Leveraging Graph Neural Networks to Forecast Electricity Consumption

Eloi Campagne, Yvenn Amara-Ouali, Yannig Goude, Argyris Kalogeratos

EDF R&D, Centre Borelli (ENS Paris-Saclay)

Context

Motivations

- Maintaining a **balance between electricity supply and demand** is important for **grid stability**
- Providing **accurate forecasts** for short-term electricity load is therefore crucial for all participants in the energy market
- The availability of new **geolocalized and individual electricity consumption data** can be exploited to further **minimize forecasting error**
- **Generalized Additive Models** are used in practice as they are **both performant and explainable**

Graph Neural Networks (1/3)

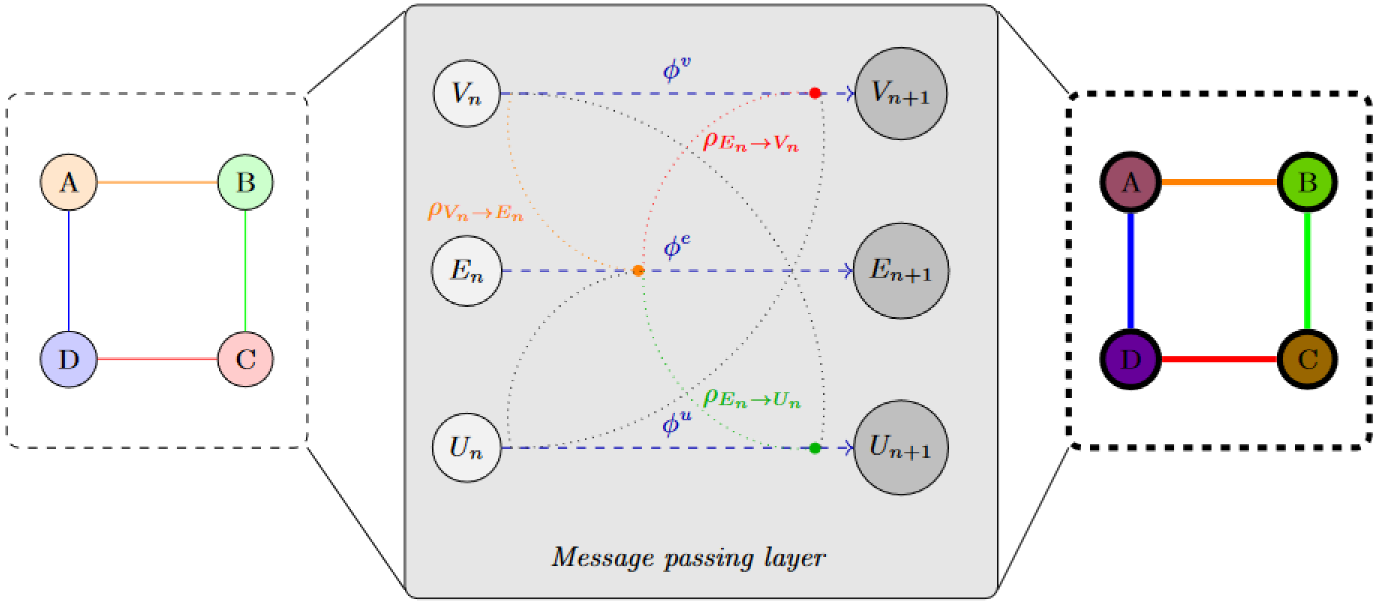


Fig. 1: Example of a message passing layer in a GNN. V_n, E_n and U_n respectively refer to node, edge, and global level at stage n. ϕ are update functions and ρ are propagation functions.

- $V_{n+1} = \phi^v \left(V_n \ ; \ \rho_{E_n \to V_n}, \rho_{U_n \to V_n}\right)$
- $E_{n+1} = \phi^e \left(E_n \ ; \ \rho_{V_n \to E_n}, \rho_{U_n \to E_n}\right)$
- $U_{n+1} = \phi^u \left(U_n \ ; \ \rho_{V_n \to U_n}, \rho_{E_n \to U_n}\right)$

$$h_v^{(\ell)} = \text{UPDATE}^{(\ell)}\left(h_v^{(\ell-1)};\ \text{AGGREGATE}^{(\ell)}\left(h_v^{(\ell-1)}; \left\{h_u^{(\ell-1)} \mid u \in \mathcal{N}_v\right\}\right)\right),$$

$$h_v^{(0)} = \mathbf{X}_v,$$

Graph Neural Networks (2/3)

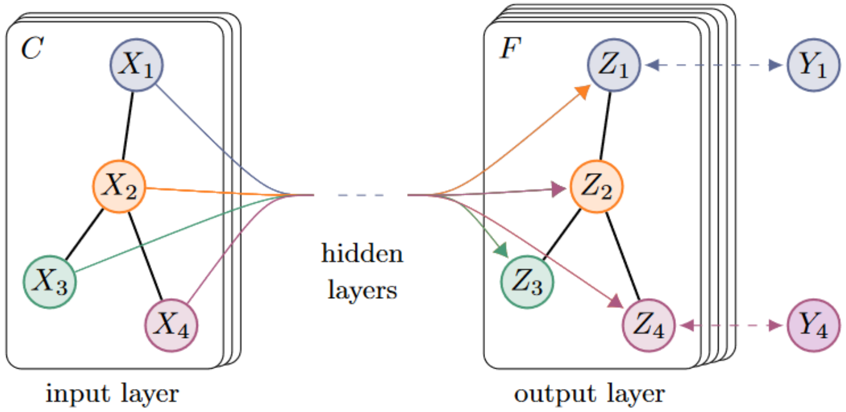
Graph Convolutional Networks (GCNs) – [Kipf T.N., Welling M. (2016)]

$$h_i^{(0)} = \mathbf{X}_i,$$
$$h_i^{(\ell+1)} = \sigma\left(\sum_{j \in \mathcal{N}_i} \frac{1}{c_{ij}} \mathbf{W}^{(\ell)} h_j^{(\ell)} + \mathbf{b}\right)$$

where $c_{ij} = \sqrt{|\mathcal{N}_i|}\sqrt{|\mathcal{N}_j|}$.

