Graph-based Forecasting with Missing Data through Spatiotemporal Downsampling

Ivan **Marisca**[1] Cesare **Alippi**[1,2] Filippo Maria **Bianchi**[3,4]

[1]IDSIA USI-SUPSI, Università della Svizzera italiana [🇨🇭] [2]Politecnico di Milano [🇮🇹] [3]UiT the Arctic University of Norway [🇳🇴] [4]NORCE, Norwegian Research Centre AS [🇳🇴]

PAPER CODE

Check out our library for STGNNs!
TorchSpatiotemporal/tsl

tl;dr

MOTIVATION

- 🙂 Graph Deep Learning excels in modeling spatial dependencies in time series forecasting.
- ☹️ Missing data affect most real-world applications. (e.g., sensors)
- ☹️ Graph-based predictors deal with complete sequences and need imputation as pre-processing.

CHALLENGES

- **Observability.** Missing data **corrupt dynamics** in the data.
- **Adaptability.** To recover corrupted dynamics, we need **different processing** for **fine-** and **coarse**-grained **scales**.
- **Scalability.** Accounting for **long-range dependencies** in both time and space might cause **efficiency** issues.

GET RID OF IMPUTATIONS, FORECAST DIRECTLY FROM INCOMPLETE TIME SERIES

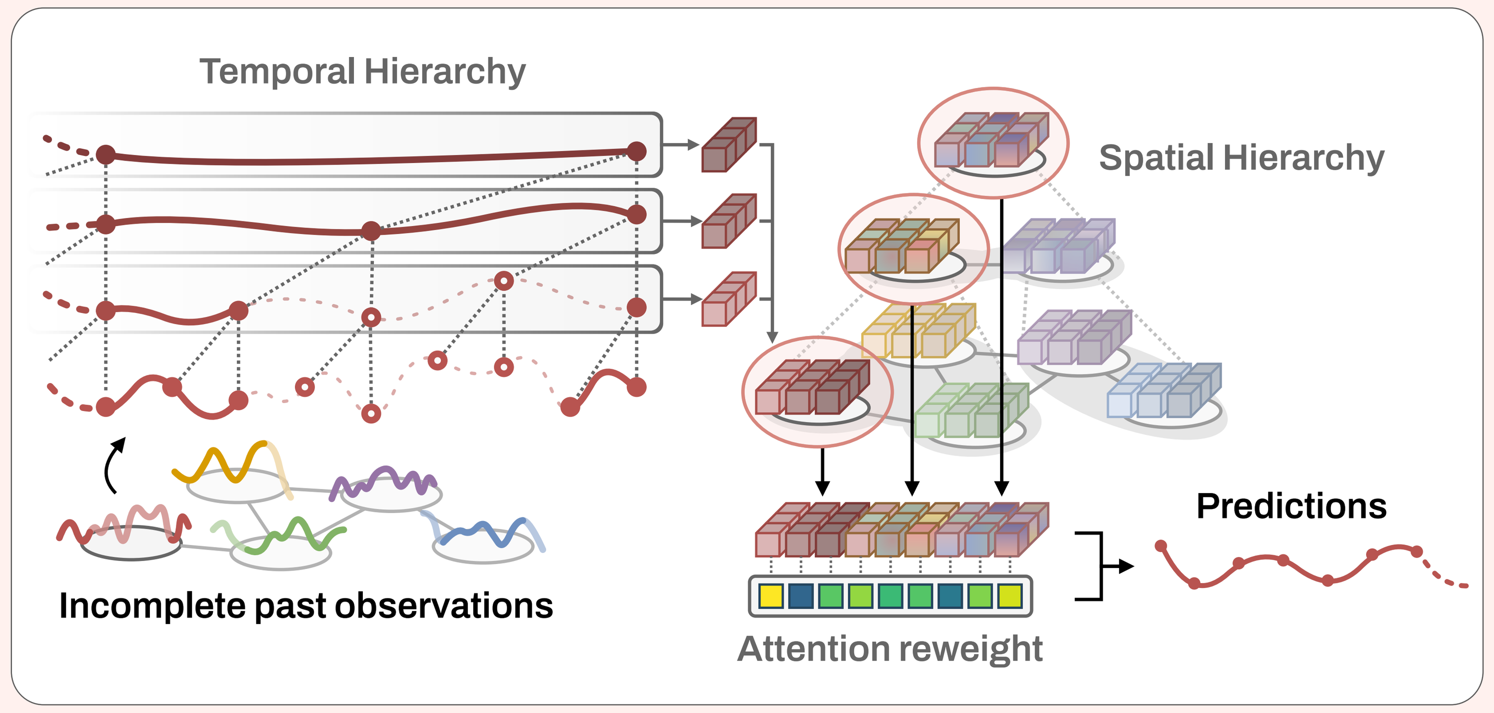
CONTRIBUTION

Obtain a hierarchy of **multi-scale representations**, each accounting for a specific **space-time resolution**.

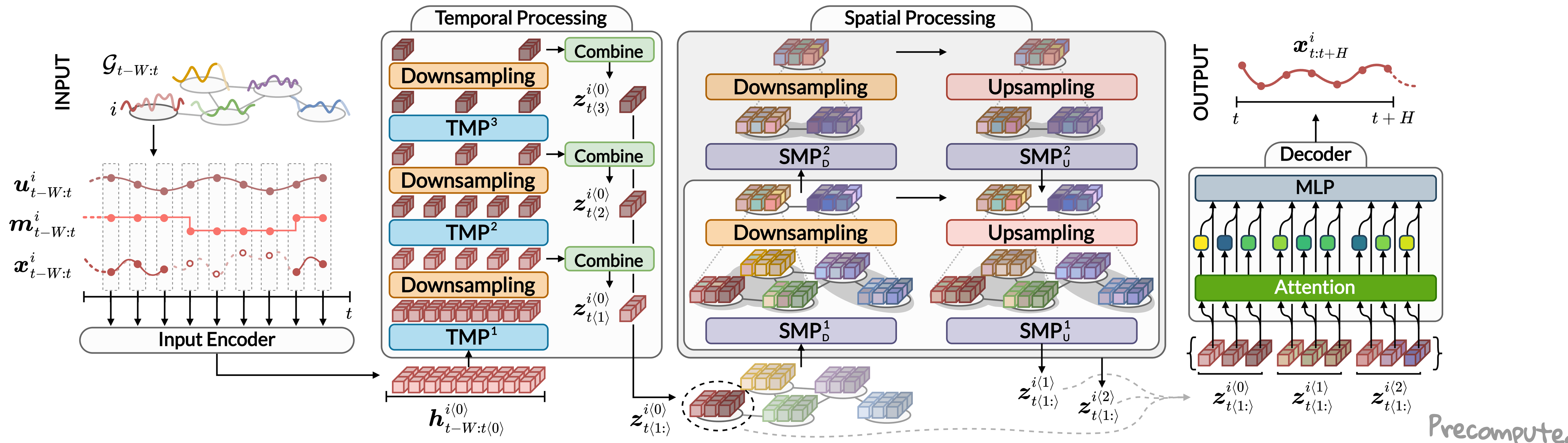
Scalability given by **factorized time-then-space** processing and precomputed **downsampling** operators.

Adaptively weigh the representations according to the **missing data pattern** in the input.

🔍 Gain **insights** on the scales through the **scores**!



HD-TTS : Hierarchical Downsampling Time-Then-Space



Temporal operators

TMP : Temporal Message Passing

RNN, 1D convolution, Transformer

Temporal Downsampling

Dilation or strided operations.

