Online hierarchical forecasting for power consumption data

ENBIS - September 13, 2021

Margaux Brégère

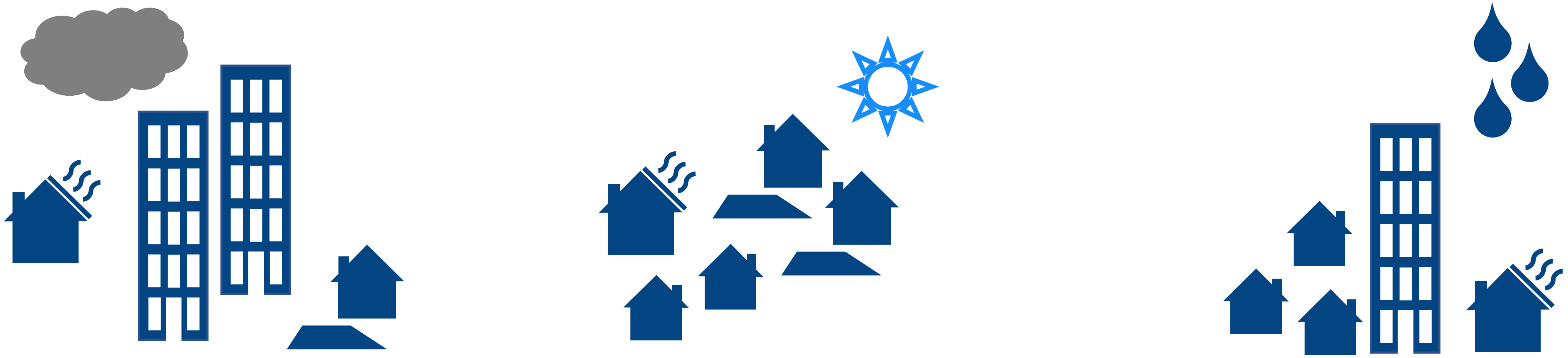
Joint work with Malo Huard

Presentation based on the article [Brégère, M., & Huard, M. (2021). Online hierarchical forecasting for power consumption data. International Journal of Forecasting.]

EDF

Motivation

Electricity forecasting at various aggregated levels



- Benchmark forecasts at each aggregated levels → Classical technics (GAM)
- Correlated time series (e.g., consumption of surrounding regions may be close) → Aggregation
- Connected times series through summation constraints (e.g., the global consumption is the sum of each region's consumption) → Projection

Literature discussion

Aggregation = combination of forecasts independently of their generating process

- Introduced by Vovk (1990), Cover (1991), and Littlestone and Warmuth, (1994).
- Effective at predicting
  - Time series (e.g., Mallet, Stoltz, & Mauricette, 2009)
  - Electricity consumption (Devaine, Gaillard, Goude, & Stoltz, 2013 and Gaillard, Goude, and Nedellec, 2016 - forecasting competition won)
- Recently extended to the hierarchical setting (Goehry, Goude, Massart, and Poggi, 2020)

Hierarchical forecasting

- Bottom-up (Dunn, Williams, and DeChaine, 1976) and top-down approaches (Gross and Sohl, 1990)
- Reconciliation of the set of forecasts with
  - Orthogonal or oblique projection (Wickramasuriya, Athanasopoulos, and Hyndman, 2019 – general minimum trace – MinT – algorithm)
  - Game-theoretically procedure (Van Erven and Cugliari, 2015)

Modeling hierarchical relationships

$$Y_t^{TOT} = y^{\cdot\,,1} + y^{\cdot\,,2} \ldots + y^{\cdot\,,k}$$

$$y_t^{\cdot\,,j} = y^{1,j} + y^{2,j} \ldots + y^{n,j}, \quad \forall j = 1, \ldots, k$$