$\boldsymbol{W}^{(\ell)}$: learned weight matrix
$\boldsymbol{b}$: learned bias vector
$\mathcal{N}_i$: neighborhood of node v_i
σ: activation function (ReLU)

- GCNs learn **representations** by **aggregating local information** through convolutions



(a) Graph Convolutional Network

Graph Neural Networks (3/3)

SAmple & AGgregatE (SAGE) – [Hamilton W.L. (2018)]

- SAGE **extends** GCNs
- New aggregation rule

$$h_i^{(\ell+1)} = \sigma\left(\mathbf{W}^{(\ell)}\left[h_i^{(\ell)} || \max\left\{\sigma\left(\mathbf{W}_{\text{pool}} h_j^{(\ell)} + \mathbf{b}\right),\ \forall v_j \in \mathcal{N}_{v_i}\right\}\right]\right)$$

- SAGE learns **aggregation functions**

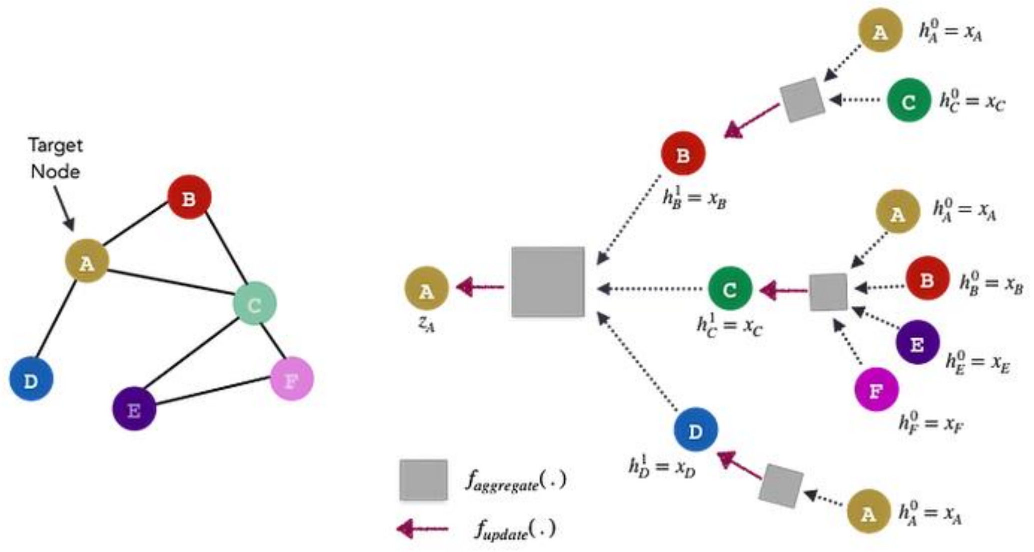


Image taken from **OhMyGraphs: GraphSAGE and Inductive Representation Learning**

Inferring Graphs from Data (1/2)

Geographical Data

- **Similarity matrix** of the geographical positions
- **Physical obstacles** (sea, mountains, etc.) **are not considered**

$$\boldsymbol{W}_\lambda = (\boldsymbol{W}_{i,j})_{1\le i,j\le 12} = \begin{cases} \exp\left\{-\dfrac{\mathbf{dist}(i,j)^2}{\sigma^2}\right\} & \text{if } \exp\left\{-\dfrac{\mathbf{dist}(i,j)^2}{\sigma^2}\right\} \ge \lambda, \\ 0 & \text{otherwise.} \end{cases}$$

