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

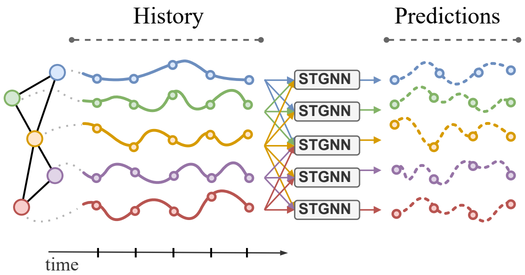
Module

Model Quality Assessment

Model quality assessment

Questions to answer

Consider a predictor $\mathcal{F}$ trained to solve a time-series forecasting problem.



1. Is the predictor optimal for the problem at hand?
2. Where does the predictor appear sub-optimal?

Remark: Multiple optimality criteria can be considered.

Performance at task

Given two predictors $\mathcal{F}_a, \mathcal{F}_b$ and performance metric M (e.g., MAE, MSE).

- we consider $\mathcal{F}_a$ better than $\mathcal{F}_b$ if $M(\mathcal{F}_a)$ is *statistically* better than $M(\mathcal{F}_b)$.
- we consider $\mathcal{F}_a$ optimal if there is no other model $\mathcal{F}_b$ better than $\mathcal{F}_a$.

Can we further improve over the best model so far $\mathcal{F}_a$?

$\rightarrow$ Either we find a new model $\mathcal{F}_*$ better than $\mathcal{F}_a$

$\rightarrow$ or we need prior knowledge about the modeled system.

Model	M
$\mathcal{F}_a$	$0.145_{\pm 0.002}$
$\mathcal{F}_b$	$0.176_{\pm 0.005}$
$\vdots$	
$\mathcal{F}_n$	$0.158_{\pm 0.004}$
$\mathcal{F}_*$	$0.139_{\pm 0.001}$

Residual correlation analysis

What correlated resduals tell us

Studying the correlation between prediction residuals $\mathbf{r}_t^i \doteq \mathbf{y}_t^i - \hat{\mathbf{y}}_t^i$ allows for testing model optimality.

If residuals are dependent
$\implies$ there is information that the model hasn't captured
$\implies$ model predictions can be improved.

Remarks: Residual correlation analysis

- ☺ Is independent of specific performance measures.
- ☹ Does not quantify how much a model can improve w.r.t. a specific performance metric.
- ☺ Does not rely on comparisons with other models.

Research focused mainly on either serial correlation (DW test, autocorrelation, etc. [1]–[3]) or spatial correlation (Moran's I, Geary's C, etc. [4], [5]).

Autocorrelation of residuals

Let residuals be $r_t^i = y_t^i - \hat{y}_t^i$.

For each node i, the lag-τ residual **autocorrelation** is

$$\rho_i(\tau) = \mathrm{Corr}\left(r_t^i, r_{t-\tau}^i\right) = \frac{\mathbb{E}\left[(r_t^i - \mathbb{E}[r_t^i])(r_{t-\tau}^i - \mathbb{E}[r_{t-\tau}^i])\right]}{\mathrm{Var}(r_t^i)}.$$

Durbin–Watson Tests for lag-1 serial correlation

$$d = \frac{\sum_{t=2}^{T}(r_t^i - r_{t-1}^i)^2}{\sum_{t=1}^{T}(r_t^i)^2} \approx 2\big(1 - \rho_i(1)\big).$$

$d \approx 2$: no autocorrelation; $d \ll 2$: positive; $d \gg 2$: negative.

Ljung–Box Portmanteau test over multiple lags simultaneously

$$Q(m) = T(T+2)\sum_{\tau=1}^{m}\frac{\hat{\rho}_i(\tau)^2}{T-\tau} \sim \chi^2(m).$$

Spatial residual analysis: Moran's I and LISA

At a fixed time t, let the residuals be r_t^i for all nodes $i = 1, \ldots, N$ and let $\mathbf{W} = [w_{ij}]$ be a spatial weights matrix (like a weighted adjacency matrix).

