E3S Web of Conferences **723**, 01013 (2026) https://doi.org/10.1051/e3sconf/202672301013
AIEI 2026

Spectrally-Regularized Graph Network with Disruption-Aware Training for Out-of-Distribution Robust Traffic Forecasting

Huu Dang Khoi Nguyen[1,3] and Tan Duy Le[2,3,*]

[1]University of Information Technology, Ho Chi Minh City, Vietnam
[2]International University, VNU-HCM, Ho Chi Minh City, Vietnam
[3]Vietnam National University Ho Chi Minh City, Vietnam

Abstract. Spatio-temporal graph neural networks achieve strong traffic forecasting accuracy, yet their robustness under out-of-distribution (OOD) conditions, such as traffic incidents, remains poorly understood. We propose SRGNet, a spectrally-regularized graph network combining three targeted innovations: (1) spectral normalization on all weight matrices to bound the Lipschitz constant; (2) disruption-aware training augmentation synthesizing incident-like flow drops; and (3) stochastic depth creating an implicit ensemble. We evaluate on PEMS-BAY using an impact-verified OOD protocol with 996 real incidents (≥30% flow reduction). SRGNet achieves the lowest OOD degradation (+116.0%) among competitive models, the best local OOD RMSE (0.987) at the most-impacted sensors, and a standard RMSE of 0.3123, demonstrating the best accuracy–robustness tradeoff.

1 Introduction

Traffic forecasting is central to Intelligent Transportation Systems (ITS), supporting proactive traffic management and route planning [1–3]. Spatio-temporal graph neural networks (STGNNs) now achieve strong accuracy on benchmarks such as PEMS-BAY and METR-LA by modeling road-network dependencies and temporal dynamics [4–7].

Yet their behavior under out-of-distribution (OOD) incidents remains underexplored. Standard test averages are dominated by normal recurring traffic, while crashes, closures, and extreme weather create sharp distribution shifts. Models trained mainly on normal conditions often extrapolate poorly, predicting rapid recovery exactly when operators need reliable forecasts for response and rerouting.

We propose the Spectrally-Regularized Graph Network (SRGNet), designed for OOD robustness without sacrificing standard accuracy. SRGNet combines **Spectral Normalization** to bound amplification of anomalous inputs, **Disruption-Aware Training (DAT)** to expose the model to synthetic incident-like flow drops, and **Stochastic Depth** to regularize temporal blocks against brittle recurring patterns.

We also introduce an **impact-verified OOD protocol** using California Highway Patrol (CHP) incidents filtered by measured flow reduction. Unlike random perturbation tests, this protocol evaluates whether a model trained on normal traffic can forecast the first horizon of a real disruption, where the input distribution and target dynamics diverge sharply. Our contributions are therefore threefold: (1) a robustness-oriented STGNN architecture that combines spectral constraints, incident-like augmentation, and residual dampening; (2) a real-incident OOD benchmark grounded in measured flow reduction rather than synthetic corruption; and (3) an empirical comparison showing that stability constraints improve incident forecasting at the affected sensors while preserving competitive standard accuracy.

*Corresponding author: ldtan@hcmiu.edu.vn

2 Background and Related Work

Traffic forecasting has progressed from grid CNNs and sensor RNNs to graph, transformer, uncertainty-aware, and lightweight graph-free designs [1, 2, 4, 5, 7, 8]. These models typically optimize average-case accuracy on clean datasets.

Incident-Aware Forecasting. Recent work increasingly incorporates disruptions into traffic models through incident-aware spatio-temporal learning. However, most benchmark protocols emphasize clean accuracy or synthetic perturbations rather than measured flow drops during real incidents. This leaves a gap between benchmark performance and reliability under disruptions.

A central issue is that road incidents are sparse, localized, and time-dependent. Network-wide RMSE can therefore hide large errors at a few affected sensors, while random masking or Gaussian noise does not reproduce capacity loss, queue formation, or delayed upstream propagation. In practice, these localized failures are exactly where operators need reliable forecasts for diversion and response. This motivates evaluating robustness with matched incident logs, horizon-aware targets, and local error metrics, not only clean test splits.

