Explainability of electricity demand forecasting models: a Shapley value approach for positive component decomposition

Margaux Brégère

MASCOT-NUM - April 2nd 2026

Laboratoire de probabilités statistique & modélisation

Joint work with **Gaspard Berthelier**

XPC

eXplainability through Positive Contributions

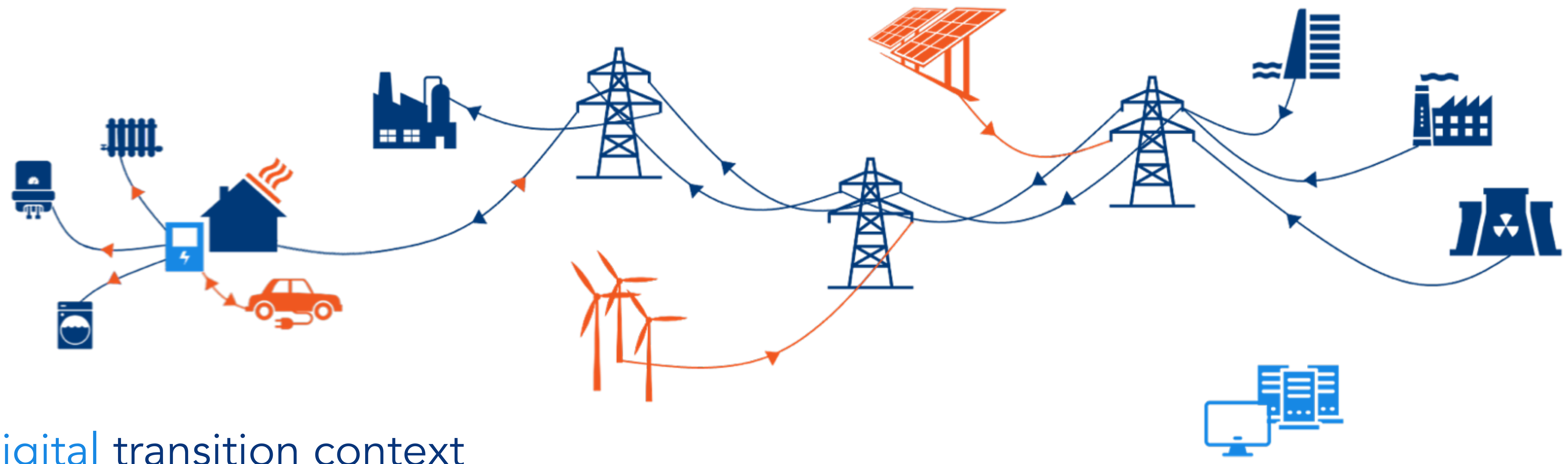
https://github.com/3gaspo/xpc

Challenges raised for a sustainable electricity system

As electricity is hard to store, balance between production and demand must be strictly maintained

Adapt production and Forecast demand

Optimization **Statistics**



The energy and digital transition context

- New uses of electricity and electrification of numerous applications
- Massive development of intermittent renewables
- Increasingly rapid availability of data, smart meters and high-performance computing resources

Raises new challenges

- Changes in electricity demand (energy crisis, sobriety, self-consumption, electric vehicles, increase from the current 450 TWh to 645 TWh according to « Energy Futures 2050 »...)
- Need for electrical flexibilities (from 13 to 17 GW in 2050)
- Explosion of artificial intelligence (increasingly complex and costly models)

As electricity is still hard to store, balance between production and demand must be strictly maintained

Optimize production

Forecast renewables

Forecast demand

Manage electrical flexibilities

→ Need of performative and **explainable** time series forecasting models

To identify model errors for effective decision-making, to interpret « anomalies » and changes in the forecasts / target variable, etc.

Interpretability and Explainability for Electrical Demand Forecasting Machine Learning Models

Interpretability or Explainability?

Interpretable models:

- Inherently understandable
- Humans can understand the prediction based on the model's design alone

Explainable models:

- Black-box models combined with external explanation methods
- Human can reason about factors influencing predictions, via post-hoc methods

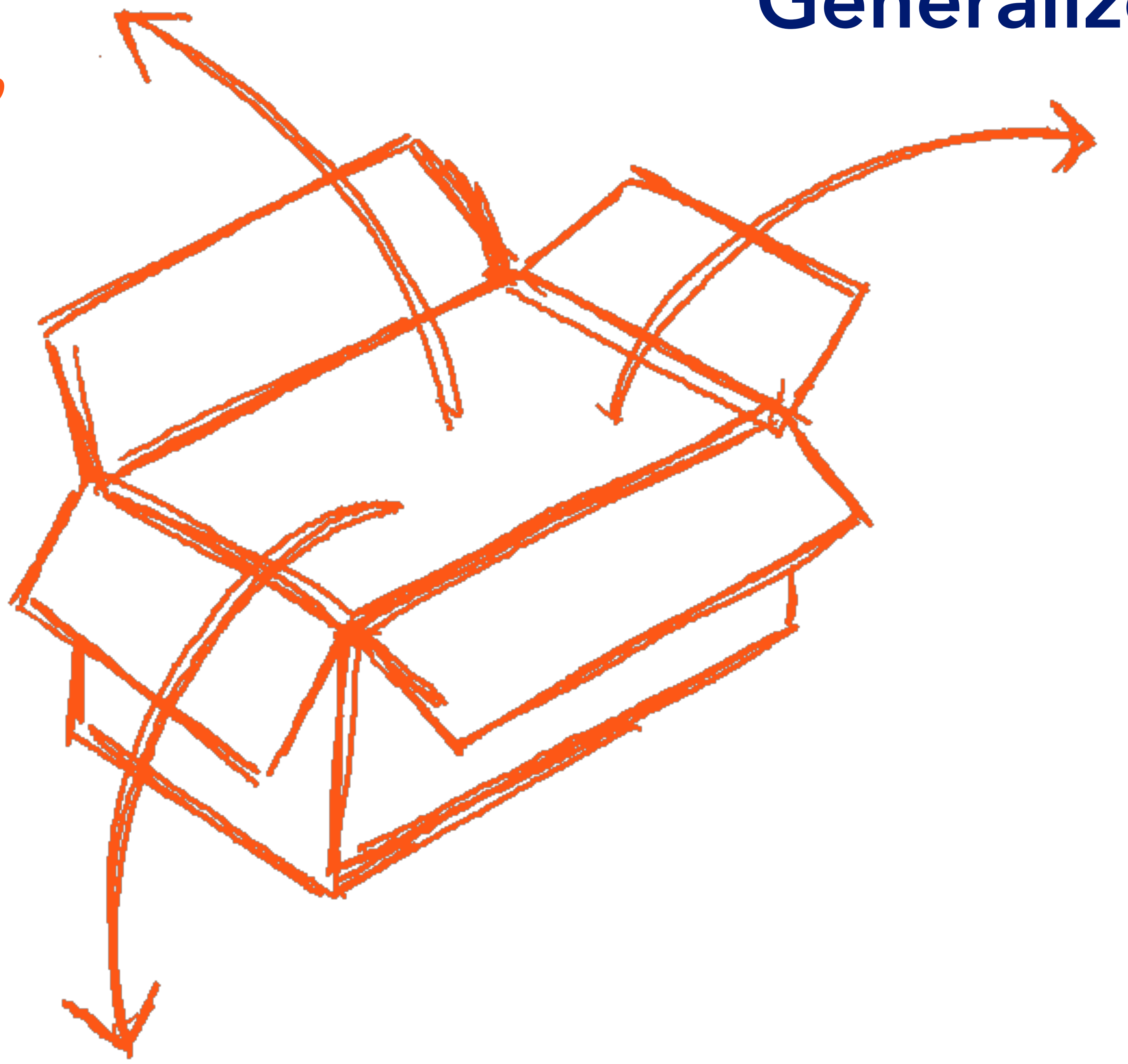
Interpretable approaches

Linear Models $\mathbb{E}[Y] = X^{\mathrm{T}}\beta$

$$\hat{Y} = X^{\mathrm{T}}\hat{\beta} \;\rightarrow\; \hat{\beta}_1, \ldots, \hat{\beta}_p$$

Generalized Additive Models $\mathbb{E}\left[g(Y)\right] = \sum_k f_k(X)$

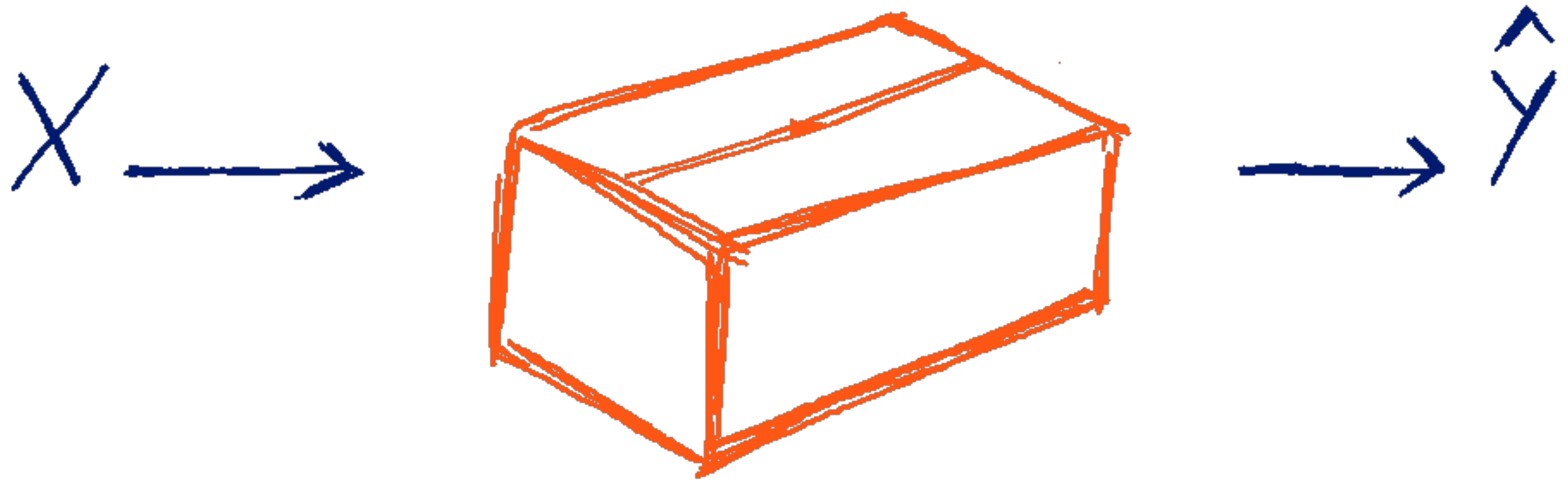
$$\hat{Y} = g^{-1}\left(\sum_k \hat{f}_k(X)\right) \;\rightarrow\; \hat{f}_k\,(X)$$



