Reconstruction of Local Energy Consumption

Dati, modelli ed indicatori per la valutazione dell'impatto dei cambiamenti climatici su cittadini e sulle imprese

Workshop-Università degli Studi della Campania

Pietro Colombo

University of Glasgow

02–04–2025

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
GRINS FOUNDATION

Context.

- Socio-economic and environmental phenomena evolve in time and space.
- Variations relate to industrial policies, sectoral investments, and physical risk can be related to local variables.

Goal. Build models that:

- Capture changes across time and geography,
- Relate observed variation to structural and financial drivers.

Why We Need Local Models

Challenge: Global models hide regional heterogeneity.

Need:

- Local models reveal provincial differences and sensitivities,
- Aggregation reconciles local forecasts to a coherent global view.

Key Idea: *Model locally, reconcile globally.*

From Complex Data to Aggregated Indicators

Approach.

- Aggregate/reduce complex data by area/period,
- Train a dynamic ML impact indicator updated with new data,
- Provide interpretation aids for adaptation measures and key drivers.

Outcome. Dynamic monitoring tool by area highlighting high-impact factors.

Electricity Demand Forecasting & Local Impact

Goal

Estimate electricity demand and its environmental sensitivity.

Data Limitation

Local load is not directly observed; only regional/global load is available.

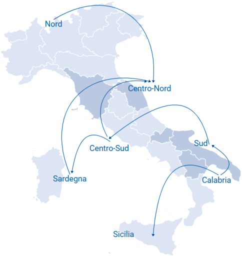

Figure 1: Bidding Zones of Terna

Reconstructing Local Load Without Local Measurements

Data

What do we have?

1. Local predictors (ERA5 weather, calendar covariates),
2. Regional/Global Load.

Problem

We miss a local response

Solution

1. We developed a strategy.
2. A testing scheme.
3. An application.

The Model: Indices, Series, and Splits

Indices. Time $t \in \mathcal{T}$ (hourly); province $i \in \mathcal{I}$; drivers $d \in \mathcal{D}$.

Observed. Global load y_t; province drivers $x_{i,d,t}$.

Structural shares. $w_i \geq 0$, $\sum_i w_i = 1$.

Splits. $\mathcal{T}_{\text{train}} = \{t : t \leq \texttt{train\_end}\}$, $\mathcal{T}_{\text{val}} = \{t : \texttt{val\_start} \leq t \leq \texttt{val\_end}\}$, $\mathcal{T}_{\text{test}} = \{t : t \geq \texttt{test\_start}\}$.

1. Forecasting Framework

- Objective: predict hourly electricity load y_t.
- Two layers:
  1. **Calendar model** $f_{\text{cal}}(z_t)$ (deterministic patterns),
  2. **Environmental model** $g_{\text{env}}(X_t)$ (meteorology on residuals).

$$\hat{y}_t^{\text{glob}} = \hat{y}_t^{\text{cal}} + \hat{r}_t$$

2. Calendar Model (Global)

Calendar covariates:

$$z_t = \text{features from } \texttt{X\_hourly} \text{ (hour, weekday, holiday, season, ...)}$$

Model:

$$y_t = f_{\text{cal}}(z_t) + \varepsilon_t^{\text{cal}}, \qquad \hat{y}_t^{\text{cal}} = f_{\text{cal}}(z_t)$$

Choices:

$$f_{\text{cal}} \in \begin{cases} \text{LM: } \beta_0 + \beta^\top z_t, \\ \text{GAM: } s(\text{hour}) + \beta^\top z_{\setminus \text{hour}}, \\ \text{QBRT: quantile GBM at } \tau = 0.5. \end{cases}$$

3. Environmental Model on Residuals (Global)

Residuals: $r_t = y_t - \hat{y}_t^{\mathrm{cal}}$

Aggregate drivers:

$$X_{d,t} = \sum_i x_{i,d,t}, \qquad X_t = (X_{d,t})_{d \in \mathcal{D}}$$

Model:

$$r_t = g_{\mathrm{env}}(X_t) + \varepsilon_t^{\mathrm{env}}, \qquad \hat{r}_t = g_{\mathrm{env}}(X_t)$$

with $g_{\mathrm{env}} \in \{\text{LM, GAM, QBRT}\}$.

