Graph Deep Learning for Time Series and Spatiotemporal Data

Daniele **Zambon**

Graph Machine Learning Group (gmlg.ch)
The Swiss AI Lab IDSIA
Università della Svizzera italiana

Università di Padova, Italy · Spring 2026

Module 3

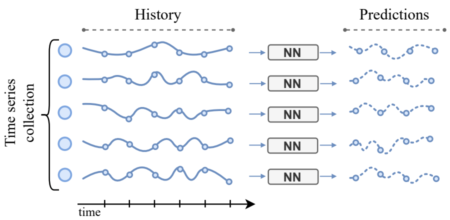
Global and Local Models

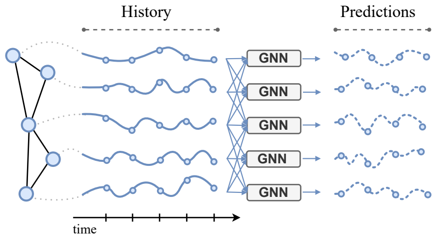
Key Concepts from Previous Lectures

Deep learning for time series

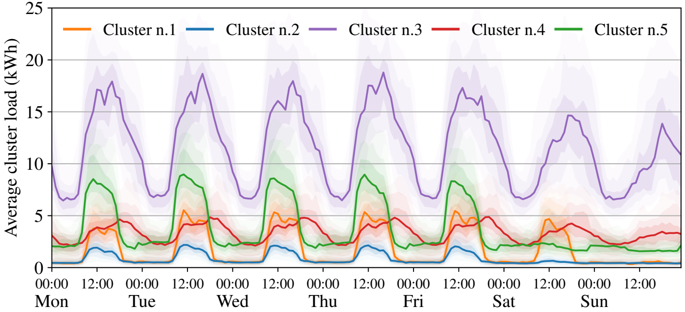
Modern deep learning methods for time series prediction tasks (e.g., forecasting) rely on a single neural network trained on a collection of related time series.

Graph deep learning provide appropriate operators to account for dependencies.





Local effects

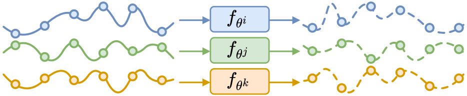


Group of time series presenting different patterns.

[23] Cini, Marisca, Zambon, and Alippi, "Taming Local Effects in Graph-based Spatiotemporal Forecasting", NeurIPS 2023.

Global and local predictors

Local models

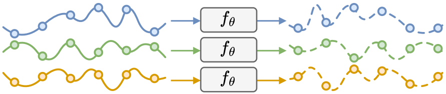


$$\hat{\boldsymbol{x}}^i_{t+h} = f\left(\boldsymbol{x}^i_{t-W:t}, \dots; \boldsymbol{\theta}^i\right)$$

☺ Tailored to each time series.

☹ Inefficient.

Global models



$$\hat{\boldsymbol{x}}^i_{t+h} = f\left(\boldsymbol{x}^i_{t-W:t}, \dots; \boldsymbol{\theta}\right)$$

☺ Sample efficient.

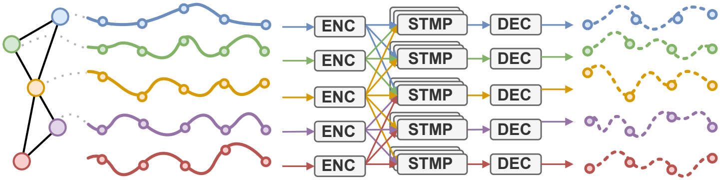
☺ Allows for more complex models.

☹ Both approaches neglect dependencies among time series.

[6] Montero-Manso *et al.*, “Principles and algorithms for forecasting groups of time series: Locality and globality”, IJF 2021.

Global and local models

Globality and locality in STGNNs

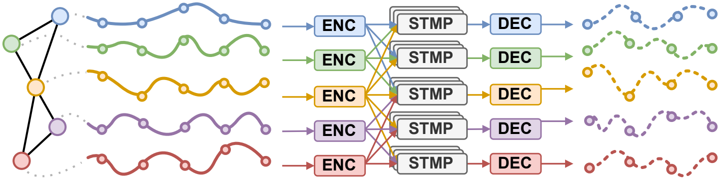


Standard STGNNs are **global** models.

- ☺ Can handle arbitrary node sets.
- ☺ Neighbors provide further conditioning on the predictions.
- ☹ Might struggle with local effects.
- ☹ Might need long windows and high model capacity.

💡 Use hybrid global-local STGNNs.