Classification And Regression Trees

$$\left(\text{variable}_k, \text{threshold}_k\right)_{k \in \text{splits}}$$

Explainable approaches



Post hoc approaches can be

model-specific (limited to specific model classes) or model-agnostic

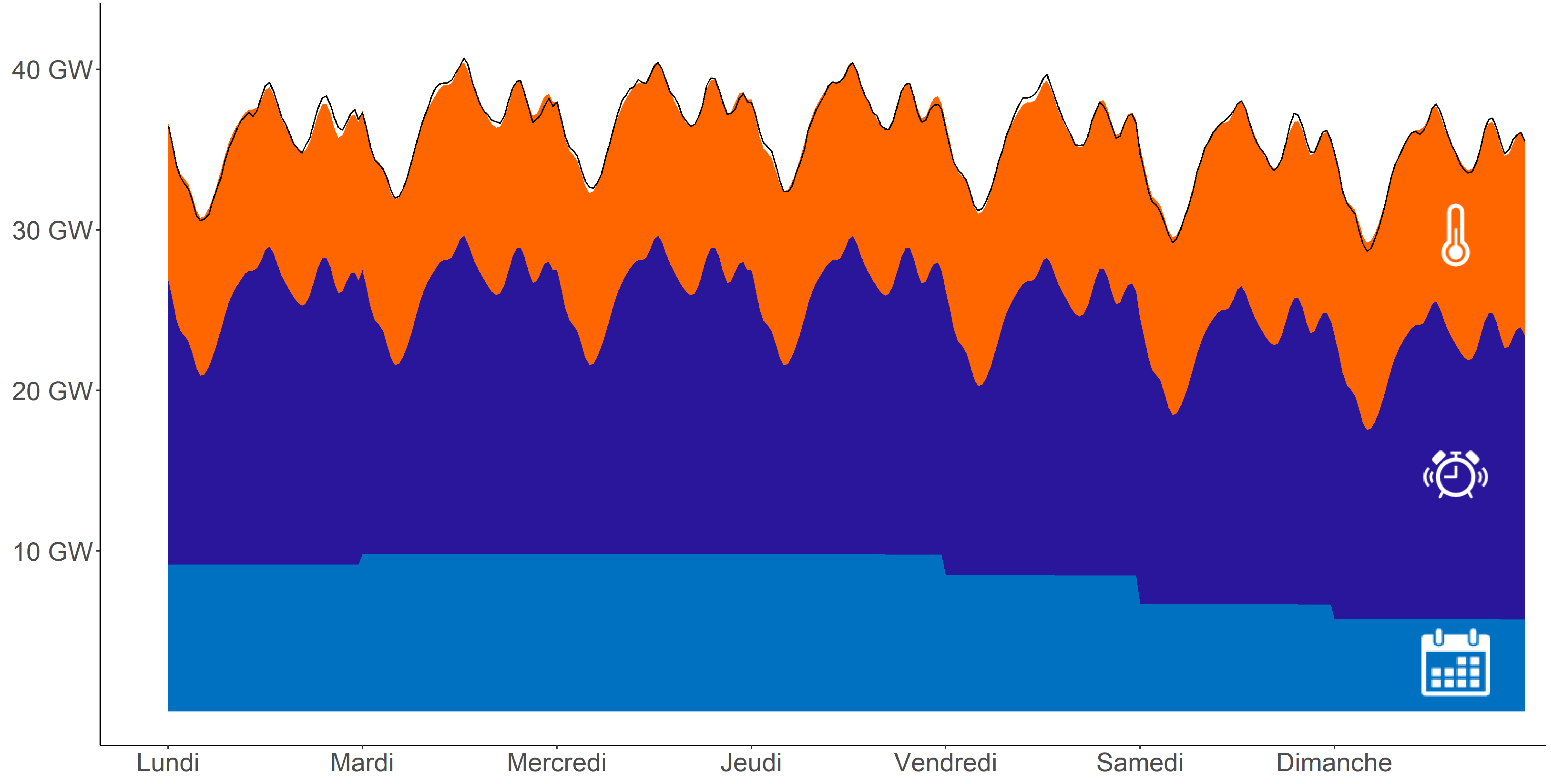
local (explain an individual prediction) or global (explain the entire model behaviour)

and may output

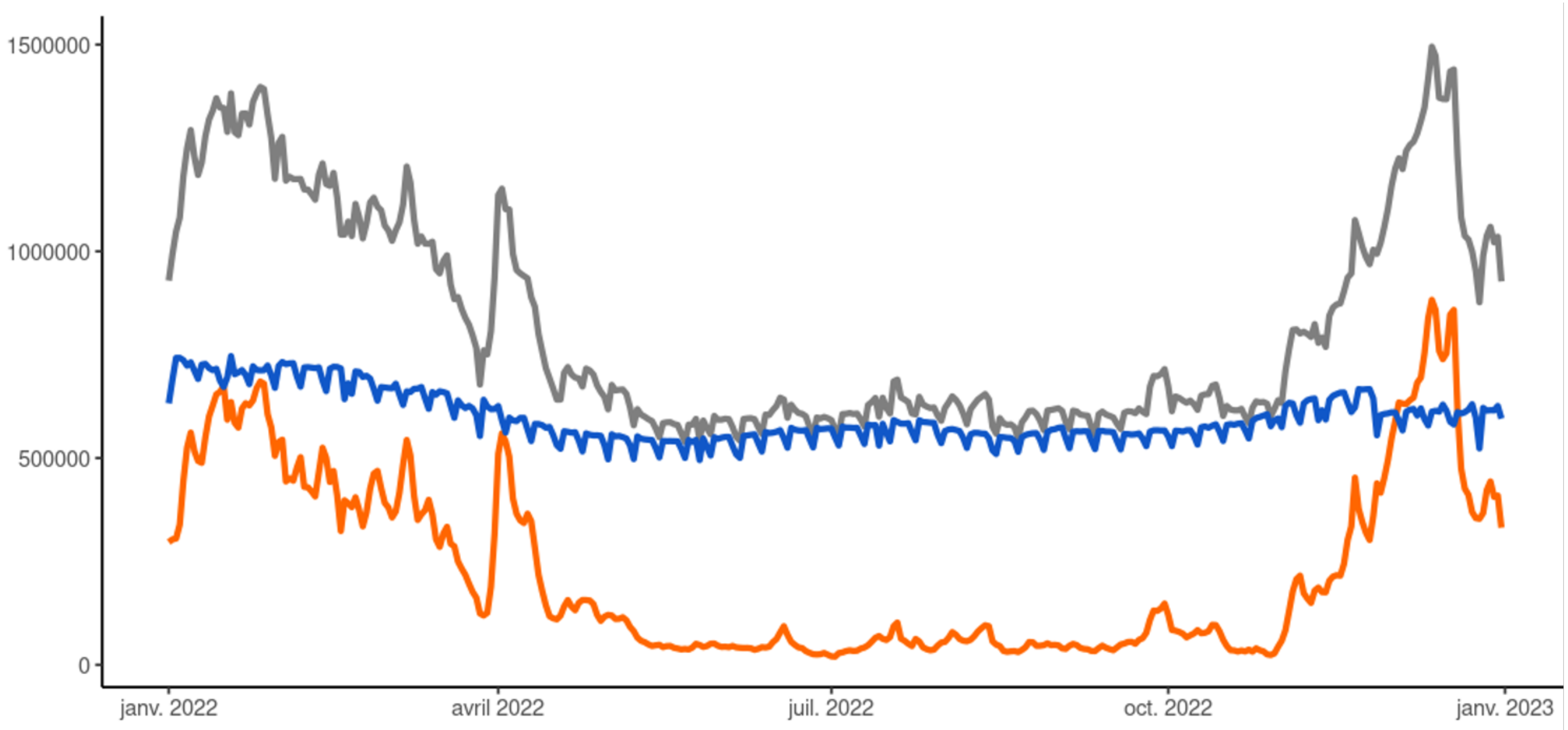
- Feature summary statistic or visualisation
- Model internals (learned weights)
- Data point (counterfactual explanations)
- Intrinsically interpretable model

A basic interpretable approach for load forecasting using generalized additive models

$$\text{Prediction} = \underbrace{\sum_{d=1}^{7} \alpha_d 1_{\text{Day}=d} + \sum_{h=1}^{48} \beta_h 1_{\text{Half-hour}=h}}_{\text{Non Climate Part}} + \underbrace{f(\text{Temperature})}_{\text{Climate Part}}$$