4. Global Forecast Combination

$$\hat{y}_t^{\text{glob}} = \hat{y}_t^{\text{cal}} + \hat{r}_t = f_{\text{cal}}(z_t) + g_{\text{env}}(X_t)$$

- f_{cal} captures recurrent time structure,
- g_{env} adjusts for short-term meteorological effects.

5. Province Reconciliation

Calendar allocation (fixed):

$$\hat{y}^{\text{cal}}_{i,t} = w_i \, \hat{y}^{\text{cal}}_t, \qquad \sum_i w_i = 1$$

Residual allocation (two options):

(A) Shares: $\hat{y}^{\text{res}}_{i,t} = \alpha_i \, \hat{r}_t$ or (B) Derivative: $\hat{y}^{\text{res}}_{i,t} = \sum_d m_{d,t} x_{i,d,t} + w_i R_t$

Reconciliation:

$$\sum_i \hat{y}_{i,t} = \sum_i (\hat{y}^{\text{cal}}_{i,t} + \hat{y}^{\text{res}}_{i,t}) = \hat{y}^{\text{cal}}_t + \hat{r}_t = \hat{y}^{\text{glob}}_t$$

Interpretation of R_t and the Calendar–Environmental Split

Context. The global residual model:

$$r_t = g_{\text{env}}(X_t)$$

is used to distribute the total residual $\hat{r}_t$ across provinces.

Derivative-based allocation:

$$\hat{y}^{\text{res}}_{i,t} = \sum_d m_{d,t}\, x_{i,d,t} + w_i R_t, \qquad R_t = \hat{r}_t - \sum_d m_{d,t} X_{d,t}$$

- $m_{d,t}$ — sensitivity (partial derivative) of the environmental model to driver d;
- $x_{i,d,t}$ — local value of driver d for province i;
- R_t — **unexplained residual share**, distributed according to structural weights w_i;
- w_i — structural weight of province i ($\sum_i w_i = 1$).

Interpretation:

- The explained part $\sum_d m_{d,t} x_{i,d,t}$ corresponds to the environmental effect;
- The term $w_i R_t$ represents the remaining structural (calendar-like) component not captured by the drivers.

6. Derivative-based Allocation (Local Gradient)

From $r_t = g_{\text{env}}(X_t)$, approximate partial derivatives:

$$m_{d,t} \approx \frac{g_{\text{env}}(X_t + \varepsilon e_d) - g_{\text{env}}(X_t - \varepsilon e_d)}{2\varepsilon}$$

Allocate by exposure:

$$y_{i,t}^{\text{res}} = \sum_d m_{d,t}\, x_{i,d,t} + w_i \left(\hat{r}_t - \sum_d m_{d,t} X_{d,t} \right) \quad \Rightarrow \quad \sum_i y_{i,t}^{\text{res}} = \hat{r}_t$$

7. Final Province Forecasts

$$\hat{y}_{i,t} = w_i \, \hat{y}_t^{\text{cal}} + y_{i,t}^{\text{res}}, \qquad \sum_i \hat{y}_{i,t} = \hat{y}_t^{\text{glob}}$$

- Interpretable decomposition (calendar vs environment),
- Sensitivity-aware local allocation,
- Exact coherence across aggregation levels.

Results: GAM–GAM Configuration

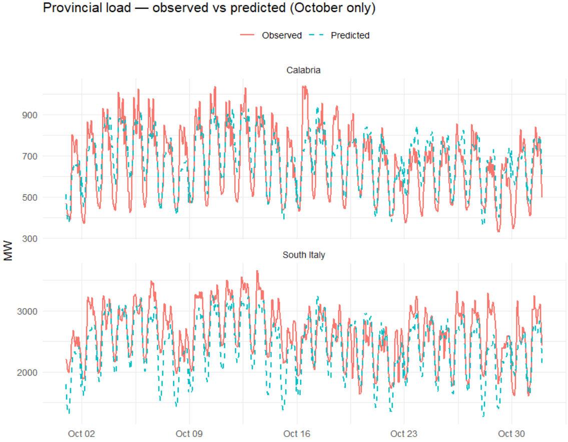

Figure 2: Both environmental and calendar models use **GAM**.