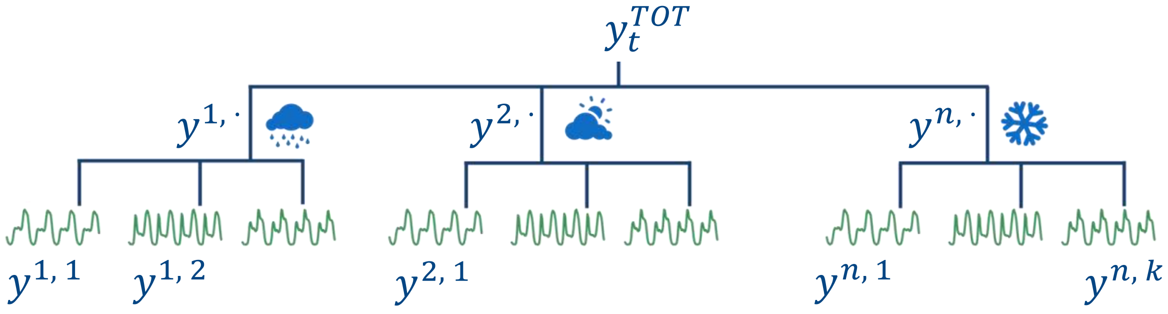
$$y_t^{TOT} = y^{1,\cdot} + y^{2,\cdot} \ldots + y^{n,\cdot}$$

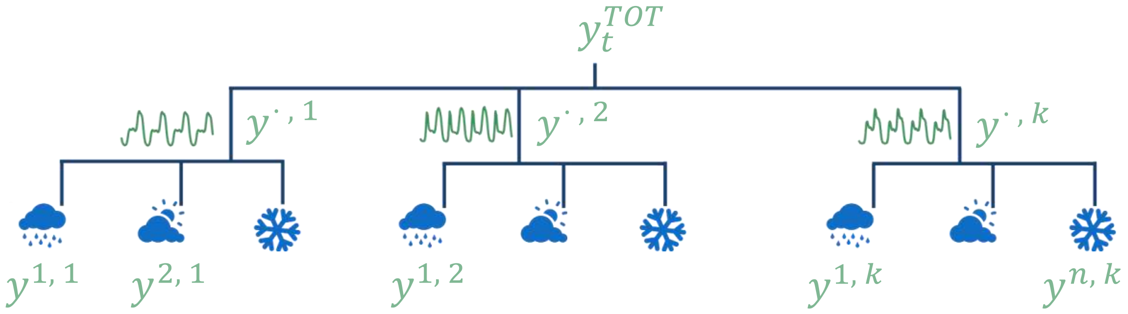
$$y_t^{i,\cdot} = y^{i,1} + y^{i,2} \ldots + y^{i,k}, \ \forall\, i = 1, \ldots, n$$

$$\mathbf{K}\mathbf{y}_t = \mathbf{0}$$

with

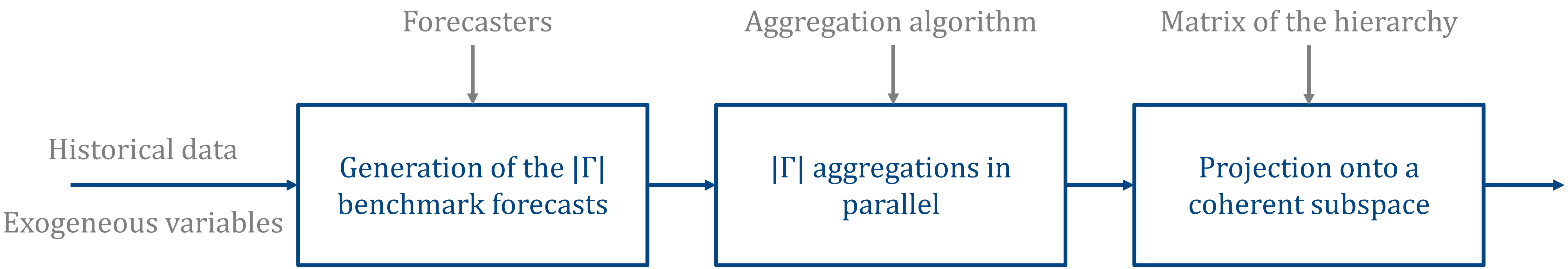
$$\mathbf{K} = \begin{pmatrix} -1 & & & & \overbrace{1 \ \cdots \ 1}^{k} & & & \\ & -1 & & & & \overbrace{1 \ \cdots \ 1}^{n} & & \\ & & \ddots & & & & \ddots & \\ & & & -1 & & & & \overbrace{1 \ \cdots \ 1}^{n} \\ -1 & & & & \overbrace{1 \ \cdots \ 1}^{n} & & & \\ & & & & -1 & 1 & & 1 \\ & & & & \ddots & \ddots & \cdots & \ddots \\ & & & & -1 & 1 & & 1 \end{pmatrix}$$





