Plugging Attention into Power Grids: Towards Transparent Forecasting

Eloi Campagne[1,3] Itai Zehavi[1,3] Yvenn Amara-Ouali[2,3] Yannig Goude[2,3] Argyris Kalogeratos[1]

[1]Centre Borelli, Ecole Normale Supérieure Paris-Saclay
[2]Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay
[3]EDF R&D

Context and Motivation

- Electricity supply-demand balance is critical for grid stability.
- There is a **lack of models leveraging spatial dependencies** between regions/meters, despite their importance for capturing correlated demand patterns [1].
- Attention-based Graph Neural Networks (GNNs) offer both predictive accuracy and a degree of **interpretability**, which remains scarce in the load forecasting literature.

Problem formulation

Problem Setup.

- $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times d \times T}$: historical node features (n nodes, d features, T time steps).
- $\mathbf{y} \in \mathbb{R}^{n \times T}$: target values.
- $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: graph, $\Phi_{\boldsymbol{\theta}}$: GNN with parameters $\boldsymbol{\theta}$.
- w, h : input and output window sizes, $\mathcal{T}$: set of timesteps to be covered.

Training Objective.

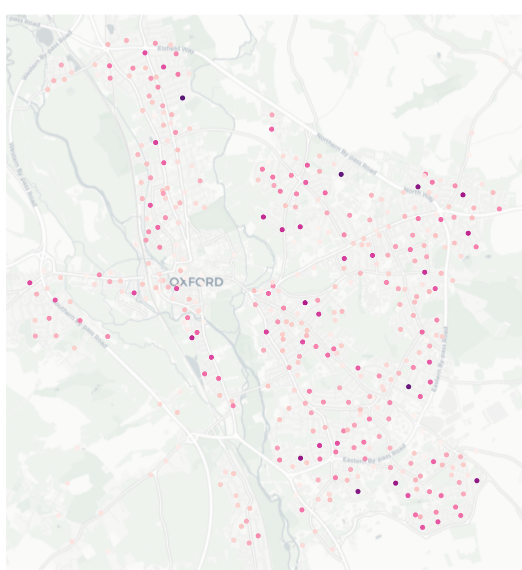
$$\min_{\boldsymbol{\theta}} \sum_{t \in \mathcal{T}} \left\| \Phi_{\boldsymbol{\theta}}(\mathcal{G},\ \tilde{\mathbf{X}}_t \cup \mathbf{y}_{t-1:t-w}) - \mathbf{y}_{t:t+h-1} \right\|^2$$

$$\left\| \hat{\mathbf{y}}_{t:t+h-1} - \mathbf{y}_{t:t+h-1} \right\|^2 = \sum_{i \in \mathcal{V}} \sum_{k=0}^{h-1} (\hat{y}_{i,t+k} - y_{i,t+k})^2$$

Main Contributions

- Systematic GNN benchmark: GCN, GraphSAGE, ChebConv, TAG, APPNP, GAT/GATv2, TransformerConv.
- Attention weights reveal **seasonal clusters** tied to weather regimes.
- Benchmarks on two datasets: **France** (RTE, regional aggregated load), **UK** (Weave, household-level consumption).
- Expert aggregation: uniform, adaptive (ML-Poly), bottom-up.

(a) France.



(b) Oxford Urban Area.

Graph Convolutions

We denote by $\mathcal{N}_v$ the neighborhood of a node v, by $\mathbf{h}_v^{(\ell)}$ the representation vector of node v at layer ℓ, by $\mathbf{A}$ the adjacency matrix, and by $\mathbf{W}$ and $\mathbf{b}$ the learned weight matrices and bias vector, respectively. Different GNN variants can be grouped into families of update rules:

Model	Example	Update rule
General	–	$\mathbf{h}_v^{(\ell+1)} = \mathsf{UPDATE}^{(\ell)}\left(\mathbf{h}_v^{(\ell)},\ \mathsf{AGGREGATE}^{(\ell+1)}\{\mathbf{h}_u^{(\ell)} \mid u \in \mathcal{N}_v\}\right)$
Spatial	GCN, SAGE	$\mathbf{h}_v^{(\ell+1)} = \sigma\left(\sum_{u \in \mathcal{N}_v} c_{uv} \mathbf{W}^{(\ell)} \mathbf{h}_u^{(\ell)} + \mathbf{b}\right)$
Attentional	GAT, GATv2, Transformer	$\mathbf{h}_v^{(\ell+1)} = \sigma\left(\sum_{u \in \mathcal{N}_v} \alpha_{uv}^{(k)} \mathbf{W}^{(k)} \mathbf{h}_u^{(\ell)} + \mathbf{b}\right)$
Spectral	TAG, Cheby	$\mathbf{h}_v^{(\ell+1)} = \sigma\left(\sum_{k=0}^{K} P_k(\mathbf{A}) \mathbf{W}_k \mathbf{h}_v^{(\ell)} + \mathbf{b}\right)$
Propagation	APPNP	$\mathbf{h}_v^{(\ell+1)} = \sigma\left((1-\alpha) \sum_{u \in \mathcal{N}_v} \hat{A}_{uv} \mathbf{h}_u^{(\ell)} + \alpha \mathbf{h}_v^{(0)} + \mathbf{b}\right)$

Numerical Results

Table 1. 'Unif.' indicates uniform averaging of expert predictions, 'Agg.' refers to online aggregation using the MLPol algorithm. 'Bottom' aggregation aggregates at the node level before summation, while 'Top' aggregation first sums expert predictions and then aggregates.