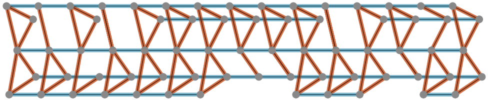
Global spatial association (Moran's I)

$$I_t = \frac{\sum_{(i,j)\in\mathcal{E}} w_{ij}(r_{t,i} - \bar{r}_t)(r_{t,j} - \bar{r}_t)}{\sum_{(i,j)\in\mathcal{E}} w_{ij}} \frac{N}{\sum_{i=1}^{(} r_{t,i} - \bar{r}_t)^2}.$$

Local indicators of spatial association (LISA)

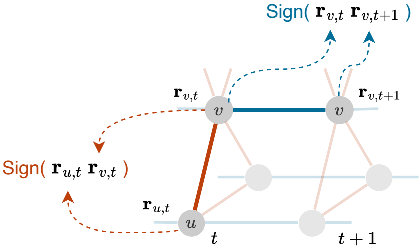
$$I_{t,i} = \frac{\sum_{j\in\mathcal{N}(i)} w_{ij}(r_{t,i} - \bar{r}_t)(r_{t,j} - \bar{r}_t)}{\sum_{j\in\mathcal{N}(i)} w_{ij}} \frac{N}{\sum_{j=1}^{N}(r_{t,j} - \bar{r}_t)^2}.$$

AZ-whiteness test: a spatio-temporal test



spatial and temporal edges.

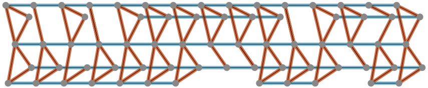
Construct a **multiplex** graph,
and compute **sign changes** for all edges.



$$\text{Sign}(a) = \begin{cases} +1 & a > 0 \\ 0 & a = 0 \\ -1 & a < 0 \end{cases}$$

[6] Zambon and Alippi, “AZ-whiteness Test: A Test for Signal Uncorrelation on Spatio-Temporal Graphs”, NeurIPS 2022.
[7] Geary, “Relative efficiency of count of sign changes for assessing residual autoregression in least squares regression” 1970.
[8] Wald *et al.*, “On a test whether two samples are from the same population” 1940.
[9] Friedman *et al.*, “Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests” 1979.

AZ-Whiteness test statistic



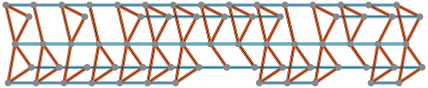
The test is defined by statistic

$$C(\{\mathbf{r}\}) = \underbrace{\sum_{t} \sum_{(i,j)\in\mathcal{E}_t} w_{ijt}\,\mathsf{sgn}(\langle \mathbf{r}_t^i, \mathbf{r}_t^j \rangle)}_{\text{spatial edge}} + \underbrace{\sum_{t} \sum_{i} w_{it}\,\mathsf{sgn}(\langle \mathbf{r}_t^i, \mathbf{r}_{t+1}^i \rangle)}_{\text{temporal edge}} \rightarrow \mathcal{N}(0,1)$$

- ☺ distribution-free and residuals can be non-identically distributed.
- ☺ computation is linear in the number of edges and time steps.

[6] Zambon and Alippi, “AZ-whiteness Test: A Test for Signal Uncorrelation on Spatio-Temporal Graphs”, NeurIPS 2022.

Assess model optimality



Model	MAE	AZ-test spatiotemporal	AZ-test temporal	AZ-test spatial
Optimal Predictor	0.319			
Model A	0.385	4.8 $_{0.000}$	8.7 $_{0.000}$	**-1.9** $_{\mathbf{0.057}}$
Model B	0.328	**-0.2** $_{\mathbf{0.777}}$	**-0.7** $_{\mathbf{0.428}}$	**0.3** $_{\mathbf{0.696}}$

Test statistic and associated p-value: Stat $_{p\text{-val}}$.

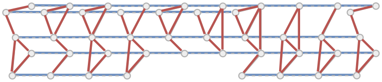
Test on synthetic forecasting problem.

[6] Zambon and Alippi, "AZ-whiteness Test: A Test for Signal Uncorrelation on Spatio-Temporal Graphs", NeurIPS 2022.

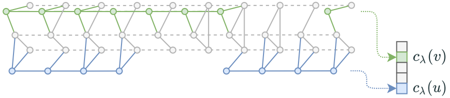
Where can we improve?

Analyzing the AZ-whiteness test statistic computed on subgraphs of the spatio-temporal graph allows for discovering insightful correlation patterns.