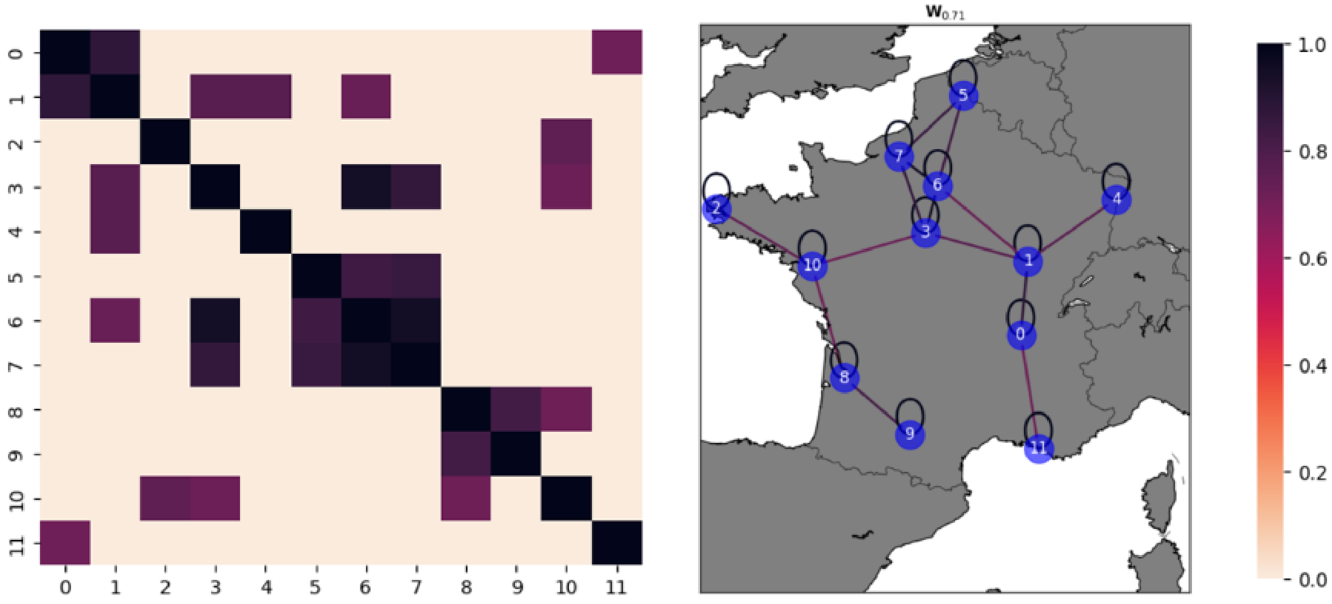


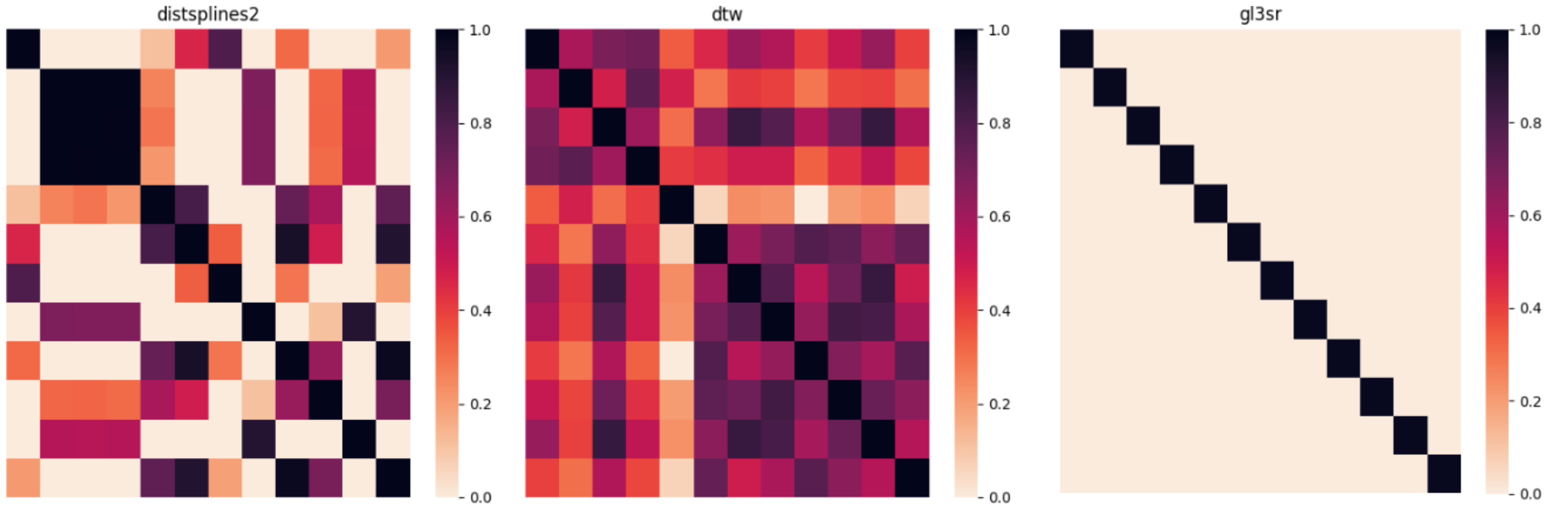
Fig. 3: Graph corresponding to $\boldsymbol{W}_{0.71}$ with $\sigma = 478.3$.

Inferring Graphs from Data (2/2)

Electricity & Weather Data

- **Project first the signal** in d-dimension into a 1-dimensional space !
- **Distance based:** Dynamic Time Warping (DTW), distance between splines
- **Optimization based:** GL3SR

$$\min_{H,\,\mathbf{U},\,\mathbf{\Lambda}} \underbrace{\|X-\mathbf{U}H\|_F^2}_{\text{quadratic approximation error}} + \underbrace{\alpha\|\mathbf{\Lambda}^{1/2}H\|_F^2}_{\text{smoothness regularization}} + \underbrace{\beta\|H\|_S}_{\text{sparsity regularization}} \quad \text{s.t.} \begin{cases} \mathbf{U}^\top\mathbf{U} = I_N,\, x_1 = \frac{1}{\sqrt{N}}\mathbf{1}_N & (a) \\ (\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top)_{i,j} \leq 0,\ i \neq j & (b) \\ \mathbf{\Lambda} = \mathrm{diag}(0, \lambda_2, \ldots, \lambda_N) \succeq 0 & (c) \\ \mathrm{tr}(\mathbf{\Lambda}) = N \in \mathbb{R}_+^* & (d) \end{cases}$$



Datasets (1/2)

Synthetic Datasets

- **Generate temperatures** and **rescale** them

$$T_j^{\mathbf{gen}}(t) = at + b_j(\cos\omega_1 t + \cos\omega_2 t) \qquad \mathbf{b} = (b_j)_{1\leq j\leq 12} \sim \mathcal{N}(\hat{\mu}, \hat{C})$$

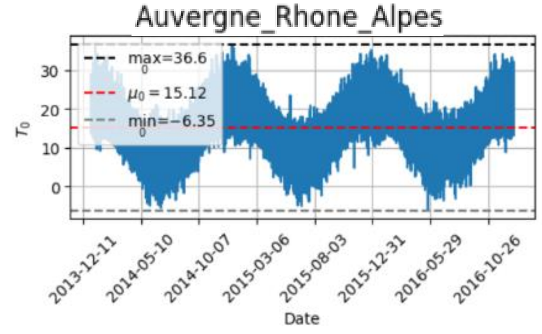
- Train **load splines** on **observed temperatures**

$$\tilde{f}_j \in \arg\min_{f_j} \left(f_j(T_j^{\mathbf{obs}}) - L_j^{\mathbf{obs}}\right)^2 \text{ with } f_j \in \mathbf{span}(s_{j,1}, \ldots, s_{j,k})$$