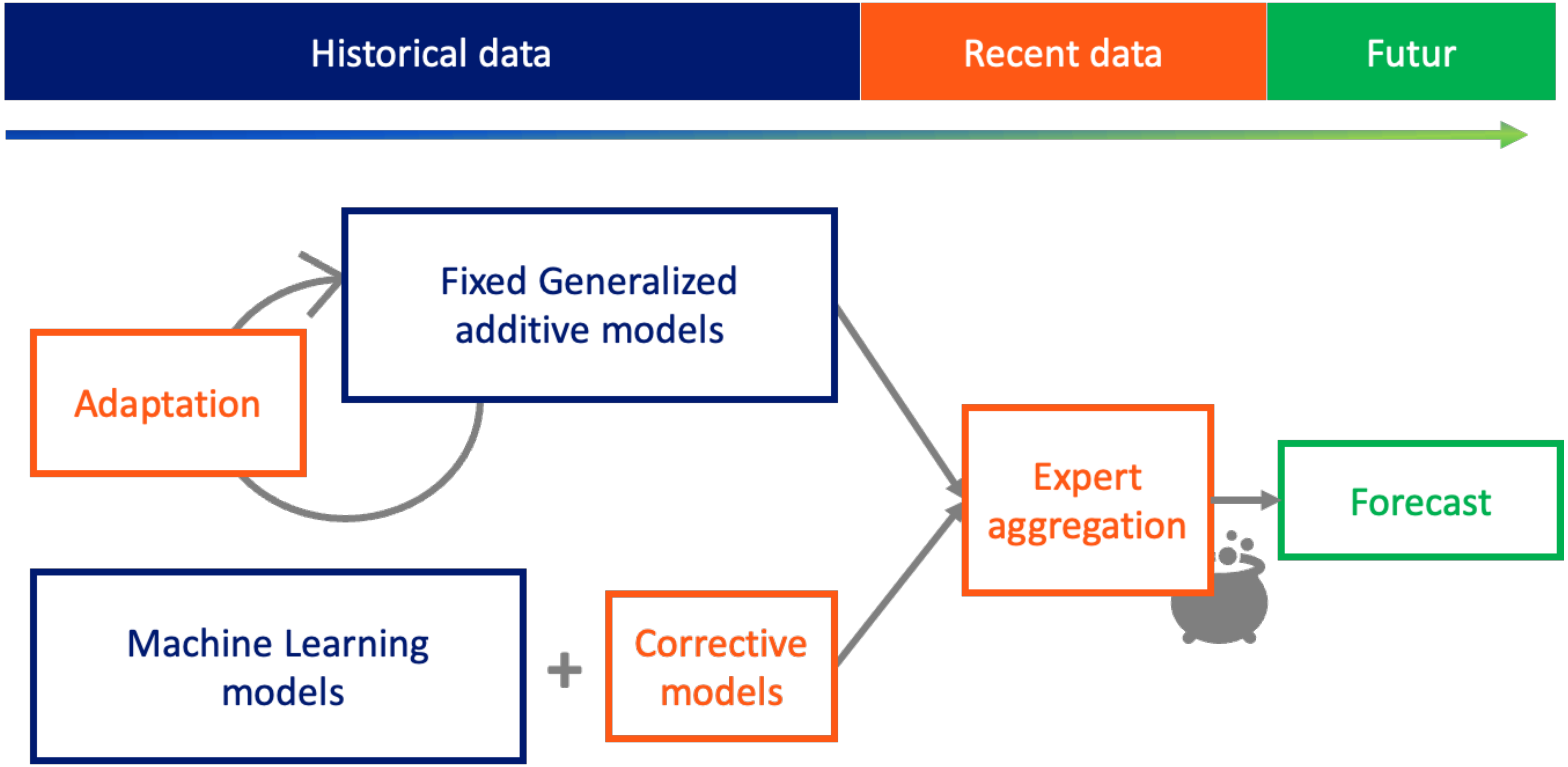


Non-independence of climatic and calendar variables / Cross-effects between climatic and calendar variables



→ Annual variation in the non-climatic component linked to thermal sensitivity in winter

How load forecasting really works



Antoniadis, A., Cugliari, J., Fasiolo, M., Goude, Y., & Poggi, J. M. (2024). *Statistical Learning Tools for Electricity Load Forecasting*. Springer International Publishing AG.

Forecasting demand at EDF relies on

- Static statistical / machine learning models (regression framework)
  - Generalized additive models
  - Deep neural networks, Random forest...
- And sequential learning updates (time series framework)
  - Corrective time series models
  - State-Space models (Kalman filter)
  - Online models aggregation

Explainability is needed to

- Understanding the model (robustness) and analysis of forecasting errors
- Interpretation of demand trends, quantification of eco-actions
- Right to explainability (customer perspective)

→Looking for a model-agnostic local method!

📘 Molnar, C. (2020). Interpretable machine learning. Lulu. com.

📘 Baur, L., Ditschuneit, K., Schambach, M., Kaymakci, C., Wollmann, T., & Sauer, A. (2024). Explainability and interpretability in electric load forecasting using machine learning techniques–a review. *Energy and AI*

Choice of the method to cut forecasts in positive contributions: **LIME or SHAP?**

Local Interpretable Model-agnostic Explanations (LIME)

- Perturb samples to generate a new dataset and get the predictions
- Weight generated data points according to their proximity to the point to explain
- Fit an interpretable model (linear regularized regression)

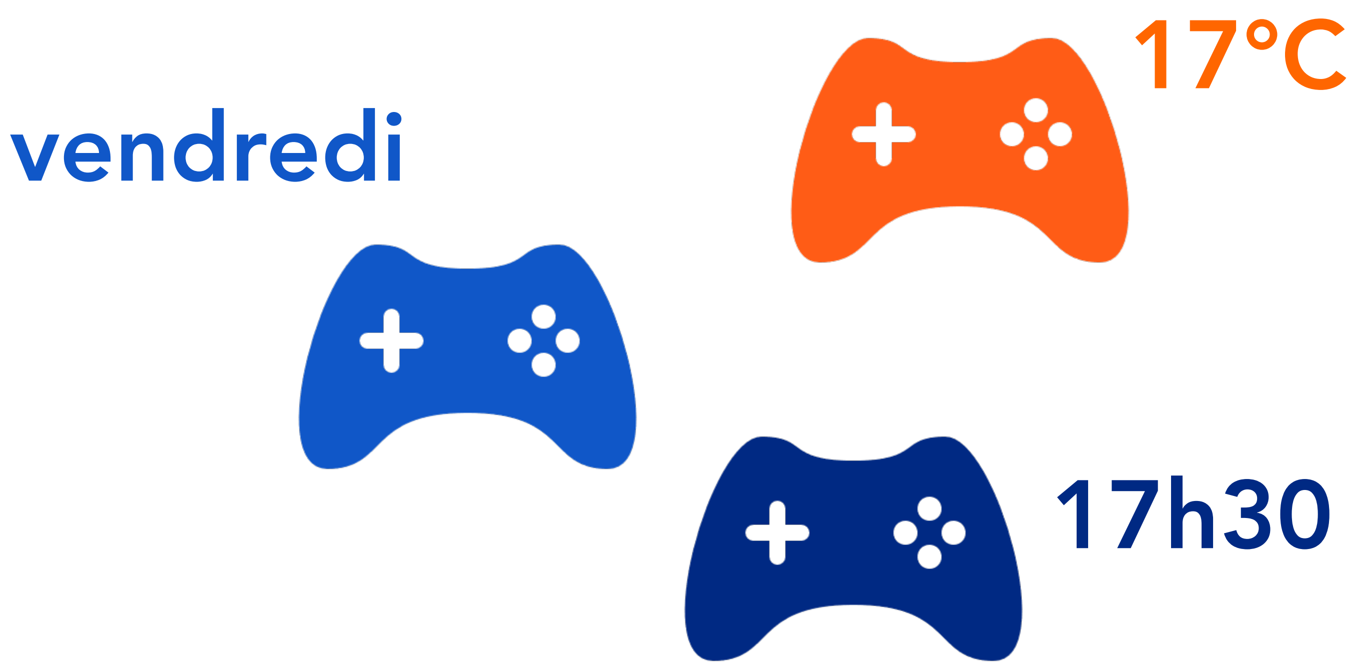
→ Implicitly assume the model to be locally linear

Features must be independent to obtain coherent explanations

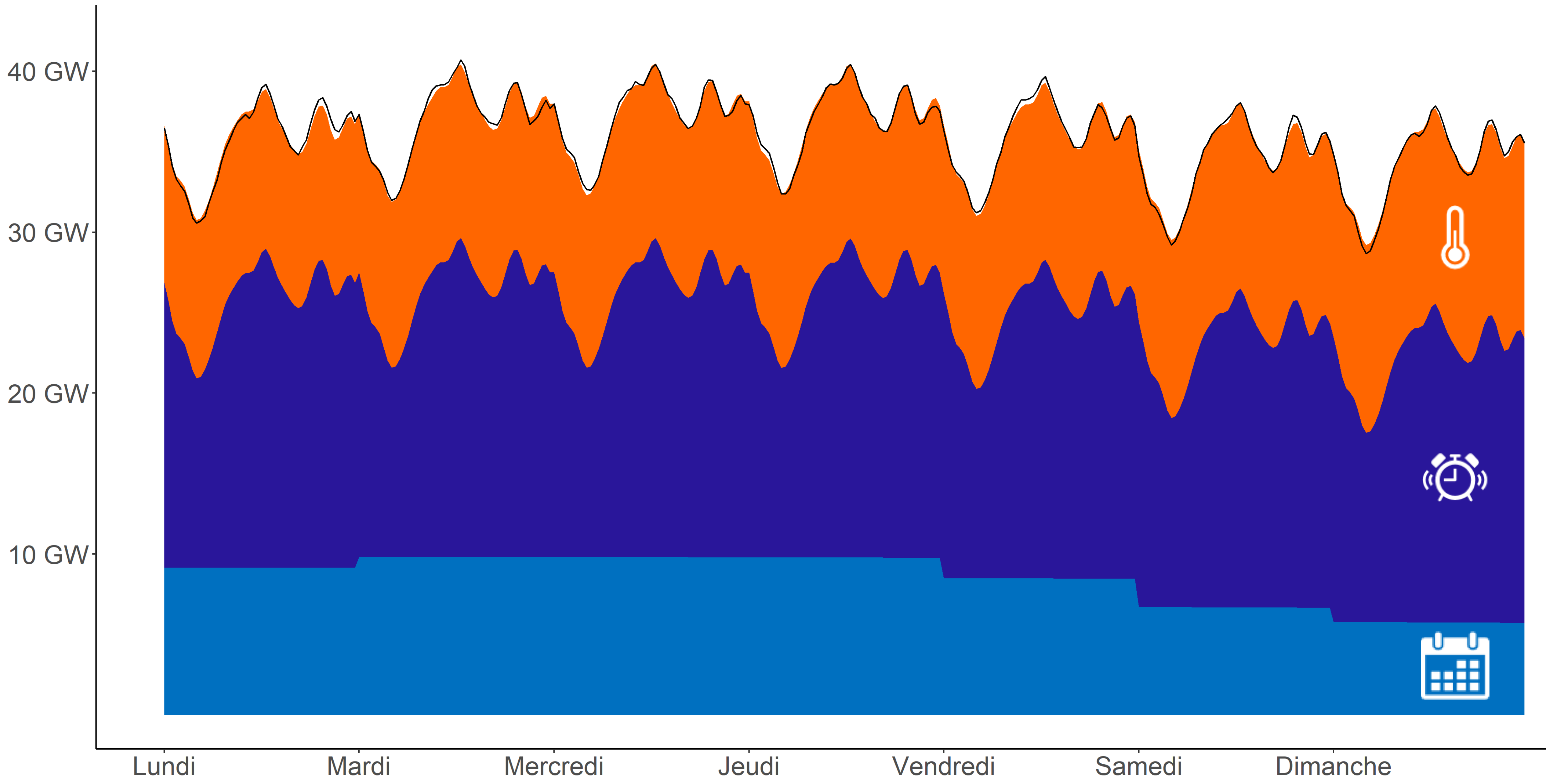
Shapley values for explainability

Shapley values (game theory - 1953)

Each explanatory variable value is a « player » in a collaborative game whose prediction is the payoff