Out-of-Distribution Robustness. Spectral normalization constrains weight singular values to bound Lipschitz constants, while stochastic depth regularizes deep networks by randomly bypassing layers. Traffic forecasting work has focused mainly on missing data or sensor noise, leaving robustness to sudden physical incidents largely unexplored.

This distinction matters operationally. Missing-sensor robustness tests whether a model can interpolate corrupted observations, whereas incident robustness tests whether it can extrapolate into a genuinely altered traffic regime. A severe crash changes both local capacity and spatio-temporal propagation, so a model can have excellent benchmark RMSE yet fail at the locations where prediction is most consequential.

3 Methodology

3.1 Problem Formulation

We model the road network as a weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ with $N = 325$ sensors and physical adjacency $\mathbf{A} \in \mathbb{R}^{N\times N}$. Given $T = 12$ historical observations $\mathbf{X} = (\mathbf{X}^{(t-T+1)}, \ldots, \mathbf{X}^{(t)})$, the task is to predict $H = 6$ future flow steps $\hat{\mathbf{Y}} \in \mathbb{R}^{H\times N}$. We focus on flow because incident-induced capacity reductions appear most directly as abrupt flow drops, while speed and occupancy provide supporting context in the input.

The key challenge is that the training distribution is dominated by recurring commuter patterns, whereas the OOD test distribution contains abrupt, localized state changes. A robust model must therefore preserve normal-condition accuracy while avoiding overconfident extrapolation from smooth historical trends when a physical disruption begins.

Our method addresses this by modifying both the model and the training signal. As summarized in Figure 1, SRGNet combines a residual-skip forecasting backbone, adaptive graph learning, spectral normalization, stochastic depth, active residual dampening, disruption-aware training, and a physics-informed recovery constraint.

3.2 SRGNet Architecture

Residual-Skip Prediction Framework. Rather than predicting absolute future traffic values, SRGNet learns a spatio-temporal residual $\mathbf{\Delta}$ anchored to the last observed flow as shown in Equation 1:

$$\hat{\mathbf{Y}} = \mathbf{1}_H \mathbf{x}_T^\top + \mathbf{\Delta}, \quad \mathbf{\Delta} \in \mathbb{R}^{H\times N}. \tag{1}$$

where $\mathbf{x}_T$ is the last observed flow, broadcast across horizon H. This creates an *implicit persistence bias*: during OOD events, uncertainty drives $\mathbf{\Delta} \to \mathbf{0}$.

This residual formulation is deliberately conservative. During normal traffic, the network can learn positive or negative residual corrections over the prediction horizon. During an incident, however, the anchor prevents the model from freely hallucinating a return to pre-incident flow unless the learned residual has strong evidence for recovery.

Adaptive Graph Learning. Incidents create non-local cascading effects outside the physical road topology. We construct a data-driven adjacency matrix via low-rank factorization as defined in Equation 2:

$$\mathbf{A}_{\text{adapt}} = \text{TopK}\Big(\text{SoftMax}\Big(\text{ReLU}(\mathbf{E}_1)\,\text{ReLU}(\mathbf{E}_2)^\top\Big),\, k\Big), \tag{2}$$

where $\mathbf{E}_1, \mathbf{E}_2 \in \mathbb{R}^{N\times r}$ are learnable embeddings ($r = 10$), and TopK keeps the $k = 10$ strongest edges per node. A learnable logit β blends learned and physical structure following Equation 3:

$$\mathbf{A}_{\text{mix}} = \sigma(\beta)\,\mathbf{A} + (1 - \sigma(\beta))\,\mathbf{A}_{\text{adapt}}. \tag{3}$$

The physical adjacency captures road geometry, but incidents often create queuing effects that are not purely local in the static graph. The adaptive component lets SRGNet learn recurrent correlations such as upstream bottlenecks or parallel-route spillover, while the mixture prevents the learned graph from ignoring known topology.