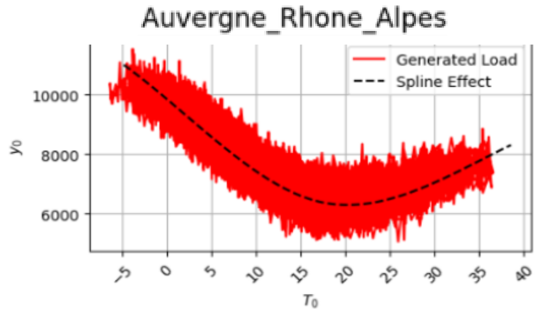
- **Evaluate** the trained splines with the **generated temperatures**

$$L_j^{\mathbf{gen}}(t) = \tilde{f}_j(T_j^{\mathbf{gen}})(t) + \varepsilon_j(t) \text{ where } \boldsymbol{\varepsilon} = (\varepsilon_j)_{1\leq j\leq 12} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

- Two covariance matrices were tested: **correlation on the space graph** and **identity**



(a) Temperature generated in Auvergne-Rhône-Alpes.



(b) Load generated with a cubic spline basis of rank 10 in Auvergne-Rhône-Alpes.

Datasets (2/2)

Real dataset

- **12 administrative regions of France** are considered
- **32 weather stations** (appearing as black dots)
- **Half-hourly** data, **train =** 2014-2018, **test** = 2019

Table 1: Features in the dataset.

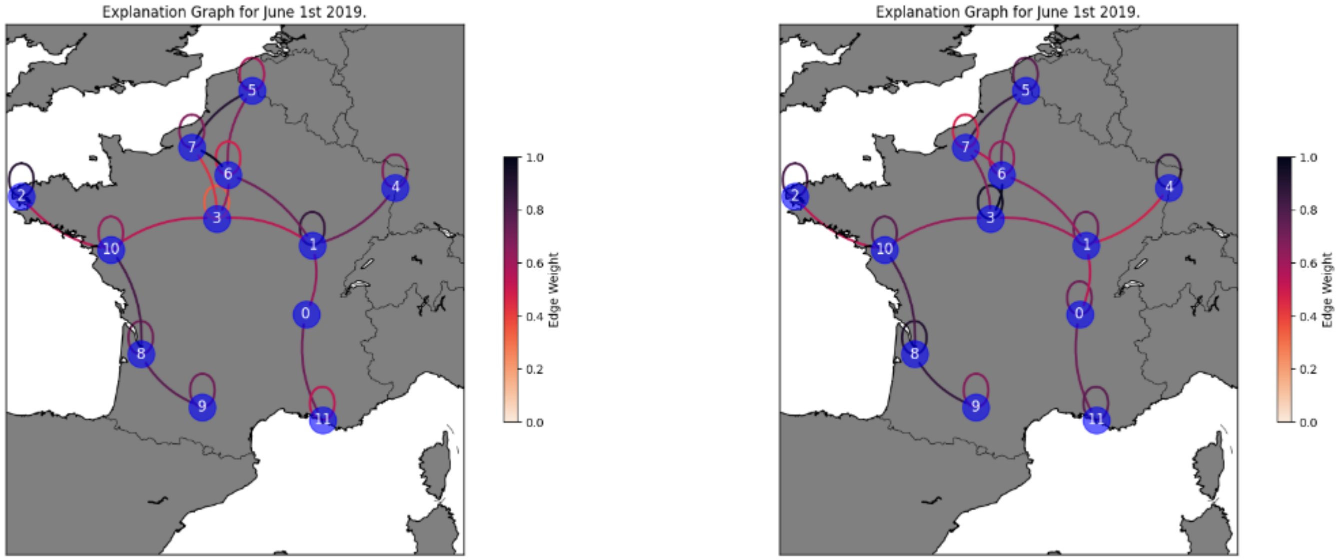
Variable	Definition
Date	Date
Region	Region
Load	Electricity consumption (in MW)
Nebulosity	Cloud cover
Wind	Wind
Temperature	Temperature (in °C)
TempMin, TempMax	Minimal and maximal values of Temp for the day
TempSmoothHigh/Low	Exponentially smoothed temperatures
Instant	Instant in the day
Posan	Position of the day in the year
DayType	Categorical variable indicating the type of the day
Weekend	Categorical variable for the weekend
Summer, Christmas	Categorical variable for summer and Christmas
Holiday_zone	Categorical variable for the other holidays

Explainability (1/2)

GNNExplainer [Ying R., et al. (2019)]

- Pinpoint a **compact subgraph that enhances a GNN's prediction certainty**
- GNNs can highlight links between nodes and **therefore important subgraphs can be extracted**:

$$\max_{\mathcal{G}_S} \ \mathrm{MI}(\boldsymbol{Y}_{\mathcal{G}}, \mathcal{G}_S) = H(\boldsymbol{Y}_{\mathcal{G}}) - H(\boldsymbol{Y}_{\mathcal{G}} \mid \mathcal{G} = \mathcal{G}_S,\ \mathbf{X} = \mathbf{X}_S)$$



(a) Synthetic dataset ($\mathbf{\Sigma} = \rho(\mathbf{W}_\lambda)$).

(b) Real dataset.

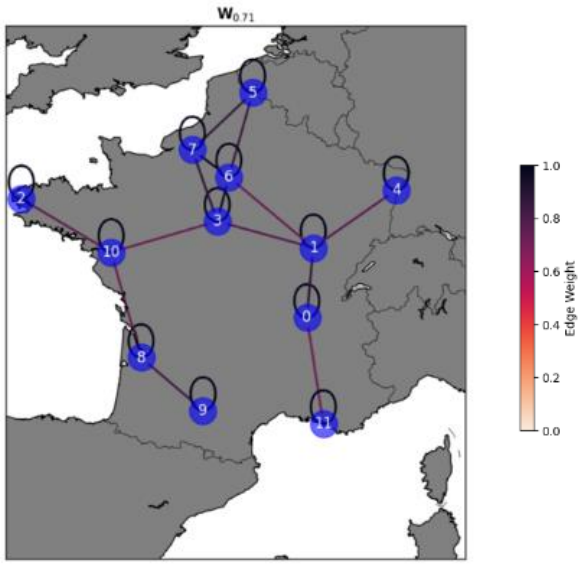


Fig. 10: Explanation graphs in June 2019 obtained from the space matrix.