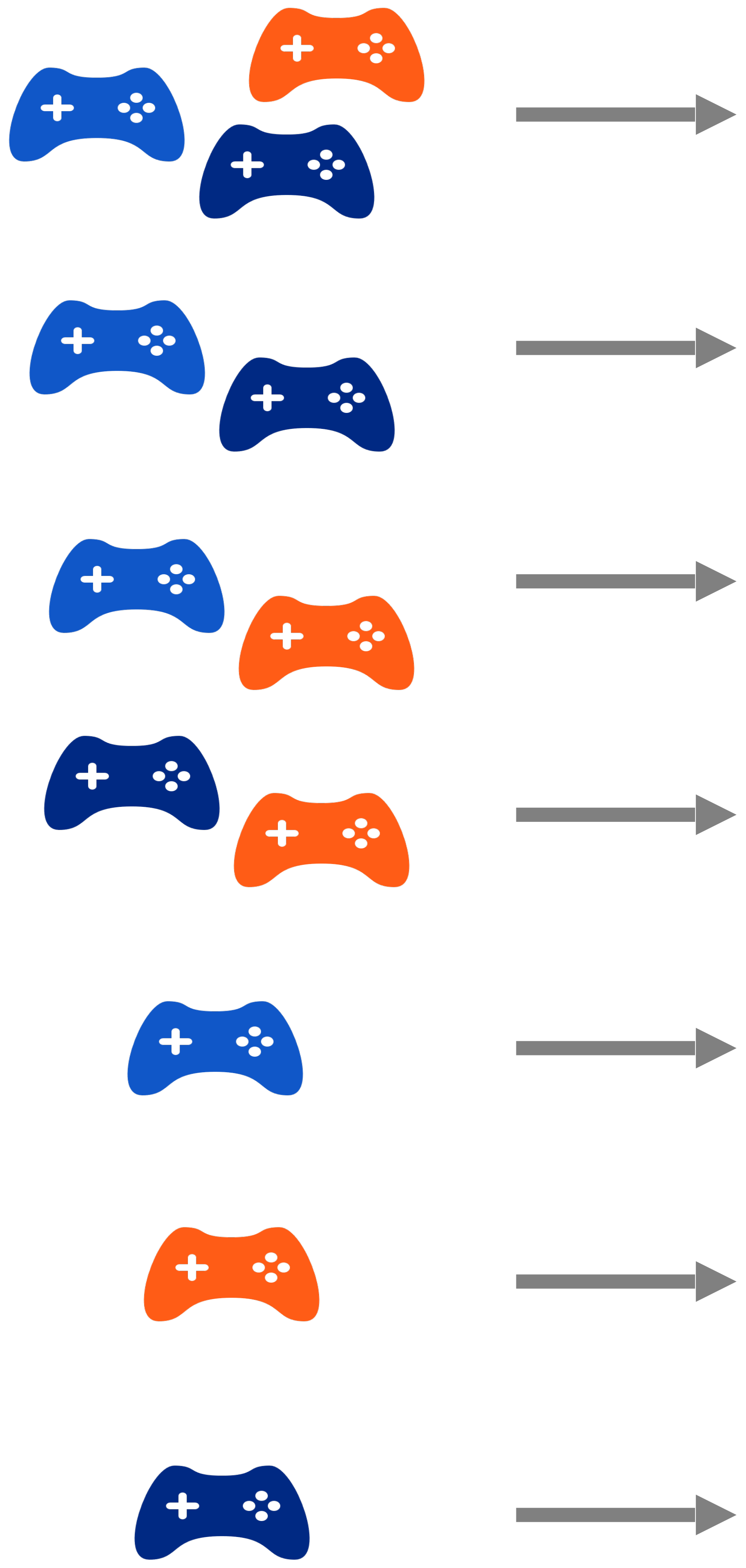


4 115 + 32 205 + 11 503 = 47 823 MW



Shapley values (game theory - 1953)

Let's assume that the game can be replayed as many times as you like for any coalition of the p players



$v(\text{vendredi}, \; 17°C, \; 17h30)$

$v(\text{vendredi}, \; 17h30)$

...

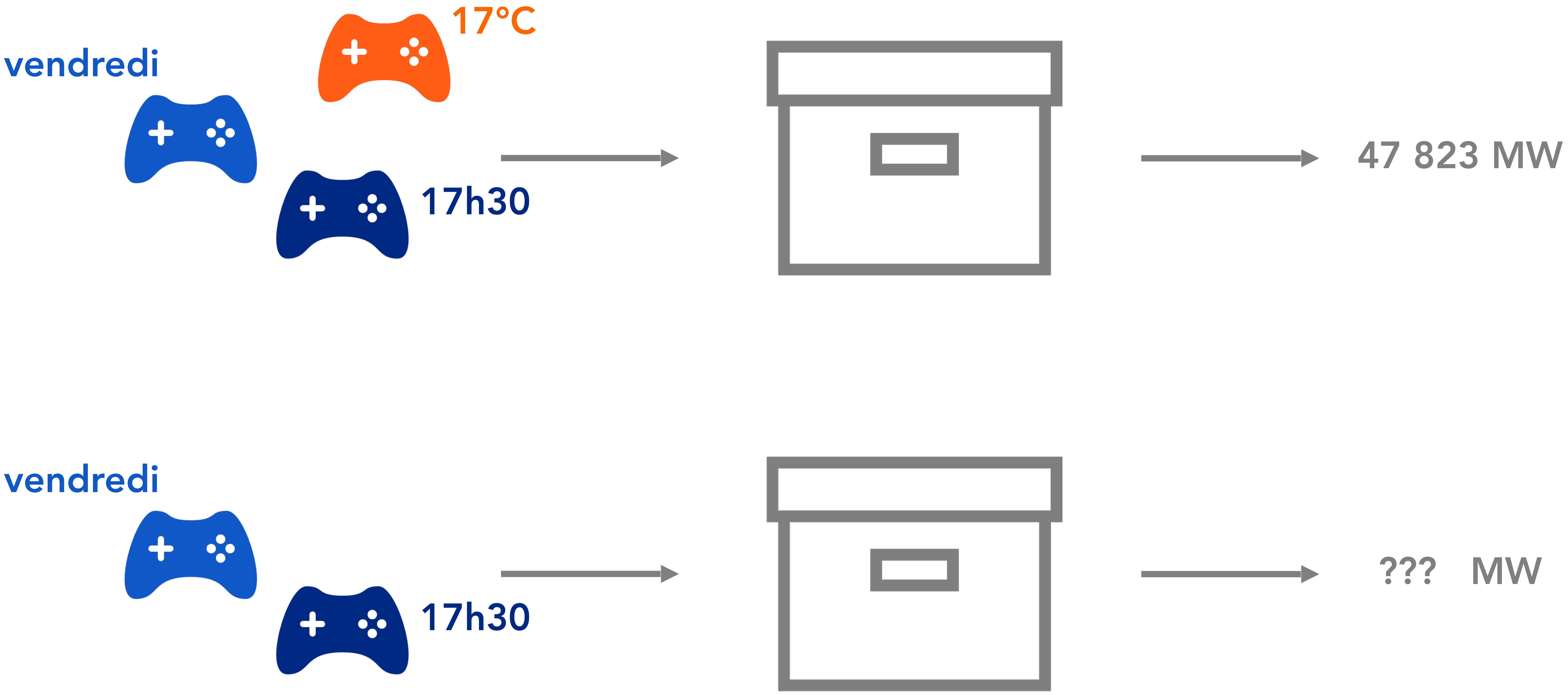
$$\phi_j = \frac{1}{\text{number of players}} \sum_{\text{Coalitions with } j} \text{Contribution of player } j$$

$$\phi_j = \frac{1}{p} \sum_{S \subset \{1,\ldots,p\} \backslash \{j\}} \frac{1}{\binom{p-1}{|S|}} \left(v\left(S \cup \{j\}\right) - v(S) \right)$$

The only method that guarantees a fair distribution (Shapley):

- Efficiency - $v\left(\{1,\ldots,p\}\right) = \sum_{j=1}^{p} \phi_j$
- Symmetry - $\forall \; S \backslash \{j, i\} \,, \;\; v(S \cup \{i\}) = v(S \cup \{j\}) \;\Rightarrow\; \phi_j = \phi_i$
- Dummy - $\forall \; S \backslash \{j\} \,, \;\; v(S \cup \{j\}) = v(S) \;\Rightarrow\; \phi_j = 0$
- Additivity - For two different games, the contributions add up.

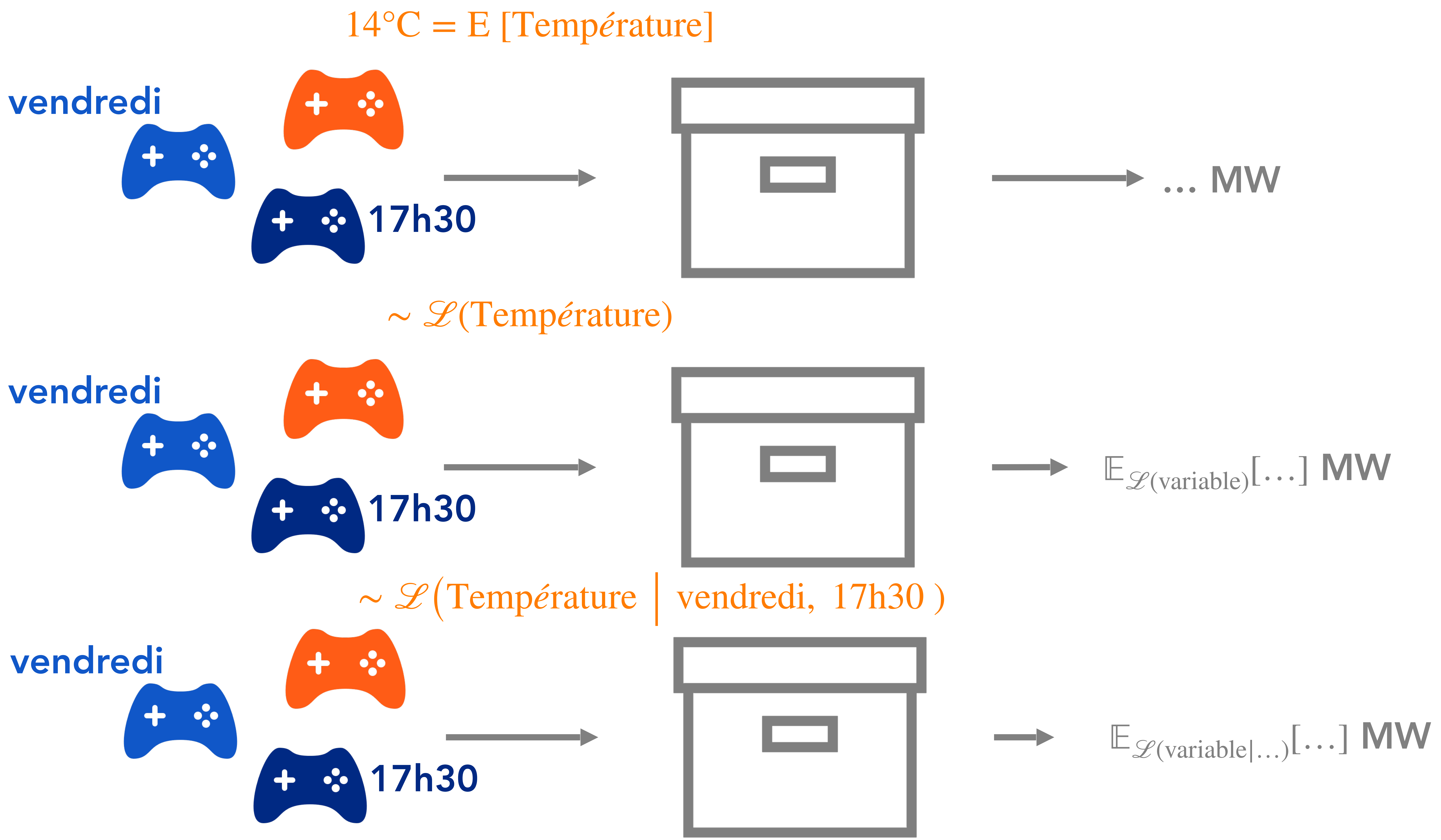
Application for ML explainability



eXplainability through Positive Contributions (XPC) package

How to replace the variable?