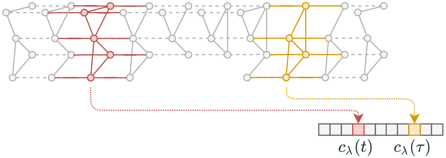
Spatial (or temporal) edges



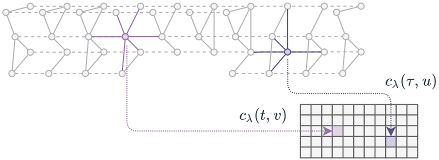
Edges related to a node



Edges related to a time step

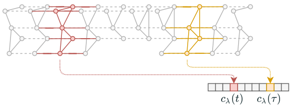


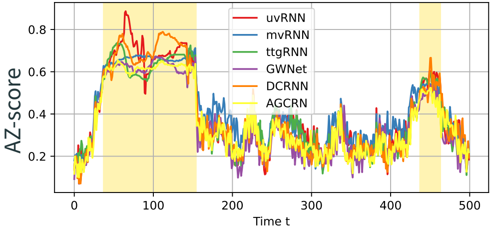
Edges related to a node and time step

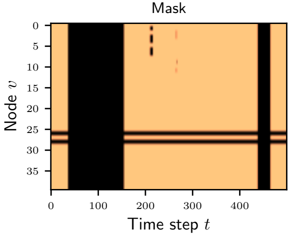


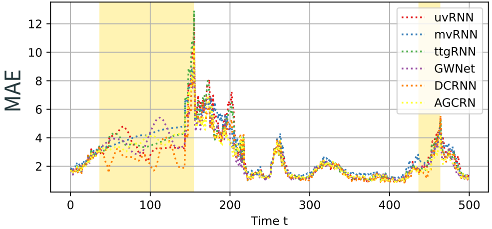
[10] Zambon and Alippi, "Assessment of Spatio-Temporal Predictors in the Presence of Missing and Heterogeneous Data", Neurocomputing 2026.

Use case in traffic forecasting









[10] Zambon *et al.*, "Assessment of Spatio-Temporal Predictors in the Presence of Missing and Heterogeneous Data", Neurocomputing

Key takeaways

- Residual correlation analysis provides complementary insights beyond traditional performance metrics.
- Incorporating relational biases improves the effectiveness of correlation detection.
- The presented approach is well-suited for real-world applications due to its mild assumptions.

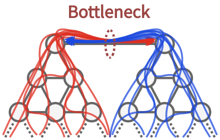
Module

Future Directions

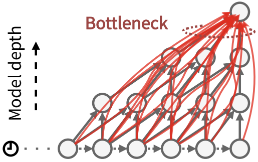
Theoretical insights on model design

GNNs suffer from over-squashing, where information is compressed and lost through bottlenecks.

STGNNs introduce a temporal dimension, compounding the issue [11], [12].



Over-squashing in GNNs



Over-squashing in TCNs

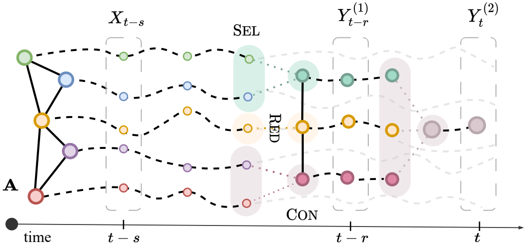
How does spatiotemporal over-squashing affect representation learning in STGNNs?

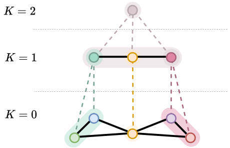
One approach to study it is through the spectral norm of the STGNN's Jacobian:

$$\left\| \nabla_{\substack{u \\ i}} \boldsymbol{h}_t^{v(L)} \right\| = \left\| \frac{\partial \boldsymbol{h}_t^{v(L)}}{\partial \boldsymbol{h}_{t-i}^{u(0)}} \right\|.$$

[13] Marisca, Bamberger, Alippi, and Bronstein, "Over-squashing in Spatiotemporal Graph Neural Networks", NeurIPS 2025.

Hierarchical processing





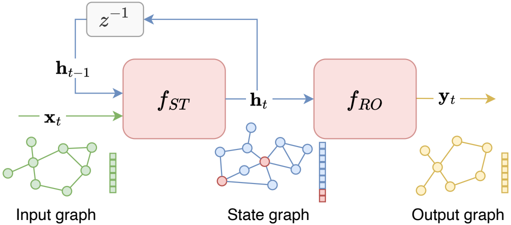
- ☹ Standard STGNNs operate at a fixed spatiotemporal scale.
- 💡 Combine hierarchical and graph-based representations.
- ☺ Exploit higher-order dependencies by operating on hierarchical representations of the input.
- ☺ Can also be used for hierarchical forecasting and to obtain reconciled predictions.