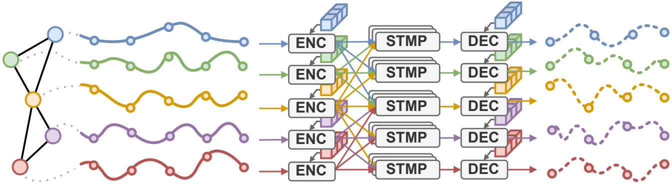
Global-local STGNNs



We can turn some global components of the architecture into local.

- ☺ Resulting models can capture local effects.
- ☹ Might require a large number of local parameters.

Global-local STGNNs with node embeddings



Node embeddings can amortize the learning of local components.

Node embeddings are a table of **learnable parameters** $Q \in \mathbb{R}^{N \times d_q}$ associated with **each node**.

- Fed into encoder/decoder, amortize the learning of local components.
- Most of the model's parameters remain shared.
- Number of parameters scales linearly with the number of time series . . .
  - $\rightarrow$ One might consider intermediate solutions, e.g., learning embeddings for clusters of time series.

[23] Cini, Marisca, Zambon, and Alippi, "Taming Local Effects in Graph-based Spatiotemporal Forecasting", NeurIPS 2023.

Some empirical results

Models	MetrLA	PemsBAY	CER-E	AQI	MetrLA	PemsBAY	CER-E	AQI
Reference arch.	Global models				+ Local node embeddings			
RNN	$3.54_{\pm.00}$	$1.77_{\pm.00}$	$4.57_{\pm.01}$	$14.02_{\pm.04}$	**$3.15_{\pm.03}$**	**$1.59_{\pm.00}$**	**$4.22_{\pm.02}$**	**$13.73_{\pm.04}$**
GCRNN-IMP	$3.35_{\pm.01}$	$1.70_{\pm.01}$	$4.44_{\pm.01}$	$12.87_{\pm.02}$	**$3.10_{\pm.01}$**	**$1.59_{\pm.00}$**	**$4.18_{\pm.01}$**	**$12.48_{\pm.03}$**
RNN+IMP	$3.34_{\pm.01}$	$1.72_{\pm.00}$	$4.39_{\pm.01}$	$12.74_{\pm.02}$	**$3.08_{\pm.01}$**	**$1.58_{\pm.00}$**	**$4.12_{\pm.03}$**	**$12.33_{\pm.02}$**
GCRNN-AMP	$3.22_{\pm.02}$	$1.65_{\pm.00}$	$4.57_{\pm.04}$	$12.29_{\pm.02}$	**$3.07_{\pm.02}$**	**$1.59_{\pm.00}$**	**$4.17_{\pm.02}$**	**$12.17_{\pm.05}$**
RNN+AMP	$3.24_{\pm.01}$	$1.66_{\pm.00}$	$4.31_{\pm.01}$	$12.30_{\pm.02}$	**$3.06_{\pm.01}$**	**$1.58_{\pm.01}$**	**$4.13_{\pm.01}$**	**$12.15_{\pm.02}$**
Baseline arch.	Original				+ Local node embeddings			
DCRNN	$3.22_{\pm.01}$	$1.64_{\pm.00}$	$4.28_{\pm.01}$	$12.96_{\pm.03}$	**$3.07_{\pm.02}$**	**$1.60_{\pm.00}$**	**$4.13_{\pm.02}$**	**$12.53_{\pm.02}$**
GraphWaveNet	$3.05_{\pm.03}$	**$1.56_{\pm.01}$**	**$3.97_{\pm.01}$**	$12.08_{\pm.11}$	**$2.99_{\pm.02}$**	$1.58_{\pm.00}$	$4.01_{\pm.01}$	**$11.81_{\pm.04}$**
AGCRN	$3.16_{\pm.01}$	**$1.61_{\pm.00}$**	$4.45_{\pm.01}$	$13.33_{\pm.02}$	**$3.14_{\pm.00}$**	$1.62_{\pm.00}$	**$4.37_{\pm.02}$**	**$13.28_{\pm.03}$**

Table 1: MAE on benchmark datasets.

[23] Cini, Marisca, Zambon, and Alippi, "Taming Local Effects in Graph-based Spatiotemporal Forecasting", NeurIPS 2023.