Results: LM–GAM Configuration

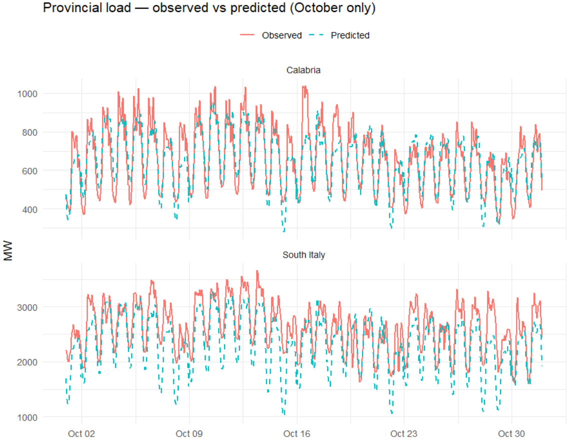

Figure 3: **GAM** for environment, **LM** for calendar.

Results: QBRT–GAM Configuration

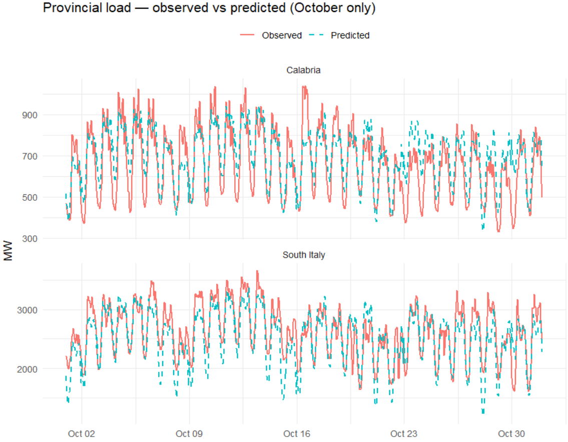

Figure 4: **GAM** for environment, **QBRT** for calendar.

Models: Comparison

Table 1: Performance metrics by province for different configurations.

Model	Province	RMSE	MAE	MAPE (%)	R^2
GAM–GAM	Calabria	96.6	76.8	12.4	0.673
	South Italy	334.0	274.0	10.7	0.701
QBRT–GAM	Calabria	103.0	81.4	13.5	0.659
	South Italy	285.0	229.0	9.07	0.722
LM–QBRT	Calabria	129.0	105.0	16.8	0.412
	South Italy	506.0	423.0	16.0	0.370

Application — Province South Italy

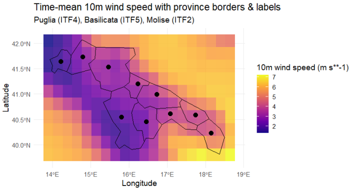

Figure 5: Provincial map: Puglia, Basilicata, Molise.

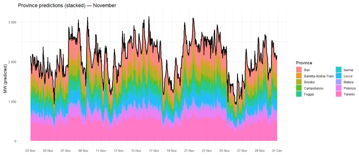

Figure 6: Predicted November load by province.

Environmental Drivers — Impact Interpretation

Table 2: Mean absolute and relative impact of environmental variables (test period).

Driver	Mean Abs. Impact [MW]	Mean Share
t2m_sd	1329	0.227
w100_sd	715	0.149
w100_mean	685	0.116
t2m_mean	532	0.050
w10_sd	505	0.127
w10_mean	470	0.112
solar_mean	108	0.013
precip_sd	56	0.075
solar_sd	42	0.014
precip_mean	39	0.118

- **Mean Abs. Impact** ($|m_{d,t}X_{d,t}|$): average magnitude of each driver's contribution, in megawatts.
- **Mean Share**: average proportion of total absolute variation in the residual load explained by each variable.
- **Interpretation:** Temperature variability (`t2m_sd`) dominates load fluctuations ($\approx 23\%$), followed by wind variability at 100 m and 10 m. Solar and precipitation variables show limited, episodic influence.