- Baseline
- Interventional
- Conditional



How to approximate probability laws?

- Interventional

$$v(x_S) = E\left[f\left(x_S, X_{\bar{S}}\right)\right] = \int f\left(x_S, z\right)\, P_{X_{\bar{S}}}(dz) \approx \frac{1}{n}\sum_{i=1}^{n} f(x_S, z_{i,\bar{S}})$$

- Conditional

$$v(x_S) = E\left[f\left(x_S, X_{\bar{S}}\right) \mid X_S = x_S\right] = \int f\left(x_S, z\right)\, P_{X_{\bar{S}}\mid X_S = x_S}(dz) \approx \frac{1}{|V_x|}\sum_{z \in V_x} f(x_S, z_{i,\bar{S}})$$

→ Computation cost: $p \times 2 \times 2^{p-1} \times n$ model calls $f(\ )$ to compute p Shapley values

MonteCarlo approach

Two sampling:

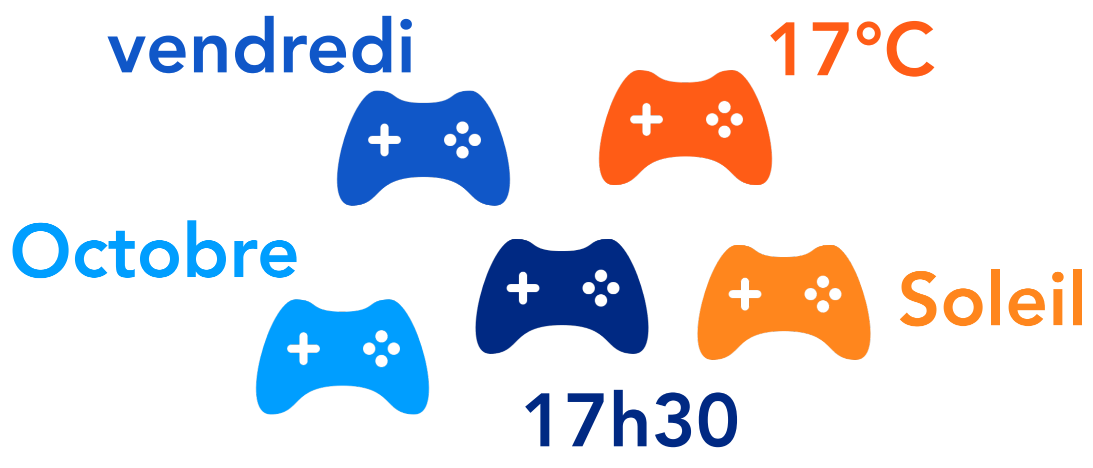
- On the coalitions (groups of players)

- On the data points

$$\phi_j = \frac{1}{\mathrm{p}} \sum_{S \subset \{1,\ldots,p\} \setminus \{j\}} \frac{1}{\binom{p-1}{|S|}} \left(v\left(S \cup \{j\}\right) - v(S) \right) \approx \frac{1}{\mathrm{p}} \frac{1}{n_1} \sum_{k=1}^{n_1} \frac{1}{n_2} \sum_{i=1}^{n_2} \left(f\left(x_{S_k}, z_{i,j}, z_{i,\bar{S}_k}\right) - f\left(x_{S_k}, x_{i,j}, z_{i,\bar{S}_k}\right)\right)$$

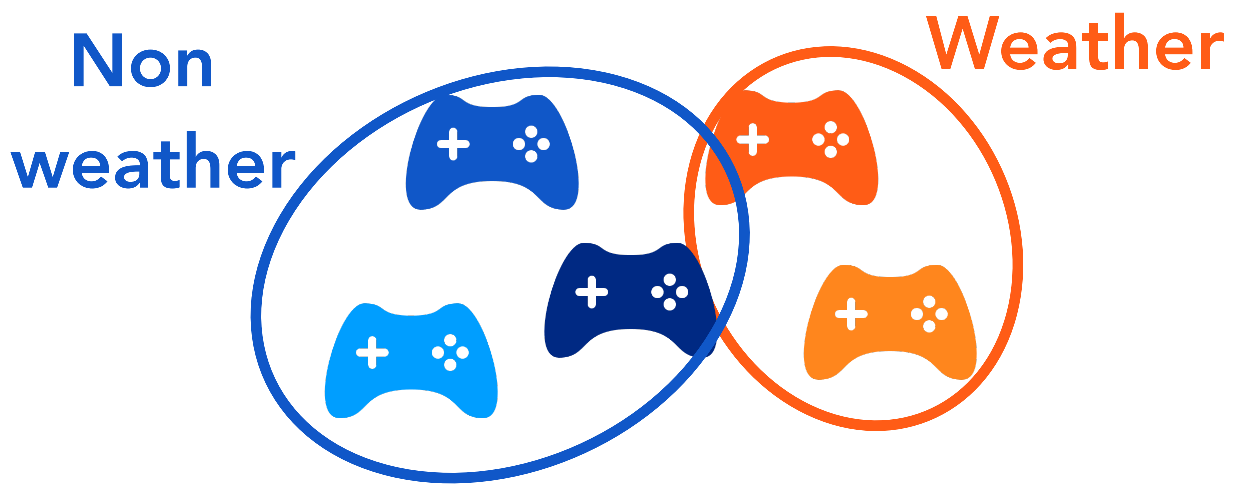
→ Computation cost: $p \times 2 \times n_1 \times n_2$ model calls $f(\,)$ to compute p Shapley values

How to aggregate the Shapley values to get a Weather and a Non weather components?