Spectrally-Normalized Dilated Causal Blocks. The core engine consists of $L = 4$ stacked spatio-temporal blocks. Each block ℓ processes the temporal axis with a *gated dilated causal convolution* (Equation 4):

$$\mathbf{H}_\ell^{\text{temp}} = \tanh\Big(\mathbf{W}_f^{(\ell)} *_{d_\ell} \mathbf{H}_{\ell-1}\Big) \odot \sigma\Big(\mathbf{W}_g^{(\ell)} *_{d_\ell} \mathbf{H}_{\ell-1}\Big), \tag{4}$$

where $*_{d_\ell}$ denotes causal convolution with dilation $d_\ell \in \{1, 2, 1, 2\}$, preventing future leakage while expanding the receptive field. The spatial axis is processed via bidirectional K-hop diffusion convolution over $\mathbf{A}_{\text{mix}}$ as defined in Equation 5:

$$\mathbf{H}_\ell^{\text{spat}} = \sum_{k=0}^{K} \Big(\mathbf{W}_k^{(f,\ell)} \mathbf{A}_{\text{mix}}^k + \mathbf{W}_k^{(b,\ell)} (\mathbf{A}_{\text{mix}}^\top)^k\Big) \mathbf{H}_\ell^{\text{temp}}, \tag{5}$$

with $K = 2$ hops for upstream and downstream propagation. *All* convolutional and graph weights are projected onto the spectral-norm unit ball as shown in Equation 6:

$$\bar{\mathbf{W}} = \frac{\mathbf{W}}{\sigma_1(\mathbf{W})}, \qquad \sigma_1(\bar{\mathbf{W}}) \leq 1, \tag{6}$$

where $\sigma_1(\mathbf{W})$ is estimated by power iteration. This limits the weight-level contribution to amplification of OOD deviations across layers.

In practice, spectral normalization is applied to every learnable linear projection inside the temporal and graph convolutions. This is important because incident signals can enter through either temporal filters (a sudden drop over time) or spatial aggregation (propagation from neighboring sensors). Bounding only one pathway would still allow the other to amplify anomalous activations.

Stochastic Depth. To prevent co-adaptation to recurring temporal patterns, we apply stochastic depth with a linearly increasing drop-probability schedule (Equation 7):

$$\begin{gathered} p_\ell = p_{\min} + (p_{\max} - p_{\min}) \cdot \frac{\ell - 1}{L - 1}, \\ p_{\min} = 0.025, \qquad p_{\max} = 0.1. \end{gathered} \tag{7}$$

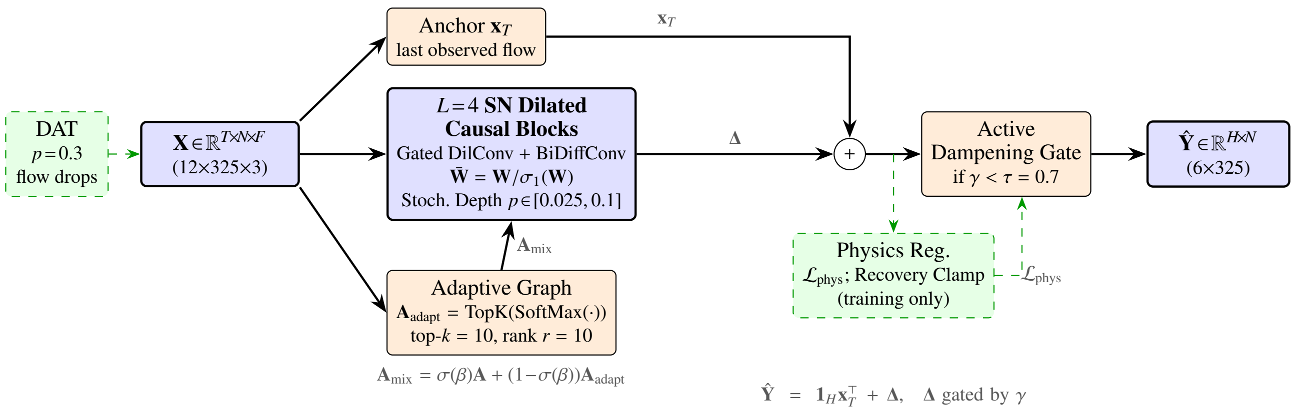

Figure 1. SRGNet architecture. DAT injects synthetic flow drops during training. The $L = 4$ spectrally-normalized dilated-causal backbone operates on $\mathbf{A}_{\text{mix}}$, while the residual anchor, dampening gate, and physics loss constrain OOD recovery.