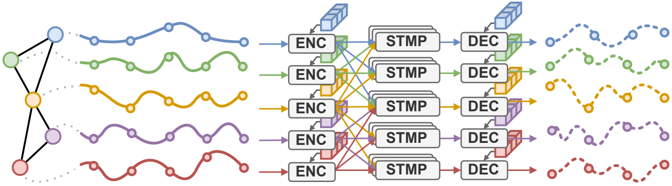
Some empirical results on synthetic data

	Global			Embeddings		
W	$d_h = 16$	$d_h = 32$	$d_h = 64$	$d_h = 16$	$d_h = 32$	$d_h = 64$
2	.5371±.0014	.4679±.0016	.4124±.0021	.3198±.0001	.3199±.0001	.3203±.0001
6	.4059±.0032	.3578±.0031	.3365±.0006	.3200±.0002	.3201±.0001	.3209±.0002
12	.3672±.0035	.3362±.0012	.3280±.0003	.3200±.0001	.3200±.0000	.3211±.0003
24	.3485±.0032	.3286±.0005	.3250±.0001	.3200±.0002	.3200±.0000	.3211±.0003

One-step-ahead MAE on GPVAR-L with varying window length W and node-embedding size d_h.

[23] Cini, Marisca, Zambon, and Alippi, “Taming Local Effects in Graph-based Spatiotemporal Forecasting”, NeurIPS 2023.

Transferability



❗ Hybrid global-local STGNNs are not inductive models.

However, the cost of transfer learning can be reduced.

- ☺ Keep shared parameters fixed and finetune local parameters only.
- ☺ Node embeddings can be regularized to facilitate transfer further.

[23] Cini, Marisca, Zambon, and Alippi, "Taming Local Effects in Graph-based Spatiotemporal Forecasting", NeurIPS 2023.
[24] Butera, Felice, Cini, and Alippi, "On the Regularization of Learnable Embeddings for Time Series Forecasting", TMLR 2025.

Transfer learning results

We consider datasets coming from four different traffic networks.

→ **One of the networks is left out** at training time and used for evaluating **transferability**.

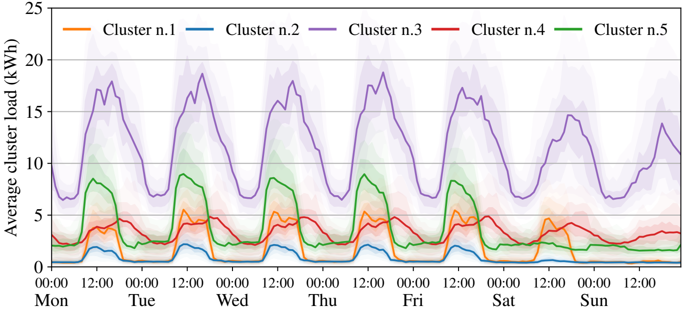
	RNN+IMP	PEMS03	PEMS04	PEMS07	PEMS08
Fine-tuning	Global	15.30 ± 0.03	21.59 ± 0.11	23.82 ± 0.03	15.90 ± 0.07
	Embeddings	14.64 ± 0.05	20.27 ± 0.11	**22.23 ± 0.08**	**15.45 ± 0.06**
	– **Variational**	**14.56 ± 0.03**	20.19 ± 0.05	22.43 ± 0.02	**15.41 ± 0.06**
	– **Clustering**	**14.60 ± 0.02**	**19.91 ± 0.11**	**22.16 ± 0.07**	**15.41 ± 0.06**
Zero-shot		18.20 ± 0.09	23.88 ± 0.08	32.76 ± 0.69	20.41 ± 0.07

Table 2: Transfer learning results (MAE) after fine-tuning on a week of data.

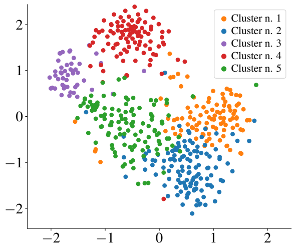
Node embeddings can be regularized through **variational** or **clustering** approaches.

[23] Cini, Marisca, Zambon, and Alippi, “Taming Local Effects in Graph-based Spatiotemporal Forecasting”, NeurIPS 2023.

Regularizing with clustering



Average cluster load by day of the week.

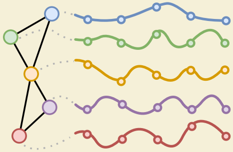


t-SNE plot of the corresponding node embeddings.

[23] Cini, Marisca, Zambon, and Alippi, “Taming Local Effects in Graph-based Spatiotemporal Forecasting”, NeurIPS 2023.

End of Part 1: what we have so far

1. We formalized the problem of processing correlated time series.
2. Graph representations allows for modeling dependencies among them.
3. We saw approaches to building spatiotemporal graph neural networks, the associated trade-offs.
4. We discussed about global/local deep learning models.



DEMO

Coding Spatiotemporal GNNs with Torch Spatiotemporal

github.com/dzambon/spatiotemporal-learning-lab

References i

[1] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, **"DeepAR: Probabilistic forecasting with autoregressive recurrent networks,"** *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020.

