our hypothesis.
  - These performance gains are achieved with high efficiency, adding only 1.1% - 4.9% more parameters than the baseline model

Queen's UNIVERSITY

CONNECTED MINDS
Neural and Machine Systems for a Healthy, Just Society

Thank You!

Questions?

Project Page: https://ameskandari.github.io/sda-grin/

Code: https://github.com/AmEskandari/sdagrin

Contact: amir.eskandari@queensu.ca | https://ameskandari.github.io