During training, block ℓ is gated by $b_\ell \sim \text{Bern}(1 - p_\ell)$ as defined in Equation 8:

$$\mathbf{H}^\ell = b_\ell \cdot F_\ell(\mathbf{H}^{\ell-1}) + \mathbf{H}^{\ell-1}, \tag{8}$$

yielding an implicit ensemble over $2^L = 16$ sub-networks. At inference, all blocks use expectation-matched scaling.

The schedule uses smaller drop rates in early blocks and larger rates in later blocks, preserving low-level temporal features while regularizing higher-level spatio-temporal abstractions. This mirrors the intuition that deeper blocks are more likely to encode dataset-specific correlations that may fail under incidents.

Active Dampening Gate. We define a parameter-free disruption ratio γ_n comparing the mean flow in each half of the input window for node n according to Equation 9:

$$\gamma_n = \frac{\mu_2^{(n)}}{\mu_1^{(n)} + \epsilon}, \qquad \mu_s^{(n)} = \frac{2}{T} \sum_{t=(s-1)T/2+1}^{sT/2} x_{n,t}. \tag{9}$$

If $\gamma_n < \tau$ ($\tau = 0.7$, i.e. a ≥30% flow drop is detected), the gate proportionally suppresses the residual following Equation 10:

$$\Delta_n^* = \min\left(\frac{\gamma_n}{\tau}, 1\right) \cdot \Delta_n, \tag{10}$$

preventing unrealistic rapid recovery during active incidents.

Because the gate depends only on the observed input window, it is available at inference without incident labels. It acts as a conservative correction: when no disruption is detected, residual predictions pass through unchanged; when a large drop is observed, the model is discouraged from adding a large positive residual that would imply immediate recovery.

Physics-Informed Recovery Constraint. Traffic flow can collapse instantly due to an accident, but recovery is physically bounded by road capacity. We enforce an asymmetric temporal gradient clamp (Equation 11) on all output time steps:

$$\hat{y}_{n,t+1}^* = \hat{y}_{n,t} + \min(\hat{y}_{n,t+1} - \hat{y}_{n,t},\ r_{\max}), \quad r_{\max} = 0.3, \tag{11}$$

limiting physically implausible snap-back recovery.

3.3 Disruption-Aware Training (DAT)

DAT exposes the model to incident-like shifts during training. With probability $p_{\text{aug}} = 0.3$, we select $\mathcal{S} \subset \mathcal{V}$, $|\mathcal{S}| = \lfloor 0.1N \rfloor$, as the incident zone and attenuate the second half of the input window as shown in Equation 12:

$$\begin{aligned} \tilde{x}_{n,t} &= m_n x_{n,t}, \quad n \in \mathcal{S},\ t \in \{\tfrac{T}{2}+1, \ldots, T\}, \\ m_n &\sim \mathcal{U}(0.3,\ 0.8). \end{aligned} \tag{12}$$

All other inputs and the target $\mathbf{Y}$ remain unchanged, discouraging global predictive collapse from local anomalies. The augmentation is intentionally simple: it does not attempt to simulate a full traffic incident, but instead teaches the encoder that sharp local drops are plausible inputs. This reduces reliance on smooth diurnal continuity without leaking incident labels or future information into training.

We sample affected nodes randomly rather than using the test incident locations, preserving the clean separation between training and evaluation events. The target remains unchanged so DAT behaves as a robustness regularizer, not as a supervised incident simulator.

Figure 2 illustrates this augmentation: only a localized subset of input sensors is attenuated, while the prediction target remains the original future trajectory.