Model	French Dataset MAPE (%)		French Dataset RMSE (MW)		UK Dataset MAPE (%)		UK Dataset RMSE (MW)	
	Unif.	Agg.	Unif.	Agg.	Unif.	Agg.	Unif.	Agg.
GCN	1.21	1.09	895	**839**	8.63	8.51	15.09	14.91
SAGE	1.12	1.14	**873**	898	8.19	8.07	14.87	14.85
GAT	**1.11**	**1.08**	883	871	7.79	8.29	14.13	14.72
GATv2	1.14	1.14	888	902	8.78	8.94	16.44	16.40
Transformer	1.14	1.15	883	910	8.36	8.48	15.32	15.53
TAG	1.14	1.14	884	881	9.23	9.41	15.79	16.19
Cheby	1.16	1.15	878	876	8.60	8.93	17.23	17.37
APPNP	1.22	1.20	944	943	**6.54**	**7.60**	**12.61**	**14.13**
Uniform	1.06	–	836	–	8.10	–	14.74	–
Bottom Aggregation	–	**1.01**	–	**789**	–	**7.68**	–	**14.35**
Top Aggregation	–	1.02	–	817	–	8.33	–	15.04

Key Results

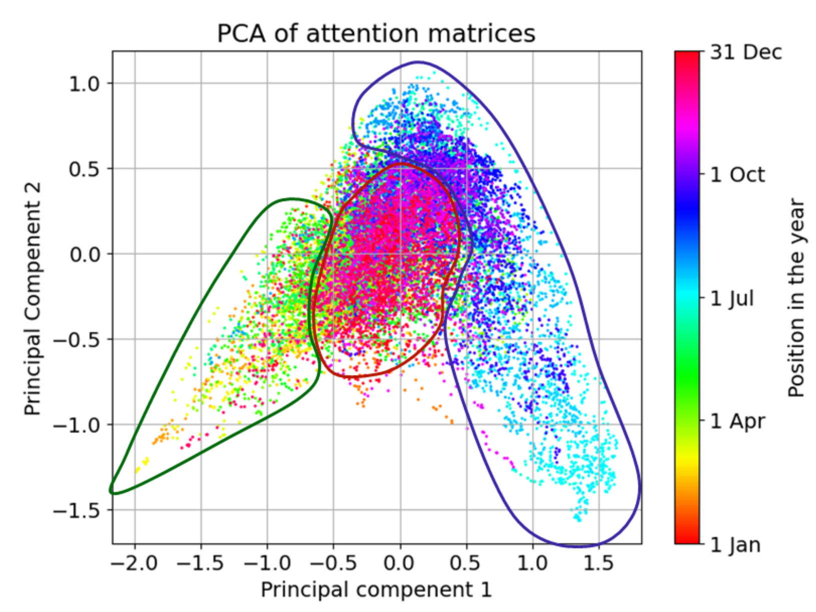
GAT achieves best MAPE, SAGE best RMSE
Simple GCN/TAG remain competitive

APPNP + uniform aggregation performs best
GAT still competitive

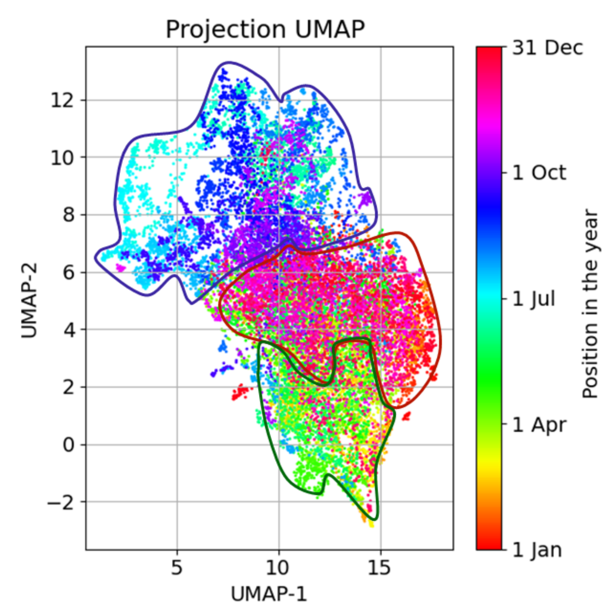
- Baseline models outperformed by GNNs.
- Complex models (GATv2, TransformerConv) do not consistently outperform simpler ones.
- **Bottom-up aggregation** outperforms top strategies, improving robustness and accuracy.

Interpretability via Attention

- Attention weights vary with season and meteorological conditions.
- PCA/UMAP projections reveal clear seasonal clusters (winter, summer, etc.).
- On synthetic datasets: Graph Attention Networks (GATs) recover latent structures even under **noise or ambiguity**.