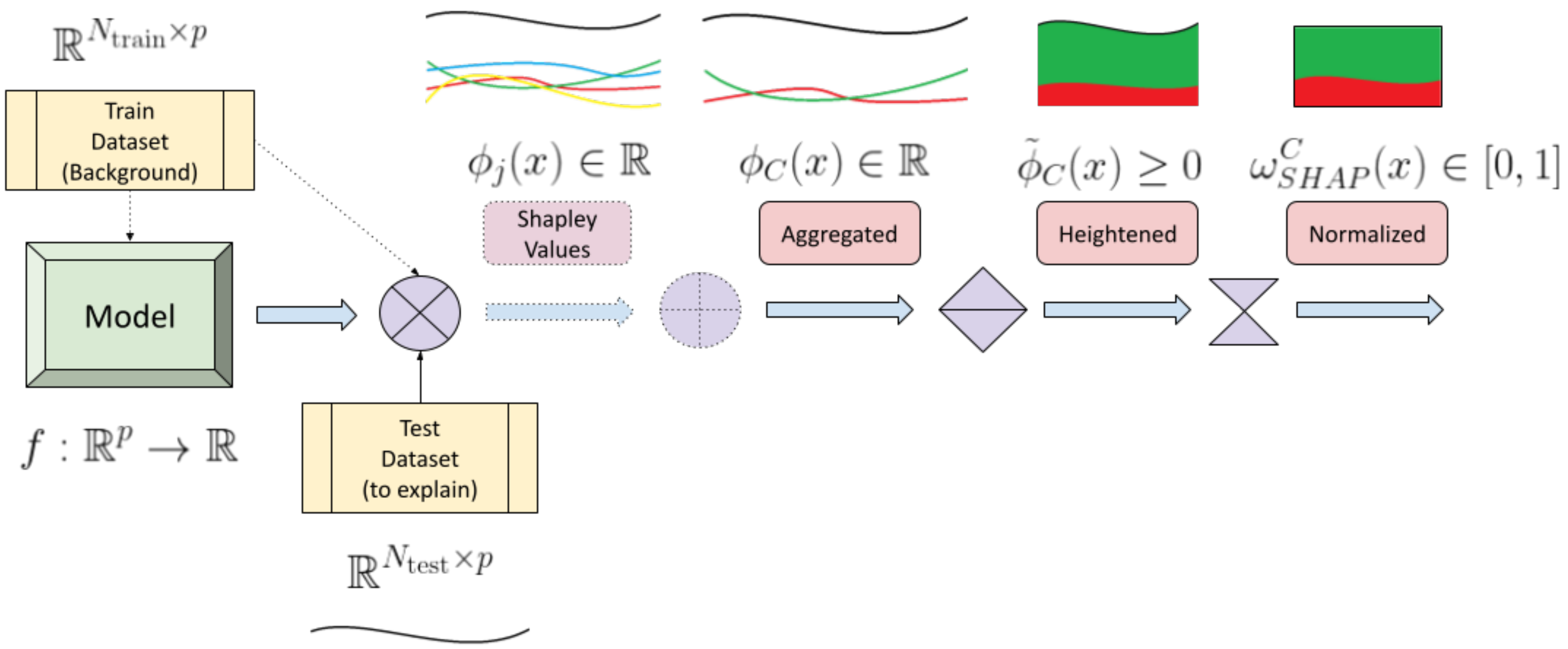
- Sum of « weather » Shapley values

- Shapley values for a game with 1 + number of non-weather variables players



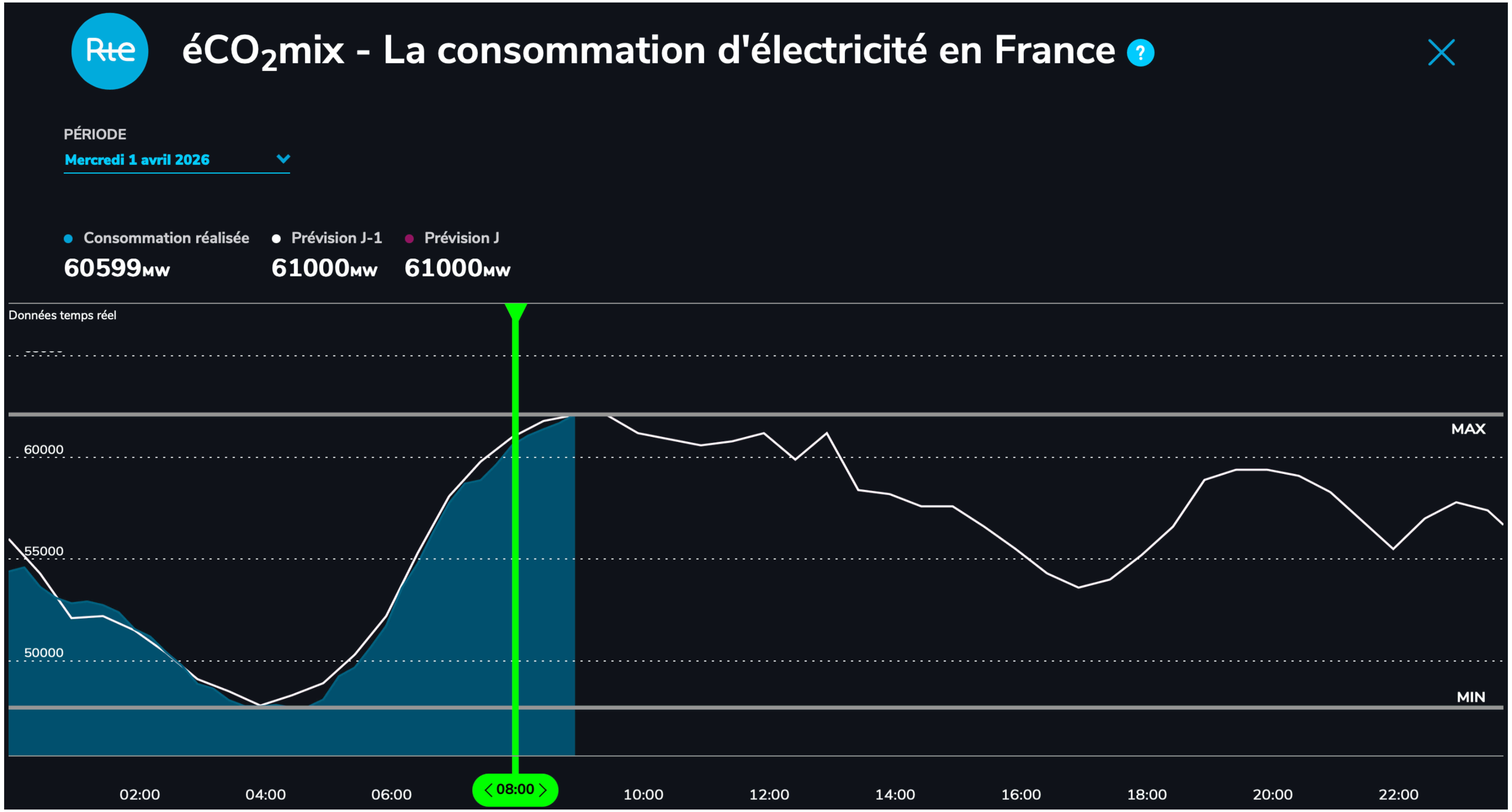
- Shapley values for a gam with 2 players

eXplainability through Positive Contributions (XPC) python package



Experiments

Application to electrical demand forecasting



Proof of concept

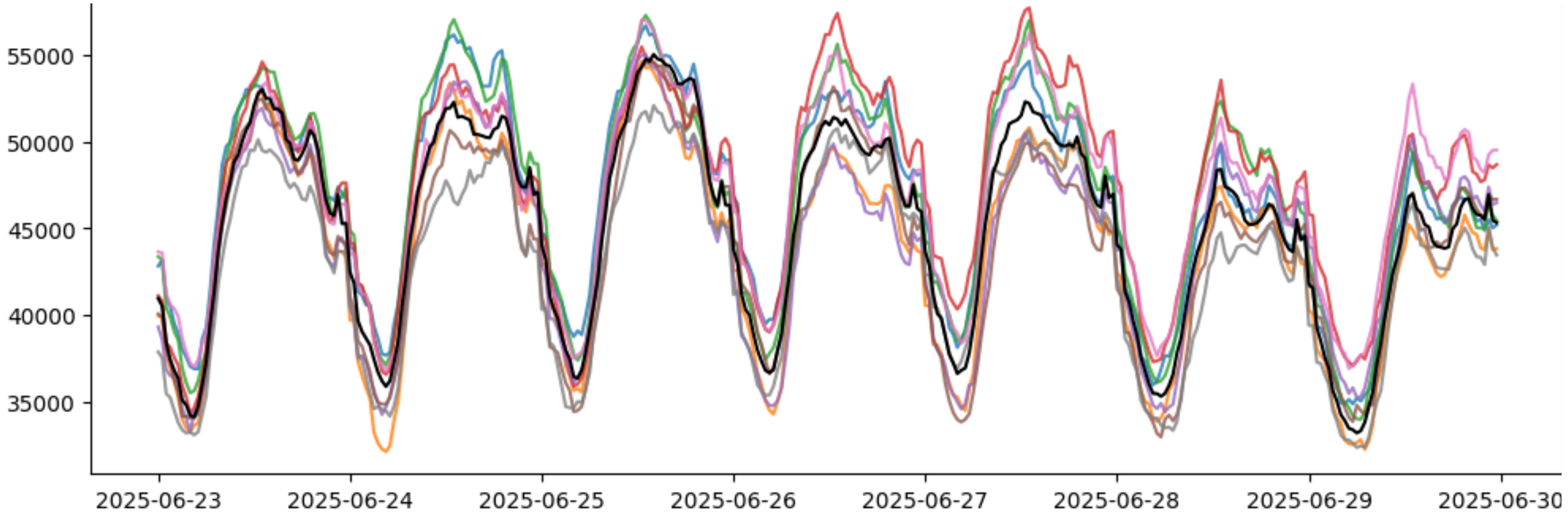
Synthetic data generation with a very parametric modeling (fitted on 2021-2025 electrical demand data) → a (rare) case of supervised learning!

- Simple additive model (MAPE 4.1%) + auto-regressive model: $Y_t = s(T_t) + s(\bar{T}_t) + s(\pi_t) + \alpha_{h_t} + \beta_{d_t} + \delta_{v_t} + \ldots + \sigma_{h_t} Z_t$ with $Z_t = \sum_{i=1}^{96} \phi_i Z_{t-i} + \varepsilon_t$
- Additive model per half-hour (MAPE 2.8%) + auto-regressive model: $Y_t = s^{h_t}(T_t) + s^{h_t}(\bar{T}_t) + s^{h_t}(\pi_t) + \beta_{d_t}^{h_t} + \delta_{v_t}^{h_t} + \ldots + \sigma_{h_t} Z_t$ $\qquad \varepsilon_t \sim \mathcal{N}(0,1)$

True decomposition:

$$\phi_t(\text{weather}) \propto \text{weather terms}$$

$$\phi_t(\text{nonweather}) \propto \text{nonweather terms}$$



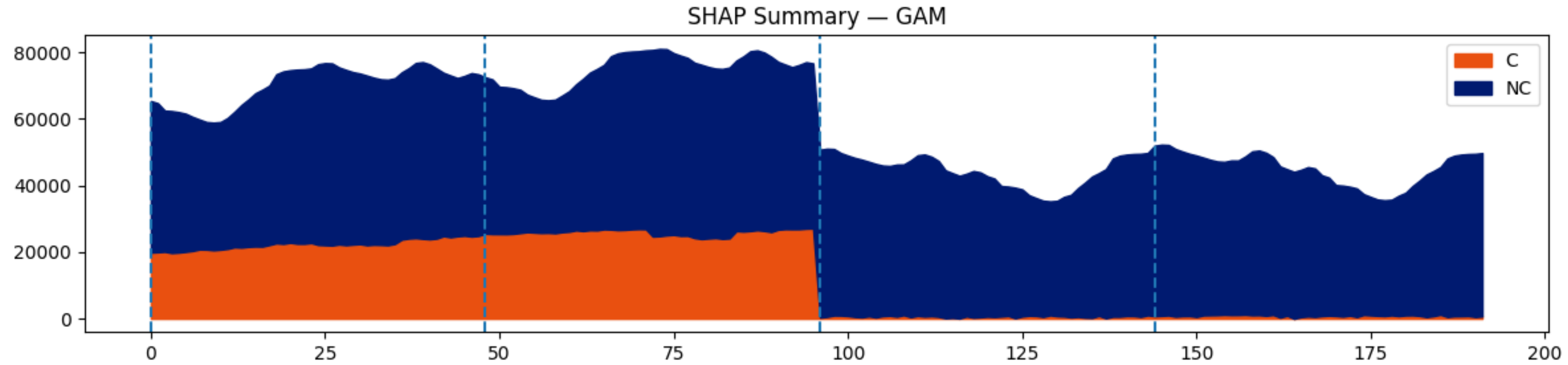
Proof of concept (auto-regressive noise induces errors)