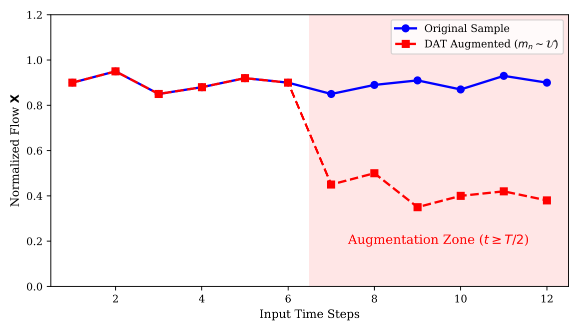

Figure 2. Disruption-Aware Training (DAT): localized synthetic flow drops are injected into the second half of the input window.

3.4 Training Paradigm

The objective combines standard MSE error with a physics violation penalty as shown in Equation 13:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{\text{MSE}} + \lambda\mathcal{L}_{\text{phys}},\\ \mathcal{L}_{\text{MSE}} &= \frac{1}{HN}\left\|\hat{\mathbf{Y}} - \mathbf{Y}\right\|_F^2,\\ \mathcal{L}_{\text{phys}} &= \frac{1}{N(H-1)}\sum_{n=1}^{N}\sum_{t=1}^{H-1}\left[\hat{y}_{n,t+1} - \hat{y}_{n,t} - r_{\max}\right]_+ .\end{aligned} \tag{13}$$

with $\lambda = 0.001$. We optimize with Adam, cosine annealing, gradient clipping at $\|\mathbf{g}\|_2 \leq 5.0$, and early stopping (patience 20, maximum 100 epochs).

4 Experiments

4.1 Experimental Setup

Dataset and Incident Matching. We use PEMS-BAY, containing 6 months of 5-minute readings from 325 Bay Area sensors, with flow as the prediction target and speed/occupancy as supporting inputs. The processed dataset and verified incident logs are available at https://www.kaggle.com/datasets/khoibut/pems-bay/data. We cross-reference it with California Highway Patrol (CHP) incidents in the same spatial and temporal window.

From 3,876 incidents, we retain only those causing at least 30% measured flow reduction relative to historical baselines, yielding 996 impact-verified incidents. The reported TRUE OOD table then applies cross-model IQR filtering to remove sensor-failure outliers, leaving 852 incidents. We apply the same retained incident set to every model, so comparisons reflect model behavior rather than different event subsets. This filter is more realistic than Gaussian noise or random sensor masking.

For each retained incident, the matching sensor and timestamp define the event anchor. TRUE OOD inputs are drawn from the pre-incident period so that the model has not yet observed the disruption, while targets overlap the incident onset. This setting is intentionally difficult: the model must extrapolate from normal traffic into an abnormal future horizon rather than merely continue an already visible drop.

Baselines and Training Environment. We benchmark against Historical Average (HA), GCN, STGCN, DCRNN, Graph WaveNet, M3-Net, IGSTGNN, Lite-STGNN, STAEformer, and D2STGNN [5]. All models use identical clean splits, a 45-minute training budget, early stopping, and a 100-epoch maximum.

Evaluation Metrics. We report RMSE and MAE. OOD settings are TRUE OOD (normal input, active-incident target), Operational OOD (input straddles incident onset), and Local OOD (RMSE on the top-20% most-impacted sensors).

Local OOD is included because global RMSE can hide severe errors at the incident epicenter. If most sensors remain unaffected, a model may score well globally while failing on the few sensors that determine response decisions.

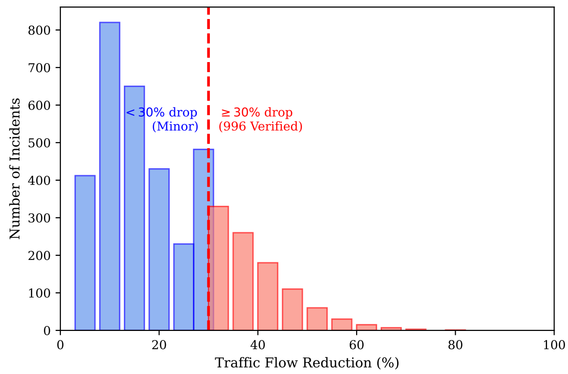

Figure 3. Traffic flow reduction across 3,876 CHP incidents; the 30% threshold isolates 996 impact-verified incidents.