Explainability (2/2)

Accumulated Local Effects

- **Impact of temperature on load** for each region
- Air conditioning and heating effects represented, but **discrepancy at extreme temperatures**

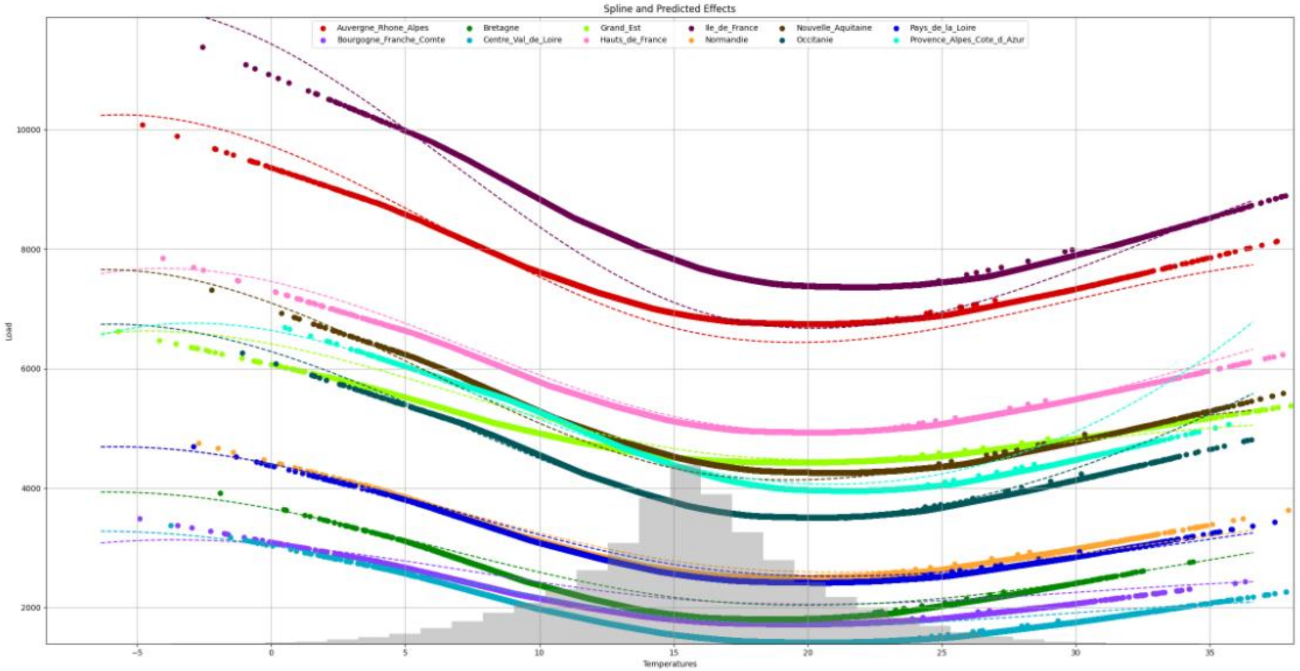


Fig. 11: Spline (dashed line) and predicted (scatter plot) effects. The distribution of the generated temperatures is represented in gray.

Results (1/2)

Table 5: Numerical performance in MAPE (%) and RMSE (MW) at national level on the test set.

Model	Real Dataset MAPE (%)	Real Dataset RMSE (MW)	Synthetic Dataset ($\boldsymbol{\Sigma} = \rho(\mathbf{W}_\lambda)$) MAPE (%)	Synthetic Dataset ($\boldsymbol{\Sigma} = \rho(\mathbf{W}_\lambda)$) RMSE (MW)	Synthetic Dataset ($\boldsymbol{\Sigma} = \boldsymbol{I}$) MAPE (%)	Synthetic Dataset ($\boldsymbol{\Sigma} = \boldsymbol{I}$) RMSE (MW)
GAM-Regions	**1.48**	**1018**	**1.11**	**662**	**1.75**	**1043**
Feed Forward	1.54	1071	3.82	3141	4.49	3213
GCN-`identity`	5.66	3949	1.43	834	2.16	1259
GCN-`space`	2.07	1452	1.26	749	1.98	1169
GCN-`distsplines`	2.04	1404	1.29	764	2.01	1185
GCN-`gl3sr`	5.95	4210	**1.25**	**743**	**1.97**	**1160**
GCN-`dtw`	**1.82**	**1276**	1.26	753	1.99	1171
SAGE-`identity`	4.38	3021	1.25	755	**1.78**	**1066**
SAGE-`space`	1.96	1350	1.29	778	1.85	1112
SAGE-`distsplines`	2.06	1410	1.22	741	1.84	1116
SAGE-`gl3sr`	**1.78**	**1234**	**1.15**	**701**	1.92	1171
SAGE-`dtw`	1.90	1335	1.21	735	1.86	1127
Mixture (Baseline)	1.31	925	1.11	662	1.76	1044
Mixture (GNNs)	1.48	1092	1.12	677	1.98	1171
Mixture (Baseline + GNNs)	**1.13**	**844**	**1.08**	**647**	**1.76**	**1050**

Results (2/2)