[14] Yu *et al.*, “ST-Unet: A spatio-temporal U-network for graph-structured time series modeling” 2019.
[15] Cini *et al.*, “Graph-based Time Series Clustering for End-to-End Hierarchical Forecasting”, ICML 2024.
[16] Marisca *et al.*, “Graph-based Forecasting with Missing Data through Spatiotemporal Downsampling”, ICML 2024.

State-space models

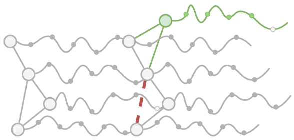
$$\begin{cases} \mathbf{h}_t = f_{\mathsf{ST}}(\mathbf{h}_{t-1}, \mathbf{x}_{t-1}, \boldsymbol{\eta}_{t-1}) \\ \mathbf{y}_t = f_{\mathsf{RO}}(\mathbf{h}_t, \boldsymbol{\nu}_t) \end{cases}$$

- Inputs $\mathbf{x}_t$, states $\mathbf{h}_t$, and outputs $\mathbf{y}_t$ are different attributed graphs.
- $\boldsymbol{\eta}_t, \boldsymbol{\nu}_t$ are noise terms at the node/edge level.



[17] Rangapuram *et al.*, "Deep State Space Models for Time Series Forecasting", NeurIPS 2018.
[18] Zambon *et al.*, *Graph State-Space Models*, Preprint 2023.
[19] Alippi *et al.*, *Graph Kalman Filters*, Preprint 2023.
[20] Buchnik *et al.*, "GSP-kalmannet: Tracking graph signals via neural-aided Kalman filtering", IEEE TSP 2024.
[21] Chouzenoux *et al.*, "Sparse graphical linear dynamical systems", JMLR 2024.

Inductive learning



In real-world applications, one often needs to

- operate under changes in the network connectivity
- make predictions for newly added nodes
- transfer the model to different sensor networks (collections of time series)

Useful in several tasks, like, forecasting, missing data imputation, and virtual sensing.

⚠ Performance can easily degrade if the data distribution of target nodes

- deviates from that at training nodes
- changes over time.

[22] Cini *et al.*, "Taming Local Effects in Graph-based Spatiotemporal Forecasting", NeurIPS 2023.
[23] Yin *et al.*, "Nodetrans: A graph transfer learning approach for traffic prediction", Preprint 2022.
[24] Prabowo *et al.*, "Traffic forecasting on new roads using spatial contrastive pre-training (SCPT)" 2024.

What have we learned?

Deep Learning for **time series** + Deep Learning on **graphs**

- Relational inductive biases allow for exploiting dependencies among the time series,
- ...while sharing most of the model parameters,
- ...and overcoming limits due to irregularities in time and space.
- Whenever possible, global-local models are a safe starting point.

Challenges. Scalability • Missing data • Latent graph learning • Model quality assessment

Resources. Tutorial paper [25] • Open-source library [26]

Credits to:


Andrea **Cini**


Ivan **Marisca**

[25] Cini, Marisca, Zambon, and Alippi, "Graph Deep Learning for Time Series Forecasting", ACM CSUR 2025.
[26] Cini and Marisca, *Torch Spatiotemporal*, https://github.com/TorchSpatiotemporal/tsl 2022.

THANK YOU!

Course feedback form

Daniele **Zambon**

Tutorial paper: **Graph deep learning for time series forecasting**,
Cini, Marisca, Zambon, and Alippi. ACM Computing surveys 2025.
Available Open Access: doi.org/10.1145/3742784

dzambon.github.io · daniele.zambon@usi.ch · gmlg.ch

References i

[1] J. Hosking, **“Equivalent Forms of the Multivariate Portmanteau Statistic,”** *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 43, no. 2, pp. 261–262, 1981.

[2] Z. Li, C. Lam, J. Yao, and Q. Yao, **“On Testing for High-Dimensional White Noise,”** *The Annals of Statistics*, vol. 47, no. 6, pp. 3382–3412, 2019.

[3] A. Bose and W. Hachem, **“A Whiteness Test Based on the Spectral Measure of Large Non-Hermitian Random Matrices,”** in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, pp. 8768–8771.

[4] P. A. P. Moran, **“Notes on Continuous Stochastic Phenomena,”** *Biometrika*, vol. 37, no. 1/2, pp. 17–23, 1950, ISSN: 0006-3444. DOI: 10.2307/2332142.