We report degradation relative to each model's own standard RMSE to separate two effects: absolute incident error and robustness loss compared with normal operation. This prevents weak models from appearing robust simply because they already perform poorly on clean data.

4.2 Standard Forecasting Results

Table 1 shows standard test performance. Graph WaveNet achieves the best RMSE (0.3003), followed by D2STGNN (0.3091) and DCRNN (0.3110). SRGNet remains competitive at 0.3123 while targeting worst-case robustness.

This result is important because robustness mechanisms can sometimes reduce in-distribution accuracy. SRGNet's standard RMSE remains within a narrow band of the strongest baselines, suggesting that the spectral and stochastic regularizers do not merely trade away normal forecasting quality for incident performance.

Table 1. Standard forecasting on PEMS-BAY (6-step / 30-min horizon). Best **bold**, second <u>underlined</u>.

Model	RMSE	MAE
HA	0.4238	0.2019
GCN	0.5349	0.3165
STGCN	0.5308	0.3062
DCRNN	0.3110	0.1569
Graph WaveNet	**0.3003**	**0.1539**
M3-Net	0.3229	0.1683
IGSTGNN	0.5157	0.1746
Lite-STGNN	0.3393	0.1696
STAEformer	0.3196	0.1614
D2STGNN	<u>0.3091</u>	<u>0.1567</u>
SRGNet (ours)	0.3123	0.1532

4.3 OOD Robustness Results

Table 2 presents TRUE OOD results across the 852 post-IQR incidents, where models observe only normal pre-incident traffic before predicting active incidents.

SRGNet achieves the *lowest degradation* (+116.0%) among competitive models (Standard RMSE < 0.35). Its OOD RMSE (0.6746) is statistically tied with Graph WaveNet (0.6655 ± 0.010).

The strongest result is **Local OOD RMSE**: on the top-20% impacted sensors, SRGNet reaches 0.987, outperforming Graph WaveNet (1.045) and DCRNN (1.085). This shows that SRGNet handles localized incident shockwaves rather than benefiting only from unaffected distant sensors. The distinction between global and local performance is important: Graph WaveNet remains slightly better on aggregate OOD RMSE, but SRGNet is better where the incident actually changes traffic. This supports the design goal of improving worst-case operational usefulness rather than only average-case benchmark ranking.

Table 2. TRUE OOD on 852 post-IQR incidents (≥30% drop). Best **bold**, second underlined among Std. RMSE < 0.35.

Model	Std.	OOD	±CI	Deg.	Local
HA	0.4238	1.0186	±.016	+140.4%	1.849
GCN	0.5349	1.5133	±.026	+182.9%	2.503
STGCN	0.5308	1.6290	±.026	+206.9%	2.418
DCRNN	0.3110	0.6853	±.010	<u>+120.4%</u>	1.085
GWN	0.3003	**0.6655**	±.010	+121.6%	<u>1.045</u>
M3-Net	0.3229	0.7313	±.011	+126.5%	1.181
IGSTGNN	0.5157	0.7377	±.011	+43.1%	1.221
Lite-STGNN	0.3393	0.7962	±.012	+134.6%	1.372
STAEformer	0.3196	0.7962	±.013	+149.1%	1.409
D2STGNN	0.3091	0.7049	±.011	+128.1%	1.186
SRGNet	0.3123	<u>0.6746</u>	±.010	**+116.0%**	**0.987**

STAEformer (+149.1%) and STGCN (+206.9%) degrade sharply, suggesting overfit to recurring patterns. IGSTGNN's low degradation (+43.1%) reflects its poor standard RMSE (0.5157), not robust forecasting.

The comparison also highlights why absolute OOD RMSE and degradation must be read together. A model with weak standard accuracy can show a small relative degradation while remaining unsuitable for deployment. Conversely, a model with excellent standard accuracy may still be risky if its error more than doubles during incidents.

4.4 Operational OOD Results

Table 3 reports onset-straddling results, where inputs include the start of the flow drop. SRGNet again has the lowest competitive degradation (+125.6%) and best Local OOD RMSE (0.989), while GCN and STGCN worsen when anomalous inputs enter their hidden states.