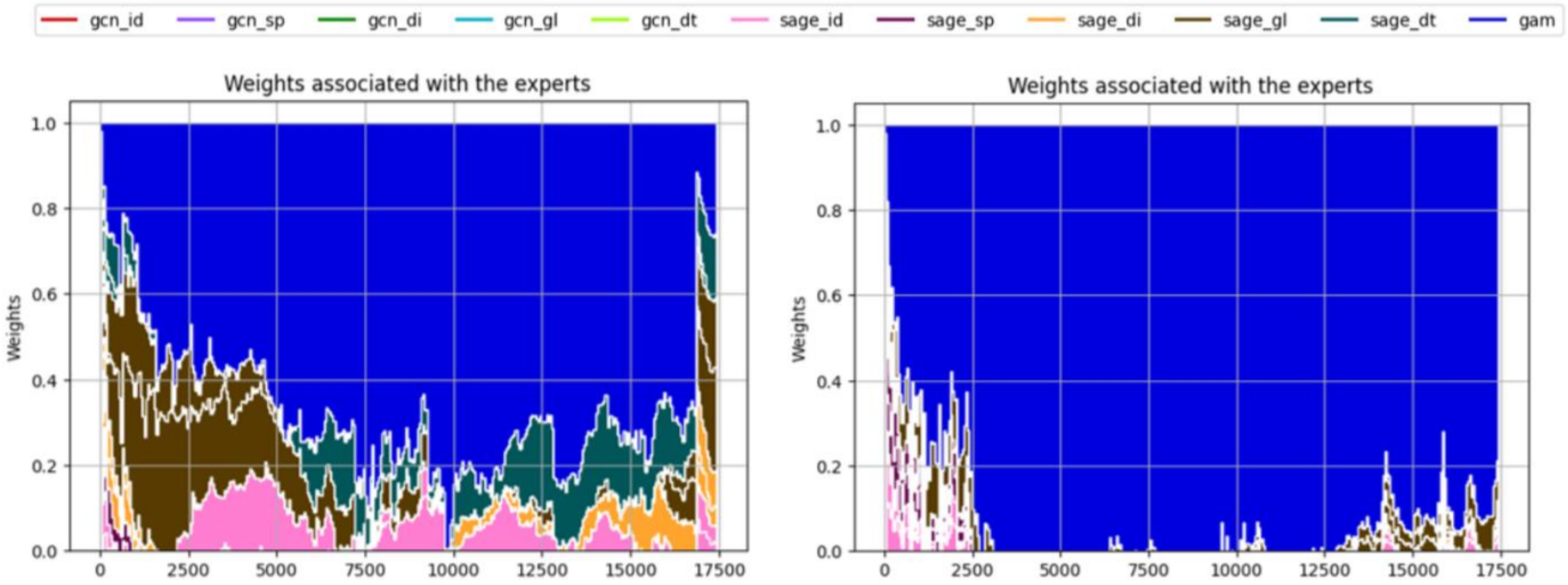


Fig. 6: Weights associated with the experts on the synthetic datasets. $\boldsymbol{\Sigma} = \boldsymbol{\rho}(\mathbf{W}_\lambda)$ (left), $\boldsymbol{\Sigma} = \boldsymbol{I}$ (right). GAM is the main expert, followed by mutliple SAGE-`gl3sr` and SAGE-`dtw`.

Conclusion

- **Expert aggregation** takes advantage of **all the qualities of the different models**
- **GNN models** bring diversity when there is an **underlying graph structure in the data**
- Focusing on **explainability helps to improve models** and **build new graphs**

Perspectives

- Apply models with **attention mechanisms** to the problem of load forecasting
- Make **models by period of the year** (summer/winter)
- Compare the results of **several explainers for greater reliability**
- Add **temporal modules** to the models