[5] A. D. Cliff and K. Ord, **“Spatial Autocorrelation: A Review of Existing and New Measures with Applications,”** *Economic Geography*, vol. 46, pp. 269–292, 1970, ISSN: 0013-0095. DOI: 10.2307/143144.

[6] D. Zambon and C. Alippi, **“AZ-whiteness test: A test for signal uncorrelation on spatio-temporal graphs,”** in *Advances in Neural Information Processing Systems*, A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, Eds., 2022.

References ii

[7] R. C. Geary, **“Relative efficiency of count of sign changes for assessing residual autoregression in least squares regression,”** *Biometrika*, vol. 57, no. 1, pp. 123–127, Apr. 1970, ISSN: 0006-3444. DOI: `10.1093/biomet/57.1.123`.

[8] A. Wald and J. Wolfowitz, **“On a test whether two samples are from the same population,”** *The Annals of Mathematical Statistics*, vol. 11, no. 2, pp. 147–162, 1940.

[9] J. H. Friedman and L. C. Rafsky, **“Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests,”** *The Annals of statistics*, pp. 697–717, 1979.

[10] D. Zambon and C. Alippi, **“Assessment of spatio-temporal predictors in the presence of missing and heterogeneous data,”** *Neurocomputing*, vol. 675, p. 132 963, 2026, ISSN: 0925-2312. DOI: `https://doi.org/10.1016/j.neucom.2026.132963`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0925231226003607`.

[11] Y. Bengio, P. Simard, and P. Frasconi, **“Learning long-term dependencies with gradient descent is difficult,”** *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[12] U. Alon and E. Yahav, **“On the bottleneck of graph neural networks and its practical implications,”** in *International Conference on Learning Representations*, 2021.

References iii

[13] I. Marisca, J. Bamberger, C. Alippi, and M. M. Bronstein, **“Over-squashing in spatiotemporal graph neural networks,”** *arXiv preprint arXiv:2506.15507*, 2025. [Online]. Available: https://arxiv.org/abs/2506.15507.

[14] B. Yu, H. Yin, and Z. Zhu, **“ST-Unet: A spatio-temporal U-network for graph-structured time series modeling,”** *arXiv preprint arXiv:1903.05631*, 2019.

[15] A. Cini, D. Mandic, and C. Alippi, **“Graph-based Time Series Clustering for End-to-End Hierarchical Forecasting,”** *International Conference on Machine Learning*, 2024.

[16] I. Marisca, C. Alippi, and F. M. Bianchi, **“Graph-based forecasting with missing data through spatiotemporal downsampling,”** in *Proceedings of the 41st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 235, PMLR, 2024, pp. 34 846–34 865.

[17] S. S. Rangapuram, M. W. Seeger, J. Gasthaus, L. Stella, Y. Wang, and T. Januschowski, **“Deep State Space Models for Time Series Forecasting,”** in *Advances in Neural Information Processing Systems*, vol. 31, Curran Associates, Inc., 2018.

[18] D. Zambon, A. Cini, L. Livi, and C. Alippi, ***Graph state-space models***, Jan. 2023. DOI: 10.48550/arXiv.2301.01741.

References iv

[19] C. Alippi and D. Zambon, ***Graph Kalman Filters*,** Mar. 2023. DOI: 10.48550/arXiv.2303.12021.

[20] I. Buchnik, G. Sagi, N. Leinwand, Y. Loya, N. Shlezinger, and T. Routtenberg, **“Gsp-kalmannet: Tracking graph signals via neural-aided kalman filtering,”** *IEEE Transactions on Signal Processing*, 2024.

[21] E. Chouzenoux and V. Elvira, **“Sparse graphical linear dynamical systems,”** *Journal of Machine Learning Research*, vol. 25, no. 223, pp. 1–53, 2024.

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

[23] X. Yin, F. Li, Y. Shen, H. Qi, and B. Yin, **“Nodetrans: A graph transfer learning approach for traffic prediction,”** *arXiv preprint arXiv:2207.01301*, 2022.

[24] A. Prabowo, H. Xue, W. Shao, P. Koniusz, and F. D. Salim, **“Traffic forecasting on new roads using spatial contrastive pre-training (scpt),”** *Data Mining and Knowledge Discovery*, vol. 38, no. 3, pp. 913–937, 2024.

References v

[25] 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.

[26] A. Cini and I. Marisca, ***Torch Spatiotemporal*,** Mar. 2022. [Online]. Available: https://github.com/TorchSpatiotemporal/tsl.