[2] K. Benidis, S. S. Rangapuram, V. Flunkert, *et al.*, **"Deep learning for time series forecasting: Tutorial and literature survey,"** *ACM Comput. Surv.*, vol. 55, no. 6, Dec. 2022, ISSN: 0360-0300. DOI: 10.1145/3533382. [Online]. Available: https://doi.org/10.1145/3533382.

[3] A. Cini, I. Marisca, D. Zambon, and C. Alippi, **"Graph Deep Learning for Time Series Forecasting,"** *ACM Comput. Surv.*, 2025, ISSN: 0360-0300. DOI: 10.1145/3742784. [Online]. Available: https://doi.org/10.1145/3742784.

[4] D. Zambon, M. Cattaneo, I. Marisca, J. Bhend, D. Nerini, and C. Alippi, ***PeakWeather*,** https://huggingface.co/datasets/MeteoSwiss/PeakWeather, 2025.

[5] M. Jin, H. Y. Koh, Q. Wen, *et al.*, **"A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection,"** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10 466–10 485, 2024. DOI: 10.1109/TPAMI.2024.3443141.

References ii

[6] P. Montero-Manso and R. J. Hyndman, **"Principles and algorithms for forecasting groups of time series: Locality and globality,"** *International Journal of Forecasting*, vol. 37, no. 4, pp. 1632–1653, 2021.

[7] R. Sen, H.-F. Yu, and I. S. Dhillon, **"Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting,"** *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[8] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, **"Neural message passing for quantum chemistry,"** in *International conference on machine learning*, PMLR, 2017, pp. 1263–1272.

[9] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, **"Geometric deep learning: Grids, groups, graphs, geodesics, and gauges,"** *arXiv preprint arXiv:2104.13478*, 2021.

[10] T. N. Kipf and M. Welling, **"Semi-supervised classification with graph convolutional networks,"** *arXiv preprint arXiv:1609.02907*, 2016.

[11] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, **"Graph attention networks,"** in *International Conference on Learning Representations*, 2018.

References iii

[12] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, **"Empirical evaluation of gated recurrent neural networks on sequence modeling,"** *arXiv preprint arXiv:1412.3555*, 2014.

[13] Y. Seo, M. Defferrard, P. Vandergheynst, and X. Bresson, **"Structured sequence modeling with graph convolutional recurrent networks,"** in *International Conference on Neural Information Processing*, Springer, 2018, pp. 362–373.

[14] Y. Li, R. Yu, C. Shahabi, and Y. Liu, **"Diffusion convolutional recurrent neural network: Data-driven traffic forecasting,"** in *International Conference on Learning Representations*, 2018.

[15] B. Yu, H. Yin, and Z. Zhu, **"Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting,"** in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 3634–3640.

[16] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, **"Graph wavenet for deep spatial-temporal graph modeling,"** in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019, pp. 1907–1913.

[17] I. Marisca, A. Cini, and C. Alippi, **"Learning to reconstruct missing data from spatiotemporal graphs with sparse observations,"** in *Advances in Neural Information Processing Systems*, 2022.

References iv

[18] Z. Wu, D. Zheng, S. Pan, Q. Gan, G. Long, and G. Karypis, **"Traversenet: Unifying space and time in message passing for traffic forecasting,"** *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[19] M. Sabbaqi and E. Isufi, **"Graph-time convolutional neural networks: Architecture and theoretical analysis,"** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 12, pp. 14 625–14 638, Dec. 2023, ISSN: 1939-3539. DOI: `10.1109/TPAMI.2023.3311912`.

[20] J. Gao and B. Ribeiro, **"On the equivalence between temporal and static equivariant graph representations,"** in *International Conference on Machine Learning*, PMLR, 2022, pp. 7052–7076.

[21] A. Cini, I. Marisca, F. M. Bianchi, and C. Alippi, **"Scalable spatiotemporal graph neural networks,"** *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 6, pp. 7218–7226, Jun. 2023. DOI: `10.1609/aaai.v37i6.25880`.

[22] V. G. Satorras, S. S. Rangapuram, and T. Januschowski, **"Multivariate time series forecasting with latent graph inference,"** *arXiv preprint arXiv:2203.03423*, 2022.

[23] A. Cini, I. Marisca, D. Zambon, and C. Alippi, **"Taming local effects in graph-based spatiotemporal forecasting,"** *arXiv preprint arXiv:2302.04071*, 2023.

References v

[24] L. Butera, G. D. Felice, A. Cini, and C. Alippi, **“On the regularization of learnable embeddings for time series forecasting,”** *Transactions on Machine Learning Research*, 2025, ISSN: 2835-8856. [Online]. Available: `https://openreview.net/forum?id=F5ALCh3GWG`.