Three-step forecasting approach

$$\mathbf{y}_t = \Big(\underbrace{y_t^{Tot}}_{\text{All households}}, \underbrace{y_t^{\cdot,1}, y_t^{\cdot,2}, \dots, y_t^{\cdot,k}}_{\text{Households of behavioral cluster}}, \underbrace{y_t^{1,\cdot}, y_t^{2,\cdot}, \dots, y_t^{n,\cdot}}_{\text{Households of same region}}, \underbrace{y_t^{1,1}, y_t^{1,2}, \dots, y_t^{i,j}, \dots, y_t^{n,k}}_{\text{Households of same behavioral cluster and region}}\Big)^{\mathrm{T}}$$



Γ = nodes of the trees

Algorithm

Input

- Set Γ and constraint matrix $\mathbf{K}$
- Benchmark forecast generation method = Generalized additive models
- Aggregation algorithm $\mathcal{A}$

Compute the orthogonal projection matrix $\Pi_{\mathbf{K}} = \left(\mathrm{I}_{|\Gamma|} - \mathbf{K}^{\mathrm{T}}\left(\mathbf{K}\mathbf{K}^{\mathrm{T}}\right)^{-1}\mathbf{K}\right)$

For $\gamma \in \Gamma$ do

- Create a copy of $\mathcal{A}$ denoted $\mathcal{A}^{\gamma}$
- For $t = 1, \ldots, T$ do
    - Generate benchmark forecasts $\mathbf{x}_t = \left(x_t^{\gamma}\right)_{\gamma \in \Gamma}$
    - For $\gamma \in \Gamma$ do
        - $\mathcal{A}^{\gamma}$ outputs $\hat{y}_t^{\gamma} = \mathbf{u}_{\mathrm{t}}^{\gamma} \cdot \mathbf{x}_t$
    - Collect forecasts $\hat{\mathbf{y}}_t = \left(\hat{y}_t^{\gamma}\right)_{\gamma \in \Gamma}$ and project them $\tilde{\mathbf{y}}_t = \Pi_{\mathbf{K}}\, \hat{\mathbf{y}}_t$
    - For $\gamma \in \Gamma$ do
        - $\mathcal{A}^{\gamma}$ observes y_t^{γ} and computes $\mathbf{u}_{\mathrm{t+1}}^{\gamma}$
    - Suffer the prediction error $\frac{1}{|\Gamma|}\sum_{\gamma \in \Gamma}\left(y_t^{\gamma} - \tilde{y}_t^{\gamma}\right)^2$