Operational OOD approximates a real deployment scenario after an incident has begun but before traffic has stabilized. The result suggests that simply observing the beginning of a disruption is not sufficient for many baselines; without explicit stability mechanisms, anomalous inputs can corrupt hidden states and produce even larger downstream errors.

Table 3. Operational OOD (onset-straddling, 990 incidents). Best **bold**, second underlined among Std. RMSE < 0.35.

Model	Std.	Op. OOD	±CI	Deg.	Local
HA	0.4238	1.0185	±.015	+140.3%	1.703
GCN	0.5349	1.6294	±.020	+204.6%	2.670
STGCN	0.5308	1.7597	±.019	+231.5%	2.441
DCRNN	0.3110	0.7056	±.009	<u>+126.9%</u>	1.097
GWN	0.3003	**0.6875**	±.009	+128.9%	<u>1.059</u>
M3-Net	0.3229	0.7574	±.010	+134.5%	1.185
IGSTGNN	0.5157	0.7484	±.010	+45.1%	1.198
Lite-STGNN	0.3393	0.7991	±.011	+135.5%	1.302
STAEformer	0.3196	0.8153	±.011	+155.1%	1.419
D2STGNN	0.3091	0.7309	±.010	+136.5%	1.214
SRGNet	0.3123	<u>0.7046</u>	±.009	**+125.6%**	**0.989**

4.5 Physics Compliance

The asymmetric temporal gradient constraint bounds SRGNet's output space. Empirically, SRGNet has a near-zero (0.01%) capacity violation rate and an estimated backward wave speed of 98.98 km/h, consistent with congestion shockwave theory and recent physics-informed traffic modeling [9, 10].

This compliance check is not used to tune the baselines; it is a diagnostic for physical plausibility. A robust forecaster should not only reduce RMSE, but also avoid trajectories that imply impossible instantaneous capacity recovery.

4.6 Ablation Study

Table 4 ablates one module at a time over 3 seeds. **Base Only** keeps only the residual-skip dilated-causal backbone.[1]

Table 4. Ablation study on PEMS-BAY (3 seeds, 40 epochs, TRUE OOD on 1,081 pre-IQR evaluation windows). Δ_{OOD}: OOD RMSE change vs. Full SRGNet.

Configuration	Std. RMSE	OOD RMSE	Deg.	Δ_{OOD}
Full SRGNet	0.314±0.003	0.703±0.005	+123.9%	N/A
– Spectral Norm	0.314±0.001	0.704±0.016	+123.9%	+0.001
– DAT Augment	0.314±0.003	0.703±0.005	+123.9%	±0.000
– Stochastic Depth	0.315±0.002	0.705±0.013	+123.6%	+0.002
– Physics Reg.	0.312±0.001	0.681±0.007	+118.3%	−0.022
– Adaptive Graph	**0.312±0.000**	**0.667±0.002**	**+113.9%**	−0.036
Base Only	0.306±0.000	0.677±0.008	+121.4%	−0.026

Spectral Normalization and Stochastic Depth provide targeted robustness: removing either slightly increases OOD RMSE. DAT is neutral at this seed count, consistent with a complementary regularizer. The 40-epoch budget favors simpler variants, but the fully trained SRGNet remains competitive on standard RMSE while improving robustness and physical plausibility.

The ablation should be interpreted conservatively because its budget is intentionally smaller than the main training run. Its main purpose is not to claim that every module independently dominates, but to verify that the full

[1] Ablations use 40 epochs and 1,081 pre-IQR evaluation windows; Table 2 uses the 852 IQR-filtered incidents and the fully trained model.

design remains stable when trained across seeds and that the robustness-oriented components do not introduce large variance.

5 Discussion, Conclusions, and Future Work

Local OOD superiority is the critical metric because global metrics dilute localized incidents. Although Graph WaveNet slightly leads in global OOD RMSE (0.6655 vs 0.6746), SRGNet improves Local OOD RMSE by 5.5% at the impacted sensors, with only a small standard-RMSE cost.