Impact overtime

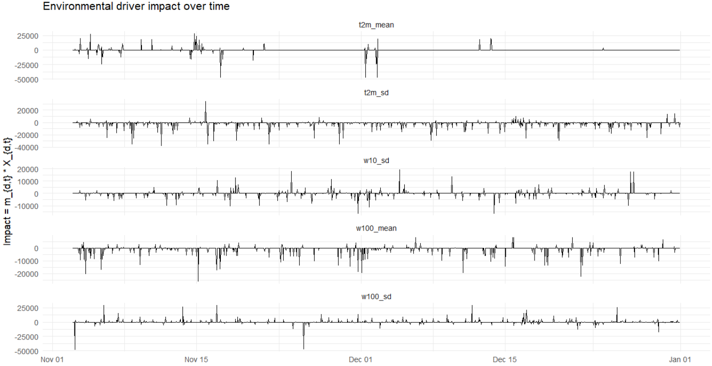

Figure 7: The impact overtime of different environmental predictor

Province-wise Dominant Environmental Predictors

Table 3: Top-ranked environmental drivers by province (mean absolute impact in MW).

Province	Driver	Mean Abs. [MW]	Mean Share
Bari	t2m_sd	125	0.166
Barletta-Andria-Trani	t2m_sd	201	0.177
Brindisi	w100_sd	78	0.094
Campobasso	t2m_sd	188	0.180
Foggia	t2m_sd	184	0.171
Isernia	w100_sd	85	0.115
Lecce	w100_sd	84	0.093
Matera	t2m_sd	221	0.183
Potenza	t2m_sd	148	0.171
Taranto	w100_sd	70	0.095

- **t2m_sd**: temperature variability dominates inland provinces (Calabria, Basilicata, Apulia).
- **w100_sd**: wind variability dominates coastal and open-area provinces.
- Mean absolute impacts are in **megawatts (MW)**, while mean shares represent the fraction of total residual variation explained.

Wind driver and Wind Installed Capacity

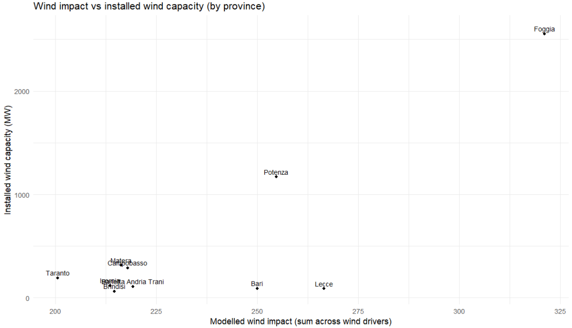

Figure 8: Comparison between installed and modelled wind capacity.

Concluding Remarks

Summary.

- Developed a **hierarchical modelling framework** combining a calendar model $f_{\text{cal}}(z_t)$ and an environmental model $g_{\text{env}}(X_t)$.
- Introduced a **derivative-based residual allocation** to ensure reconciliation between provincial and aggregate forecasts.
- Demonstrated good predictive performance and **interpretability through sensitivity measures**.

Empirical Insights.

- Temperature variability (`t2m_sd`) and wind intensity (`w100_mean`, `w100_sd`) are the key environmental drivers.
- The ranking of wind-related impacts is broadly consistent with **Terna's installed wind capacity** by province.
- This coherence validates the model's ability to **capture physical exposure and infrastructure effects**.

Next Steps.

- Extend the analysis to **photovoltaic and hydro components**.
- Incorporate **uncertainty quantification** via probabilistic forecasting.
- Use reconciled forecasts for **policy evaluation and investment planning**.