Spatial operators

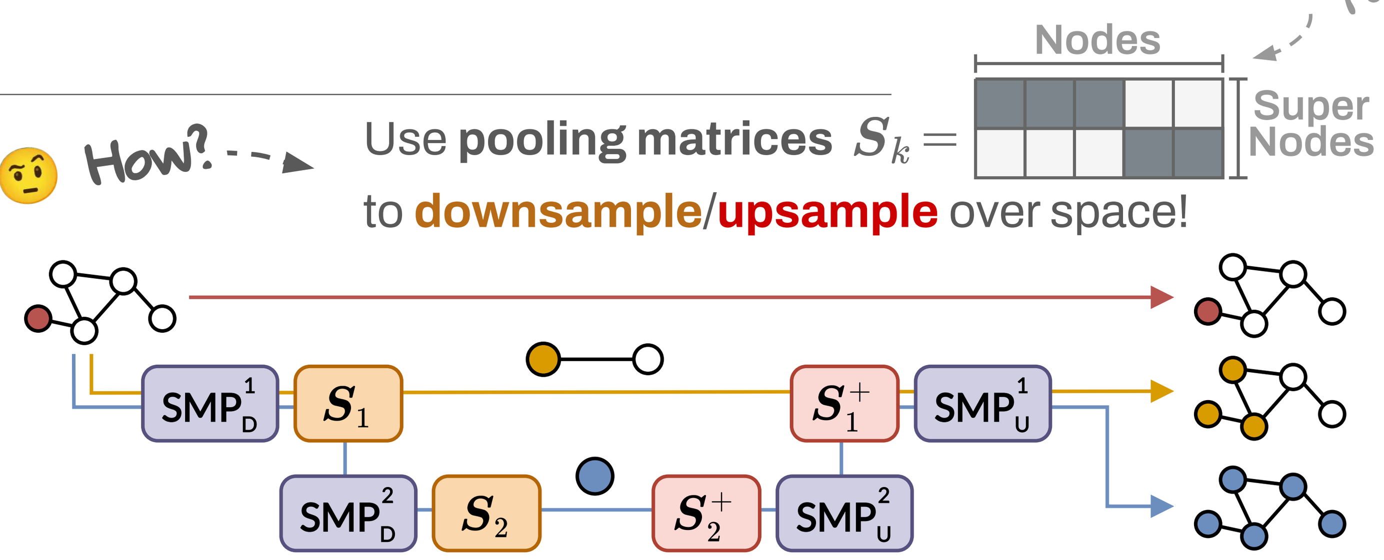
SMP : Spatial Message Passing

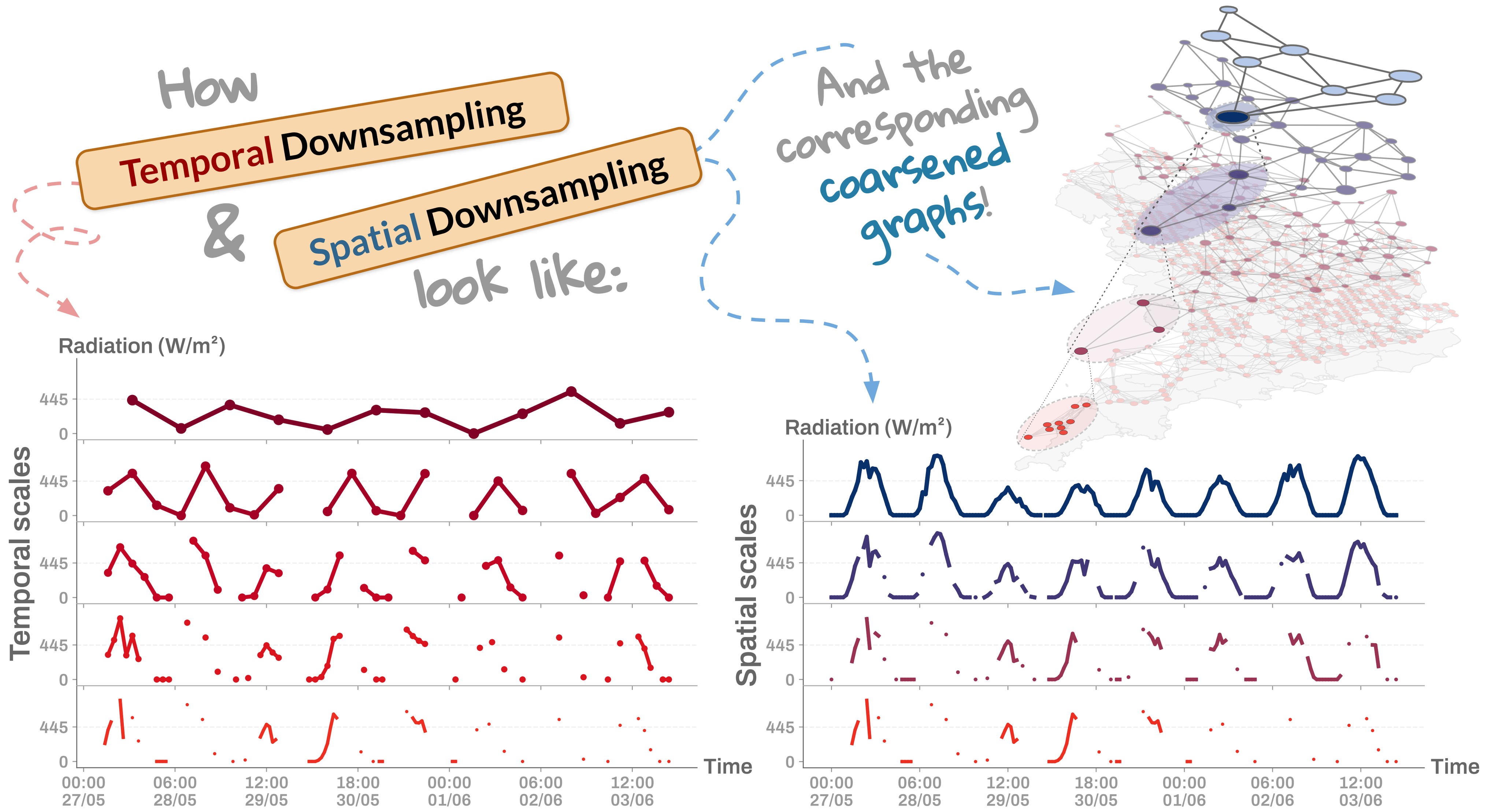