Assessment of the forecasts

To minimize the average prediction error

$$\tilde{L}_T = \frac{1}{|\Gamma|}\sum_{t=1}^{T}\frac{1}{|\Gamma|}\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\|^2 = \frac{1}{T|\Gamma|}\sum_{t=1}^{T}\sum_{\gamma\in\Gamma}\left(y_t^{\gamma} - \tilde{y}_t^{\gamma}\right)^2$$

is equivalent to minimize, for a given set D, the regret

Linear combination of benchmark forecasts

$$R_T(D) = T|\Gamma| \times \tilde{L}_T - \underbrace{\min_{\mathbf{U}\in C_{|D}} \sum_{t=1}^{T}\|\mathbf{y}_t - \mathbf{U}^T\mathbf{x}_t\|^2}_{\text{Approximation error}}$$

with C the set of matrices $\mathbf{U}$ such that the predictions $\mathbf{U}^T\mathbf{x}_t$ satisfy the hierarchical constraints and $C_{|D}$ the set such that the matrices also have all their rows in D:

γ – row of $\mathbf{U}$

$$C_{|D} = \left\{\mathbf{U} \mid \forall\, \mathbf{x} \in \mathbb{R}^{|\Gamma|}, \mathbf{K}\mathbf{U}^{T}\mathbf{x} = \mathbf{0} \quad \text{and} \quad \forall\, \gamma \in \Gamma, u^{\gamma} \in D\right\}$$

Theorem

If, for any D such that, for any $\gamma \in \Gamma$, for $T > 0$, for any $\mathbf{x}_1, \ldots, \mathbf{x}_T$ and $y_1^\gamma, \ldots y_T^\gamma$, Algorithm $\mathcal{A}^\gamma$ provides a regret bound of the following form

$$\sum_{t=1}^{T} \left(y_t^\gamma - \hat{y}_t^\gamma\right)^2 - \min_{\mathbf{u}^\gamma \in D} \sum_{t=1}^{T} \left(y_t^\gamma - \mathbf{u}^\gamma \cdot \mathbf{x}_t\right)^2 \leq B$$

then,

$$R_T(D) = \sum_{t=1}^{T} \|\mathbf{y}_\mathrm{t} - \tilde{\mathbf{y}}_\mathrm{t}\|^2 - \min_{\mathbf{U} \in \mathcal{C}_{|D}} \sum_{t=1}^{T} \|\mathbf{y}_\mathrm{t} - \mathbf{U}^T \mathbf{x}_t\|^2 \leq |\Gamma| B$$

Sketch of the proof

Pythagorean theorem

$$\sum_{t=1}^{T} \|\mathbf{y}_{\mathrm{t}} - \tilde{\mathbf{y}}_{\mathrm{t}}\|^2 - \sum_{t=1}^{T} \|\mathbf{y}_{\mathrm{t}} - \mathbf{U}^T \mathbf{x}_t\|^2 \leq \sum_{t=1}^{T} \|\mathbf{y}_{\mathrm{t}} - \hat{\mathbf{y}}_t\|^2 - \sum_{t=1}^{T} \|\mathbf{y}_{\mathrm{t}} - \mathbf{U}^T \mathbf{x}_t\|^2$$

$$= \sum_{\gamma \in \Gamma} \sum_{t=1}^{T} \left| y_t^\gamma - \hat{y}_t^\gamma \right|^2 - \sum_{\gamma \in \Gamma} \sum_{t=1}^{T} \left| y_t^\gamma - \mathbf{u}^\gamma \cdot \mathbf{x}_t \right|^2$$

$$\leq |\Gamma| B$$

Regret bound of the aggregation algorithm

Example of an aggregation algorithm: ML-Poly

Polynomially weighted average forecaster with multiple learning rates with gradient trick (Gaillard, 2015) competes against the best convex combination of benchmark forecast

$D =$ simplex of dimension $|\Gamma|$

Under boundedness assumptions on observations y_t^γ and benchmark forecasts x_t^γ, the regret satisfies

$$R_T \leq O\left(|\Gamma|^{2/3}\sqrt{T \ln T}\right)$$

The underlying real data set

Electrical consumption records of 1 545 households over the period from April 20, 2009 to July 31, 2010.