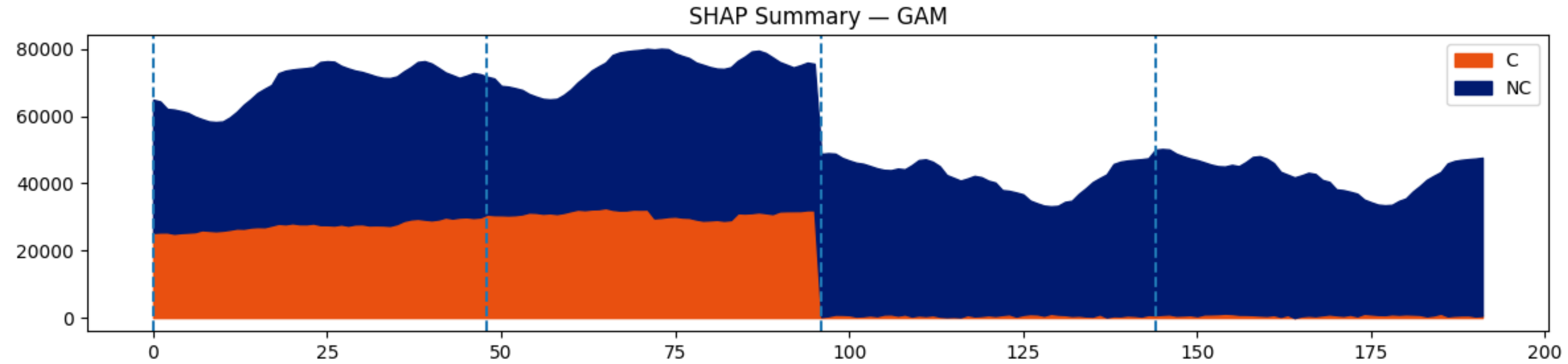
Simple additive model + AR

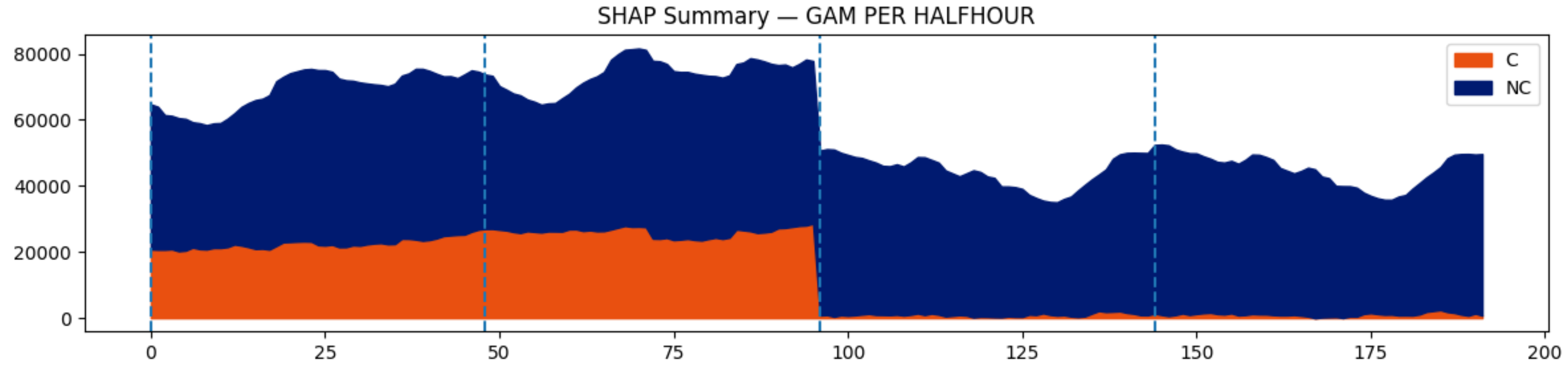
Error (%) Model	Train (2021–23)	Test (2024–25)	Efficiency (4 days)	Shapley (winter days)
GAM	3.93	4.73	0.44	2.72
GAM with cross-effect	3.92	4.74	0.45	2.37
GAM per half-hour	**3.87**	**4.67**	**0.50**	**2.16**
Random Forest	0.59	5.56	0.77	4.32
Gradient Boosting	4.13	4.83	0.42	5.39

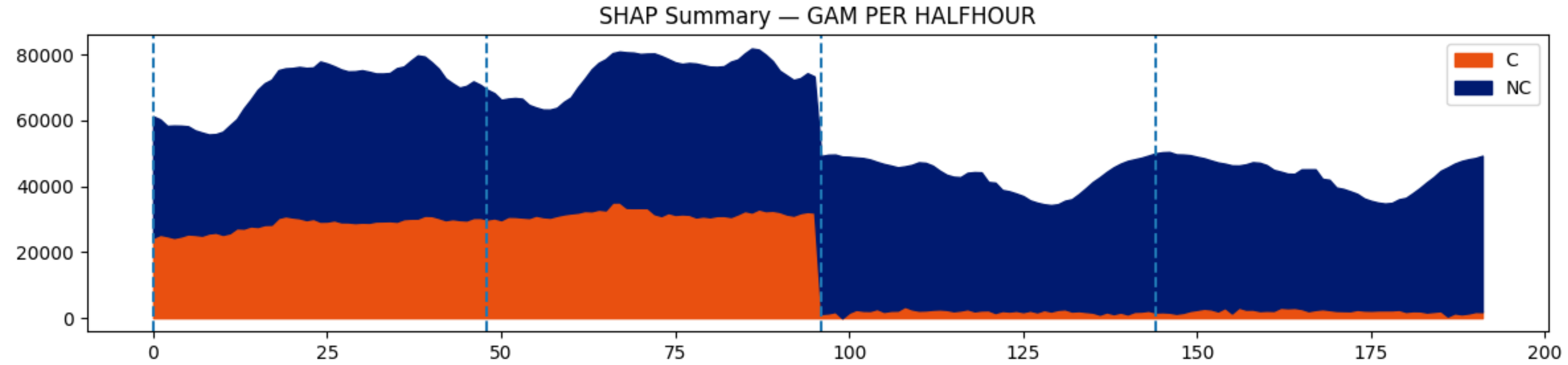
Additive model per half-hour+ AR

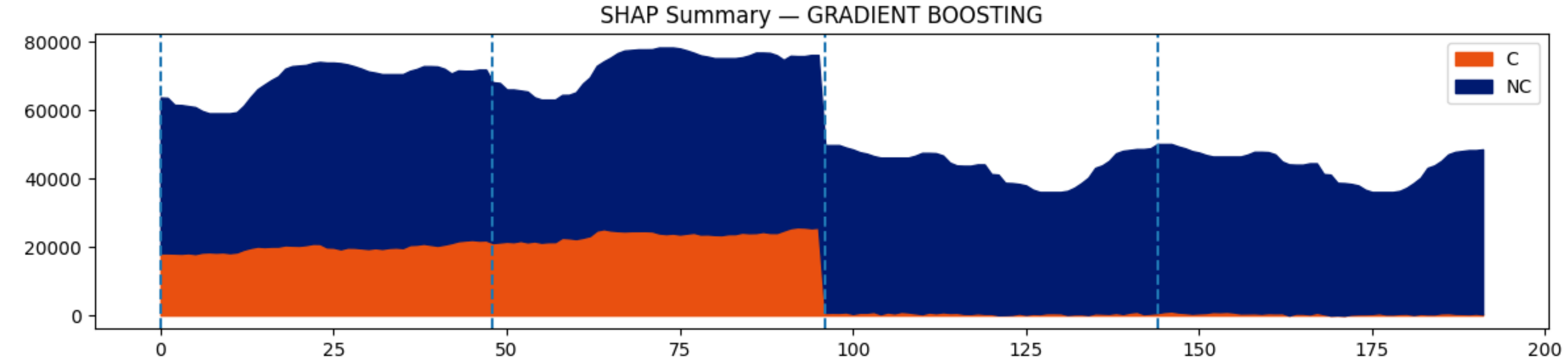
Error (%) Model	Train (2021–23)	Test (2024–25)	Efficiency (4 days)	Shapley (winter days)
GAM	4.01	4.45	0.71	3.02
GAM with cross-effect	3.92	4.41	0.85	3.16
GAM per half-hour	**2.53**	**3.23**	**0.85**	**2.53**
Random Forest	0.42	3.78	0.90	2.99
Gradient Boosting	3.38	3.84	0.67	4.17

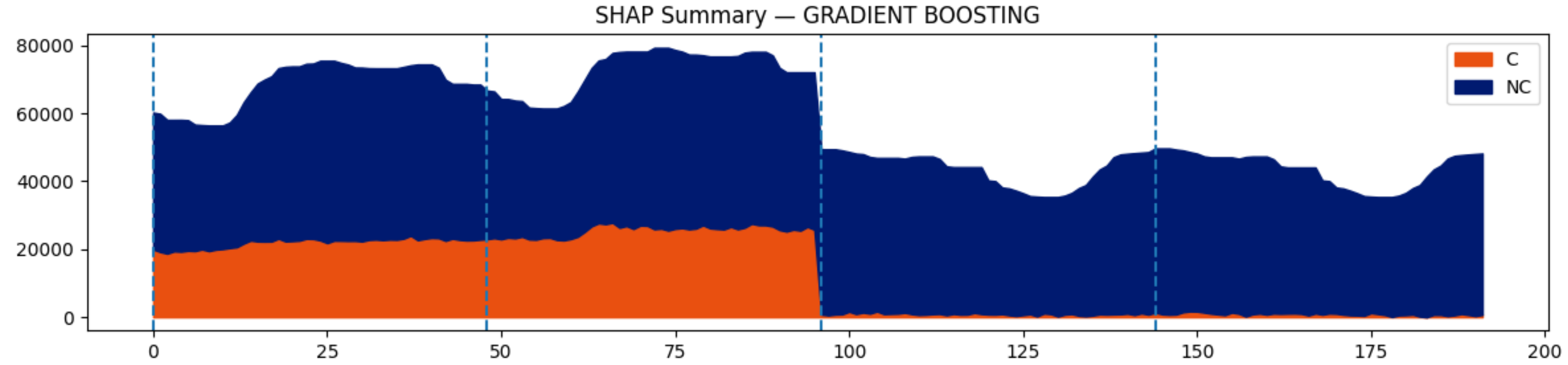












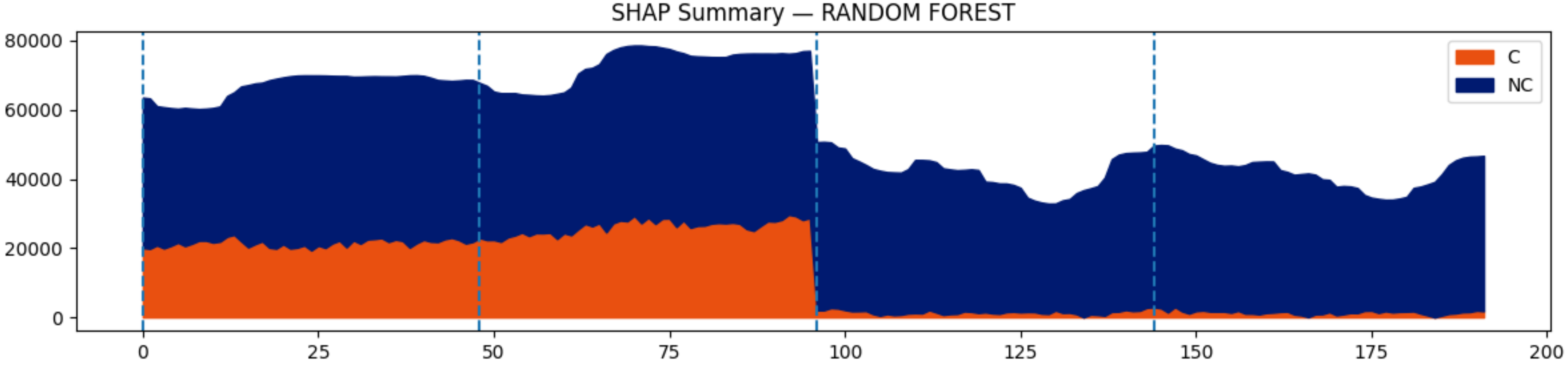
$n_1(\text{coalition}) \times n_2(\text{sample}) = 10\,000$