References

[1] Agarwal, R., Melnick, L., Frosst, N., Zhang, X., Lengerich, B., Caruana, R., Hinton, G.: Neural Additive Models: Interpretable Machine Learning with Neural Nets (2021)
[2] Alasali, F., Nusair, K., Alhmoud, L., Zarour, E.: Impact of the Covid-19 pandemic on electricity demand and load forecasting. Sustainability (2021)
[3] Antoniadis, A., Gaucher, S., Goude, Y.: Hierarchical transfer learning with applications for electricity load forecasting (2022)
[4] Apley, D.W., Zhu, J.: Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society Series B: Statistical Methodology (2020)
[5] Brégère, M., Huard, M.: Online hierarchical forecasting for power consumption data. International Journal of Forecasting (2022)
[6] Cesa-Bianchi, N., Lugosi, G.: Prediction, learning, and games. Cambridge university press (2006)
[7] Chen, K., Chen, F., Lai, B., Jin, Z., Liu, Y., Li, K., Wei, L., Wang, P., Tang, Y., Huang, J., et al.: Dynamic spatio-temporal graph-based cnns for traffic flow prediction. IEEE Access (2020)
[8] Daigavane, A., Ravindran, B., Aggarwal, G.: Understanding convolutions on graphs. Distill (2021)
[9] Doumèche, N., Allioux, Y., Goude, Y., Rubrichi, S.: Human spatial dynamics for electricity demand forecasting: the case of france during the 2022 energy crisis (2023)
[10] Fan, S., Hyndman, R.J.: Forecast short-term electricity demand using semi-parametric additive model. IEEE (2010)
[11] Fasiolo, M., Wood, S.N., Zaffran, M., Nedellec, R., Goude, Y.: Fast calibrated additive quantile regression. Journal of the American Statistical Association (2021)
[12] Feik, M., Lerch, S., Stühmer, J.: Graph Neural Networks and Spatial Information Learning for Post-Processing Ensemble Weather Forecasts (2024)
[13] Gaillard, P., Goude, Y., Nedellec, R.: Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. International Journal of forecasting (2016)
[14] Gaillard, P., Stoltz, G., Van Erven, T.: A second-order bound with excess losses. In: Conference on Learning Theory. PMLR (2014)
[15] Garreau, D., Jitkrittum, W., Kanagawa, M.: Large sample analysis of the median heuristic (2018)
[16] Gilbert, C., Browell, J., Stephen, B.: Probabilistic load forecasting for the low voltage network: Forecast fusion and daily peaks. Sustainable Energy, Grids and Networks (2023)
[17] Guo, S., Lin, Y., Feng, N., Song, C., Wan, H.: Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: Proceedings of the AAAI conference on artificial intelligence (2019)
[18] Hamilton, W.L., Ying, R., Leskovec, J.: Inductive Representation Learning on Large Graphs (2018)
[19] Hastie, T., Tibshirani, R.: Generalized additive models (with discussion), Statistical Science. (1986)
[20] Humbert, P., Le Bars, B., Oudre, L., Kalogeratos, A., Vayatis, N.: Learning laplacian matrix from graph signals with sparse spectral representation. The Journal of Machine Learning Research (2021)
[21] Hyndman, R.J., et al.: Another look at forecast-accuracy metrics for intermittent demand. Foresight: The International Journal of Applied Forecasting (2006)
[22] Jiang, H., Dong, Y., Dong, Y., Wang, J.: Power load forecasting based on spatial–temporal fusion graph convolution network. Technological Forecasting and Social Change (2024)
[23] Jiang, J., Han, C., Wang, J.: Buaa_bigscity: Spatial-Temporal Graph Neural Network for Wind Power Forecasting in Baidu KDD CUP 2022 (2023)
[24] Jiang, P., Van Fan, Y., Klemeš, J.J.: Impacts of Covid-19 on energy demand and consumption: Challenges, lessons and emerging opportunities. Applied energy
[25] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2014)
[26] Kipf, T.N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks (2017)
[27] LeCun, Y., Bengio, Y., et al.: Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks (1995)
[28] Long, M., Cao, Y., Wang, J., Jordan, M.I.: Learning Transferable Features with Deep Adaptation Networks (2015)
[29] Marino, D.L., Amarasinghe, K., Manic, M.: Building energy load forecasting using deep neural networks. In: Annual conference of the IEEE industrial electronics society (2016)
[30] Massaoudi, M., Abu-Rub, H., Refaat, S.S., Chihi, I., Oueslati, F.S.: Deep Learning in Smart Grid Technology: A Review of Recent Advancements and Future Prospects. IEEE Access (2021)
[31] Obst, D., de Vilmarest, J., Goude, Y.: Adaptive Methods for Short-Term Electricity Load Forecasting During Covid-19 Lockdown in France. IEEE Transactions on Power Systems (2021)
[32] Pierrot, A., Goude, Y.: Short-term electricity load forecasting with generalized additive models. Proceedings of ISAP power (2011)
[33] Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., Bronstein, M.: Temporal graph networks for deep learning on dynamic graphs (2020)
[34] Salvador, S., Chan, P.: FastDTW: Toward accurate dynamic time warping in linear time and space. In: KDD workshop on mining temporal and sequential data (2004)
[35] Shi, H., Xu, M., Li, R.: Deep learning for household load forecasting—A novel pooling deep RNN. IEEE Transactions on Smart Grid (2017)
[36] Thielmann, A., Kruse, R.M., Kneib, T., Säfken, B.: Neural Additive Models for Location Scale and Shape: A Framework for Interpretable Neural Regression Beyond the Mean (2024)
[37] de Vilmarest, J., Browell, J., Fasiolo, M., Goude, Y., Wintenberger, O.: Adaptive probabilistic forecasting of electricity (net-)load. IEEE Transactions on Power Systems (2023)
[38] de Vilmarest, J., Goude, Y.: State-Space Models Win the IEEE DataPort Competition on Post-covid Day-ahead Electricity Load Forecasting (2021)
[39] Williams, S., Short, M.: Electricity demand forecasting for decentralised energy management. Energy and Built Environment (2020)
[40] Wood, S.N.: Generalized additive models: an introduction with R. CRC press (2017)
[41] Wood, S.N., Augustin, N.H.: GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. Ecological modelling (2002)
[42] Wood, S.N., Pya, N., Säfken, B.: Smoothing parameter and model selection for general smooth models. Journal of the American Statistical Association (2016)
[43] Ying, Z., Bourgeois, D., You, J., Zitnik, M., Leskovec, J.: GNNExplainer: Generating explanations for graph neural networks. Advances in neural information processing systems (2019)

Thank you!

Eloi Campagne, Yvenn Amara-Ouali, Yannig Goude, Argyris Kalogeratos

EDF R&D, Centre Borelli (ENS Paris-Saclay)

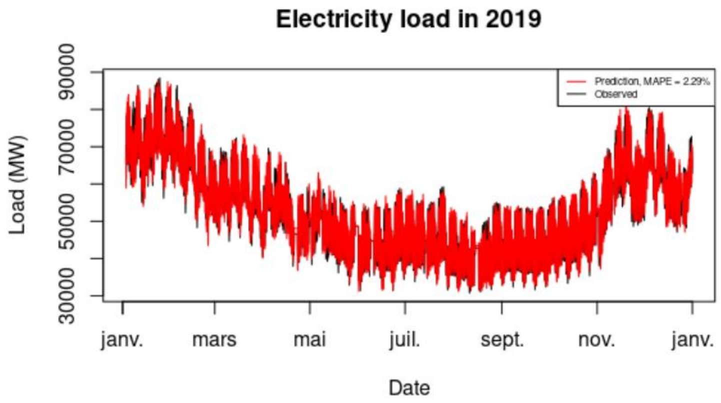
Appendix (1/4)