Your friendly neighborhood GNN

Spatial Down-/& Up-/sampling

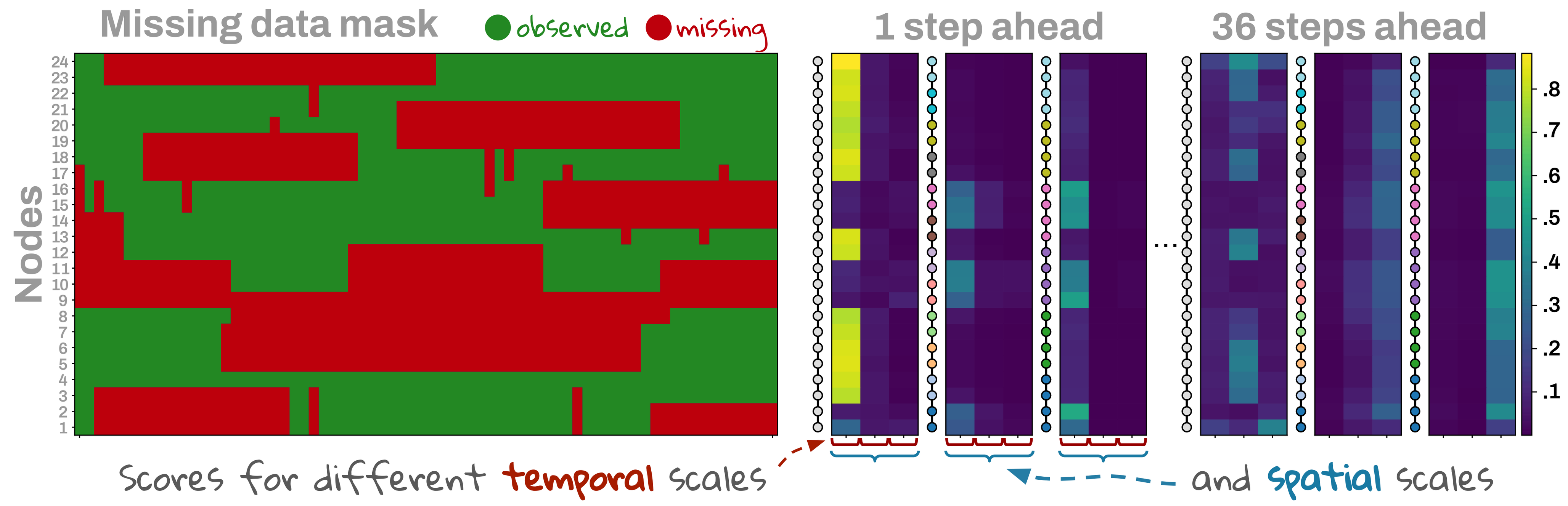
We adopt **graph pooling**.