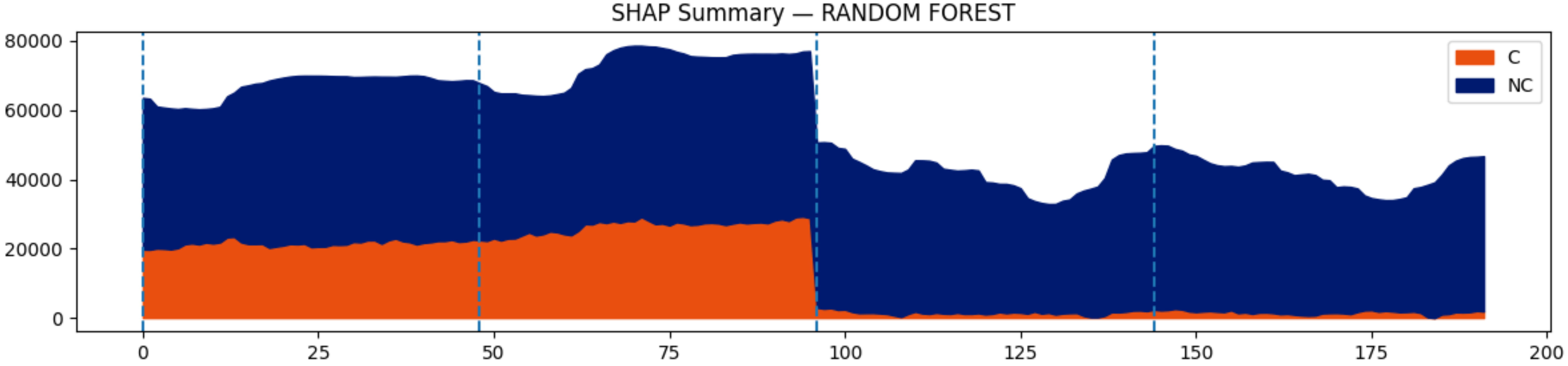
$n_1 = 10 \quad n_2 = 100$

Error (%) Model	Efficiency (4 days)	Shapley (winter days)
GAM	0.75	**3.00**
GAM with cross-effect	1.18	**3.13**
GAM per half-hour	1.08	**2.50**
Random Forest	1.59	3.19
Gradient Boosting	0.94	4.20

$n_1 = 100 \quad n_2 = 10$

Error (%) Model	Efficiency (4 days)	Shapley (winter days)
GAM	**0.71**	3.02
GAM with cross-effect	**0.85**	3.16
GAM per half-hour	**0.85**	2.53
Random Forest	**0.90**	**2.99**
Gradient Boosting	**0.67**	**4.17**

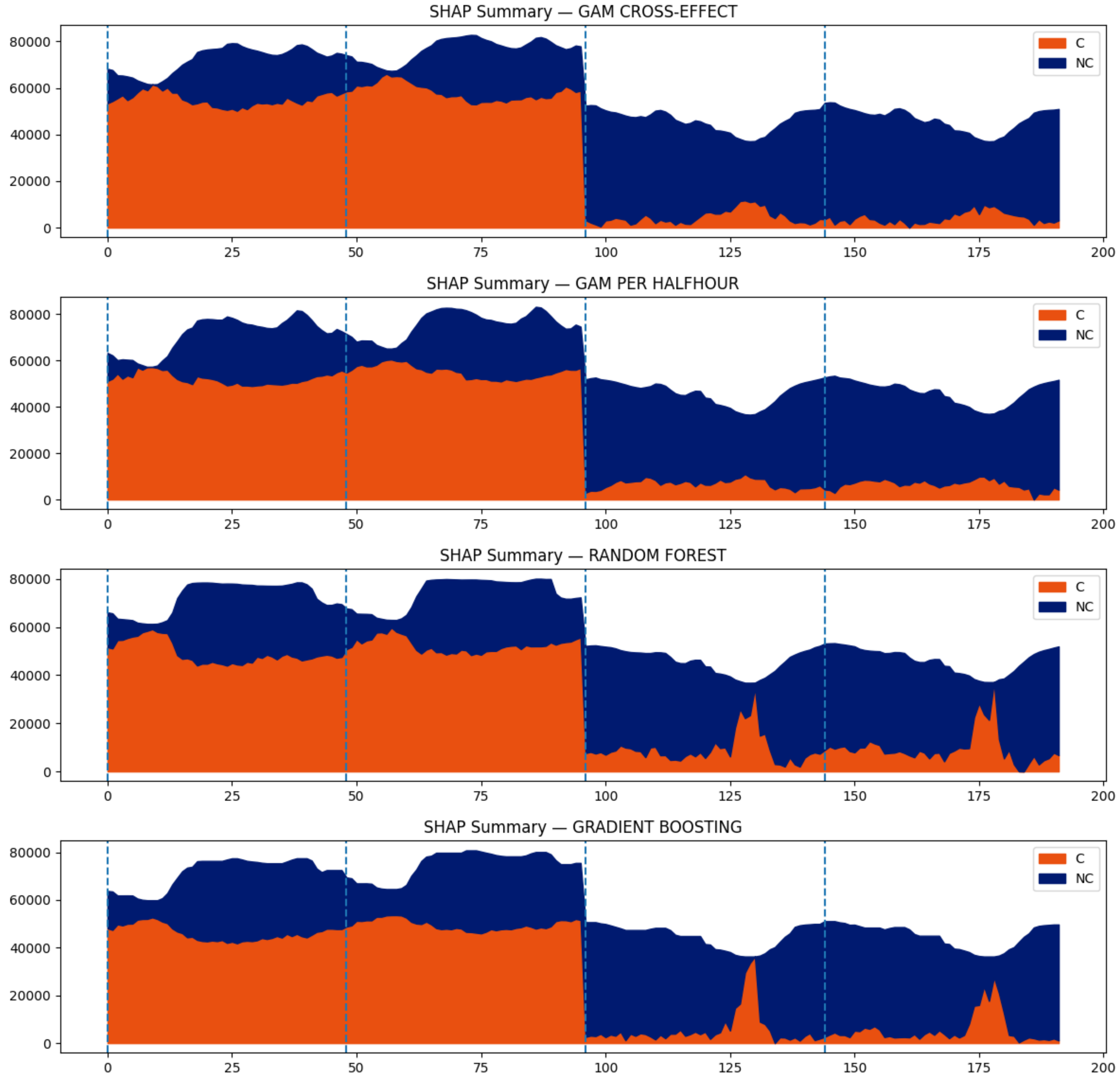




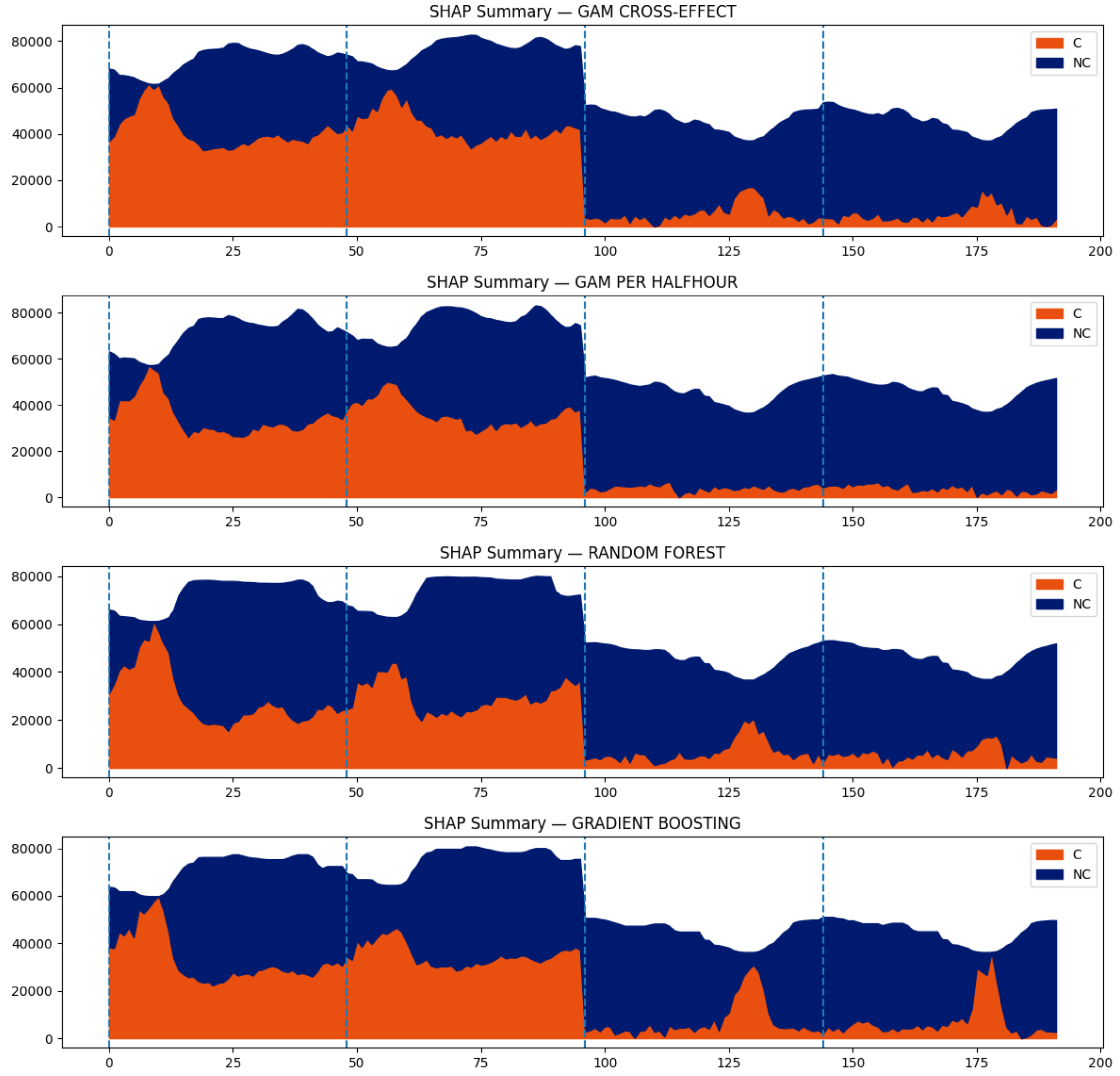
On real data

Error (%) Model	Train (2021–23)	Test (2024–25)	Efficiency Int.	Efficiency Cond.
GAM	4.11	4.62	0.46	13.6
GAM with cross-effect	4.00	4.63	0.81	14.2
GAM per half-hour	**2.73**	**3.36**	**0.59**	**14.8**
Random Forest	0.36	4.17	0.79	17.1
Gradient Boosting	3.25	4.21	0.47	14.4

Interventional



Conditional



Prospects

- Improving neighborhood definition to stabilize and accelerate conditional setting
- Applications to models in online settings: with $\left((Y_s, X_s)_{s \leq t-D}, X_t\right)$ in input
- Paper to come…

Thank you for your attention!

Question?