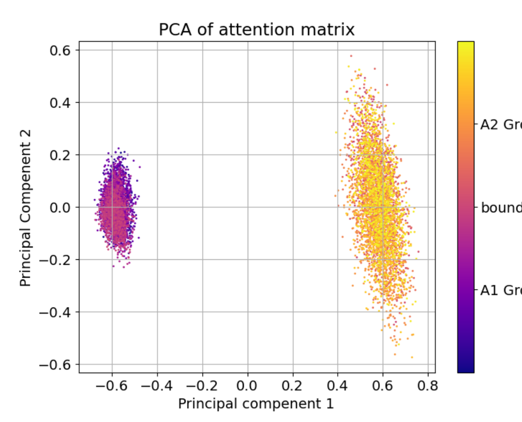


(a) PCA.

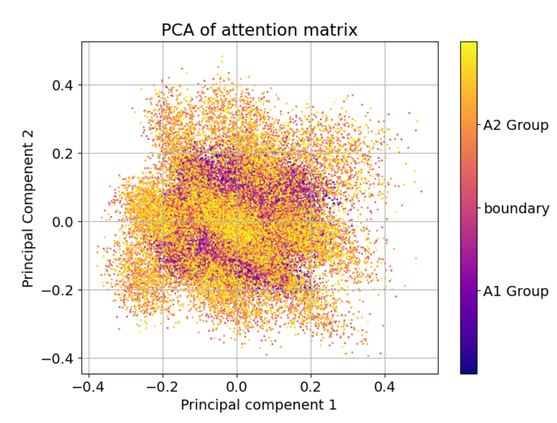


(b) UMAP.

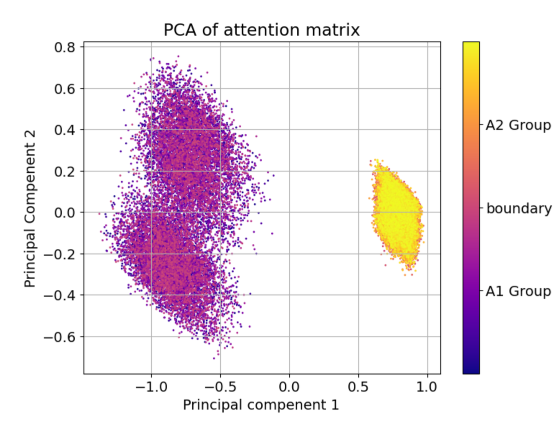
Figure 2. PCA and UMAP projections of attention matrices on the French dataset reveal seasonal clustering: spring (green), summer (blue), and winter (red).



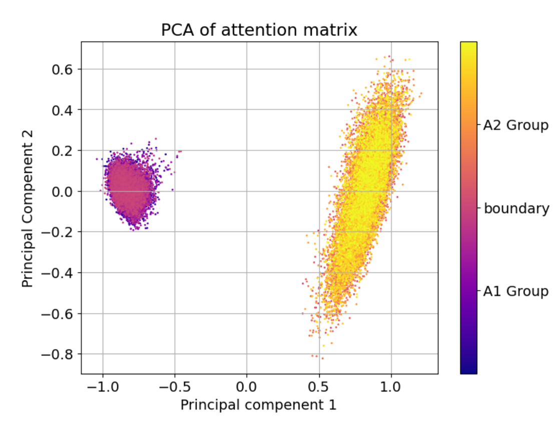
(a) Explicit coupling: Layer 2.



(b) Implicit coupling: Layer 0.



(c) Implicit coupling: Layer 1.



(d) Implicit coupling: Layer 2.

Figure 3. PCA of attention matrices on the synthetic dataset with two coupling modes ($\mathbf{A}_1$ or $\mathbf{A}_2$). Colors reflect the coupling matrix in effect. (a) Explicit coupling information: the mode is explicitly given to the model as a binary input. (b–d) Implicit coupling information: the mode in effect is indicated implicitly by the combined values of two input binary variables. The subfigures show the PCA of attention matrices across GAT layers.

Conclusions and Perspectives

- **Simplicity beats complexity** in low-data, noisy regimes.
- Attention provides both **predictive gains** and **explainability**.
- Expert aggregation enhances **robustness** across settings.
- Future directions: joint learning of graph structure, causal interpretations of attention.

References

[1] Eloi Campagne, Yvenn Amara-Ouali, Yannig Goude, and Argyris Kalogeratos.
Leveraging Graph Neural Networks to Forecast Electricity Consumption.
In *Proceedings of the Machine Learning for Sustainable Power Systems Workshop at ECML PKDD*, Vilnius, Lithuania, 2024.

[2] Eloi Campagne, Yvenn Amara-Ouali, Yannig Goude, and Argyris Kalogeratos.
Plugging Attention into Power Grids: Towards Transparent Forecasting.
In *Proceedings of the Machine Learning for Sustainable Power Systems Workshop at ECML PKDD*, Porto, Portugal, 2025.