🤨 How? Use **pooling matrices** S_k = to **downsample**/**upsample** over space!





Interpretability of Decoder Weights



Some Empirical Results

MAE on real-world datasets with different missing data distributions.

	AQI		EngRAD (NEW DATASET!)		PV-US			
Model	**Original**	**+ Point**	**Block–T**	**Block–ST**	**Block–T**	**Block–ST**	**Batch/s**	**GPU RAM**
GRU	$18.17_{\pm 0.03}$	$19.19_{\pm 0.06}$	$5.30_{\pm 0.03}$	$5.42_{\pm 0.02}$	$3.98_{\pm 0.02}$	$4.14_{\pm 0.02}$	$11.59_{\pm 0.04}$	12.01 GB
DCRNN	$16.99_{\pm 0.09}$	$17.51_{\pm 0.08}$	$5.14_{\pm 0.06}$	$5.33_{\pm 0.05}$	$3.54_{\pm 0.01}$	$3.76_{\pm 0.00}$	$1.36_{\pm 0.01}$	19.72 GB
AGCRN	$17.19_{\pm 0.06}$	$17.92_{\pm 0.05}$	$4.84_{\pm 0.01}$	$5.10_{\pm 0.06}$	$4.06_{\pm 0.01}$	$4.20_{\pm 0.04}$	$1.15_{\pm 0.01}$	23.40 GB
GRIN-P	$16.85_{\pm 0.05}$	$17.59_{\pm 0.06}$	$4.91_{\pm 0.04}$	$5.05_{\pm 0.00}$	$3.62_{\pm 0.02}$	$3.85_{\pm 0.07}$	$1.52_{\pm 0.00}$	17.28 GB
GWNet	$15.89_{\pm 0.04}$	$16.39_{\pm 0.14}$	$4.59_{\pm 0.04}$	$4.76_{\pm 0.03}$	$\mathbf{3.48}_{\pm 0.05}$	$3.71_{\pm 0.03}$	$2.12_{\pm 0.00}$	16.02 GB
T&S-IMP	$16.54_{\pm 0.03}$	$17.13_{\pm 0.05}$	$4.98_{\pm 0.01}$	$5.15_{\pm 0.03}$	$3.60_{\pm 0.02}$	$3.82_{\pm 0.03}$	$2.68_{\pm 0.00}$	7.03 GB
T&S-AMP	$16.15_{\pm 0.02}$	$16.58_{\pm 0.10}$	$4.93_{\pm 0.02}$	$5.11_{\pm 0.05}$	N/A	N/A	N/A	N/A
TTS-IMP	$16.25_{\pm 0.01}$	$16.90_{\pm 0.26}$	$4.81_{\pm 0.07}$	$5.08_{\pm 0.04}$	$3.50_{\pm 0.01}$	$3.66_{\pm 0.02}$	$18.84_{\pm 0.14}$	12.81 GB
TTS-AMP	$15.63_{\pm 0.06}$	$16.15_{\pm 0.05}$	$4.70_{\pm 0.00}$	$4.81_{\pm 0.06}$	$\mathbf{3.46}_{\pm 0.03}$	$3.65_{\pm 0.05}$	$14.26_{\pm 0.08}$	12.81 GB
HD-TTS-IMP	$15.50_{\pm 0.07}$	$15.94_{\pm 0.10}$	$\mathbf{4.48}_{\pm 0.01}$	$\mathbf{4.64}_{\pm 0.03}$	$\mathbf{3.47}_{\pm 0.01}$	$\mathbf{3.62}_{\pm 0.02}$	$7.11_{\pm 0.03}$	10.86 GB
HD-TTS-AMP	$\mathbf{15.35}_{\pm 0.01}$	$\mathbf{15.76}_{\pm 0.07}$	$4.53_{\pm 0.03}$	$\mathbf{4.65}_{\pm 0.04}$	$\mathbf{3.47}_{\pm 0.02}$	$\mathbf{3.61}_{\pm 0.02}$	$6.21_{\pm 0.02}$	10.86 GB

- 🙂 Improvements in **forecasting accuracy** and **computational efficiency**.
- 😉 Especially in challenging settings with **blocks of missing values** in space and time.