Generalized Additive Models

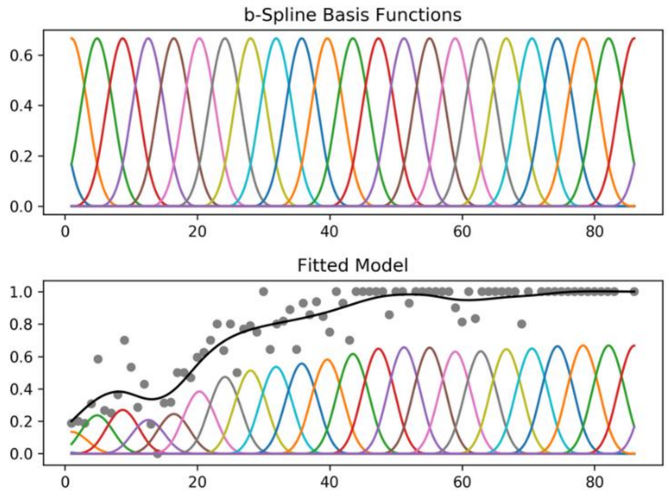
$y_t = \beta_0 + \sum_{j=1}^{d} f_j(x_{t,j}) + \varepsilon_t$ where β_0 is the intercept, and (ε_t) is an i.i.d. random noise.

$$f_j(x) = \sum_{k=1}^{m_j} \beta_{j,k} B_{j,k}(x)$$

with coefficient $\boldsymbol{\beta}_j$ where m_j is the chosen spline basis dimension



(a) Load prediction for 2019 using GAM.



(b) A GAM model and its spline basis.

Appendix (2/4)

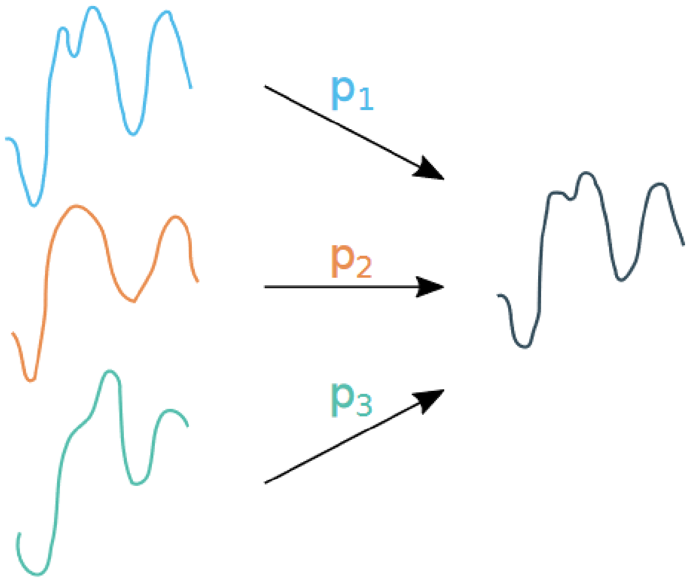
Aggregation of Experts

- Exponentially Weighted Average (EWA)

$$\widehat{p}_{k,t} = \frac{e^{-\eta \sum_{s=1}^{t-1} \ell_s(x_{k,s})}}{\sum_{i=1}^{K} e^{-\eta \sum_{s=1}^{t-1} \ell_s(x_{i,s})}}$$

- Polynomial weighted averages with multiple learning rates (ML-Poly)

  - set $\eta > 0$
  - set initial weights to $p_{j,1} = 1/N$
  - initialize $\widehat{y}_1 = \sum_{j=1}^{N} p_{j,1} f_{j,1}$
  - for $t = 2, \ldots, T$
    - for each expert j, pick the learning rates: $\eta_{j,t-1} = 1/\left(1 + \sum_{s=1}^{t-1} (l(\widehat{y}_s, y_s) - l(f_{j,s}, y_s))^2\right)$
    - update the weights: $p_{j,t} = \eta_{j,t-1} \frac{R_t(\delta_j)^+}{R_t(MLpol)^+}$
    - then aggregation: $\widehat{y}_t = \sum_{j=1}^{N} p_{j,t} f_{j,t}$



$$\hat{y}_t = \sum_k p_{k,t} x_{k,t}$$

Appendix (3/4)

Parametrization

Table 3: Hyperparameters of the GNN models for the synthetic dataset ($\boldsymbol{\Sigma} = \boldsymbol{\rho}(\mathbf{W}_\lambda)$).

Model	Graph structure	batch_size	n_layers	hidden_channels	n_epochs	# parameters	Training time
GCN	Identity	512	3	50	127	2701	$\sim 390s$
GCN	Space	1024	3	64	158	4353	$\sim 480s$
GCN	DistSplines	1024	3	64	168	4353	$\sim 510s$
GCN	GL3SR	1024	3	64	147	4353	$\sim 450s$
GCN	DTW	1024	3	64	146	4353	$\sim 450s$
SAGE	Identity	512	4	50	9	10351	$\sim 50s$
SAGE	Space	512	4	50	12	10351	$\sim 60s$
SAGE	DistSplines	512	4	50	11	10351	$\sim 50s$
SAGE	GL3SR	256	3	50	3	5301	$\sim 10s$
SAGE	DTW	512	4	50	12	10351	$\sim 50s$

Table 4: Hyperparameters of the GNN models for the synthetic dataset ($\boldsymbol{\Sigma} = \boldsymbol{I}$).

Model	Graph structure	batch_size	n_layers	hidden_channels	n_epochs	# parameters	Training time
GCN	Identity	512	3	50	127	2701	$\sim 390s$
GCN	Space	1024	3	64	158	4353	$\sim 480s$
GCN	DistSplines	1024	3	64	168	4353	$\sim 510s$
GCN	GL3SR	1024	3	64	147	4353	$\sim 450s$
GCN	DTW	1024	3	64	146	4353	$\sim 450s$
SAGE	Identity	512	4	50	9	10351	$\sim 50s$
SAGE	Space	1024	3	64	21	8577	$\sim 60s$
SAGE	DistSplines	1024	3	50	6	5301	$\sim 15s$
SAGE	GL3SR	1024	4	50	10	10351	$\sim 40s$
SAGE	DTW	512	4	50	13	10351	$\sim 50s$

Appendix (4/4)

Parametrization