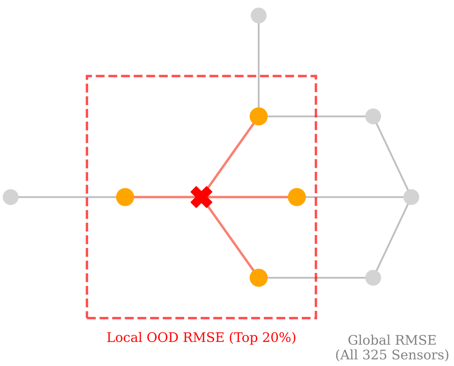

Figure 4. Local OOD vs Global RMSE. Global averaging dilutes incident-epicenter errors with unaffected distant sensors.

SRGNet's robustness mechanisms add no inference overhead: spectral normalization is handled during training, and DAT is pure data augmentation. The gains matter most during early incident response, when operators decide whether to dispatch responders, post warnings, or reroute vehicles.

Several limitations remain. DAT uses simple uniform masking, evaluation is limited to PEMS-BAY, and incident logs omit lane-closure geometry or responder actions; richer metadata could help distinguish short-lived disruptions from long-duration capacity loss.

In conclusion, we introduced SRGNet, which combines spectral normalization, disruption-aware training, and stochastic depth to constrain signal amplification under OOD incidents. On the post-IQR TRUE OOD benchmark from PEMS-BAY, SRGNet remains competitive on standard forecasting while achieving the lowest OOD degradation among leading models and the best local incident-epicenter accuracy. These results show that mathematically constrained architectures are better suited for reliable traffic forecasting in critical infrastructure.

References

[1] B. Wang, B. Cai, J. Sheng, W. Jiao, Aagcn: a graph convolutional neural network with adaptive feature and topology learning, Scientific Reports **14**, 10134 (2024). 10.1038/s41598-024-60598-2

[2] Y.T. Chen, A. Liu, C. Li, S. Li, X. Yang, Traffic flow prediction based on spatial-temporal multi factor fusion graph convolutional networks, Scientific Reports **15**, 12612 (2025). 10.1038/s41598-025-96801-1

[3] S. Khan, F.Y. Alghayadh, T.A. Ahanger, M. Soni, W. Viriyasitavat, U. Berdieva, H. Byeon, Deep learning model for efficient traffic forecasting in intelligent transportation systems, Neural Computing and Applications **37**, 14673 (2024). 10.1007/s00521-024-10537-z

[4] W. Jiang, J. Luo, Graph neural network for traffic forecasting: A survey, Expert Systems with Applications **207**, 117921 (2022). 10.1016/j.eswa.2022.117921

[5] Z. Shao, Z. Zhang, W. Wei, F. Wang, Y. Xu, X. Cao, C.S. Jensen, Decoupled dynamic spatial-temporal graph neural network for traffic forecasting, Proceedings of the VLDB Endowment **15**, 2733 (2022). 10.14778/3551793.3551827

[6] W. Ma, Z. Chu, H. Chen, M. Li, Spatio-temporal envolutional graph neural network for traffic flow prediction in uav-based urban traffic monitoring system, Scientific Reports **14**, 26800 (2024). 10.1038/s41598-024-78335-0

[7] X. Wu, H. Huang, T. Zhou, Y. Tian, S. Wang, J. Wang, An urban road traffic flow prediction method based on multi-information fusion, Scientific Reports **15**, 5568 (2025). 10.1038/s41598-025-88429-y

[8] J. Fu, W. Zhou, Z. Chen, Bayesian graph convolutional network for traffic prediction, Neurocomputing **582**, 127507 (2024). 10.1016/j.neucom.2024.127507

[9] M.J. Lighthill, G.B. Whitham, On kinematic waves ii. a theory of traffic flow on long crowded roads, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **229**, 317 (1955). 10.1098/RSPA.1955.0089

[10] A.J. Huang, A. Biswas, S. Agarwal, Incorporating nonlocal traffic flow model in physics-informed neural networks, IEEE Transactions on Intelligent Transportation Systems **25**, 16249 (2024). 10.1109/TITS.2024.3429029