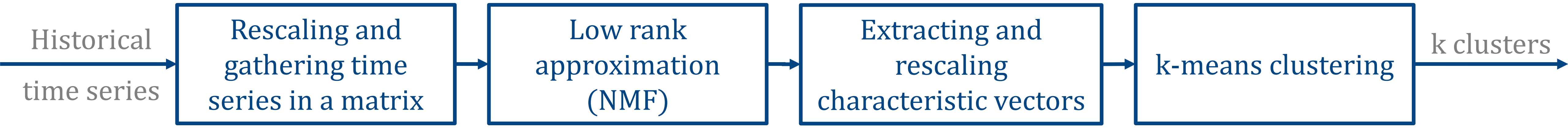
Variable	Description	Range / Value
Date	Current time	From April 20, 2009 to July 31, 2010 (half-hourly)
Consumption	Power consumption	From 0.001 to 900 kWh
Region	UK NUTS of level 3	UK- H23, -J33, -L15, -L16, -L21, -M21, or -M27
Temperature	Air temperature	From -20 $^\circ$C to 30 $^\circ$C
Visibility	Air visibility	From 0 to 10 (integer)
Humidity	Air humidity percentage	From 0% to 100%
Half-hour	Half-hour of the day	From 1 to 48 (integer)
Day	Day of the week	From 1 (Monday) to 7 (Sunday) (integer)
Position in the year	Linear values	From 0 (Jan 1, 00:00) to 1 (Dec 31, 23:59)
Smoothed temperature	Exponential smoothing	From -20 $^\circ$C to 30 $^\circ$C

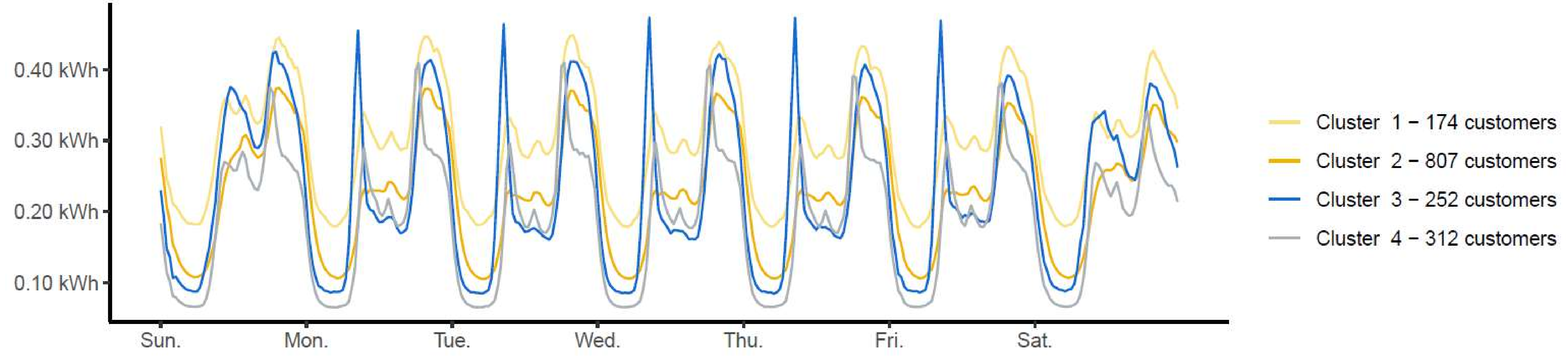
From Energy Demand Research Project (Power consumption of ~18,000 UK households at half-hourly steps over two years) — Date, Consumption, Region, Temperature

From NOAA (National Oceanic and Atmospheric Administration) — Visibility, Humidity

Created — Half-hour, Day, Position in the year, Smoothed temperature

Behavioral segmentation of the households





Experiment design

- Double segmentation:
  - Geographical, based on region information
  - Behavioral
- Meteorological data:
  - One per region
  - Convex combination of local meteorological variables for levels containing several regions
- Benchmark creation: Generalized Additive Models
- Observation and benchmark standardization for Aggregation (ML-Poly)
- Operational constraint: Half-hourly predictions with one-day-delayed observations

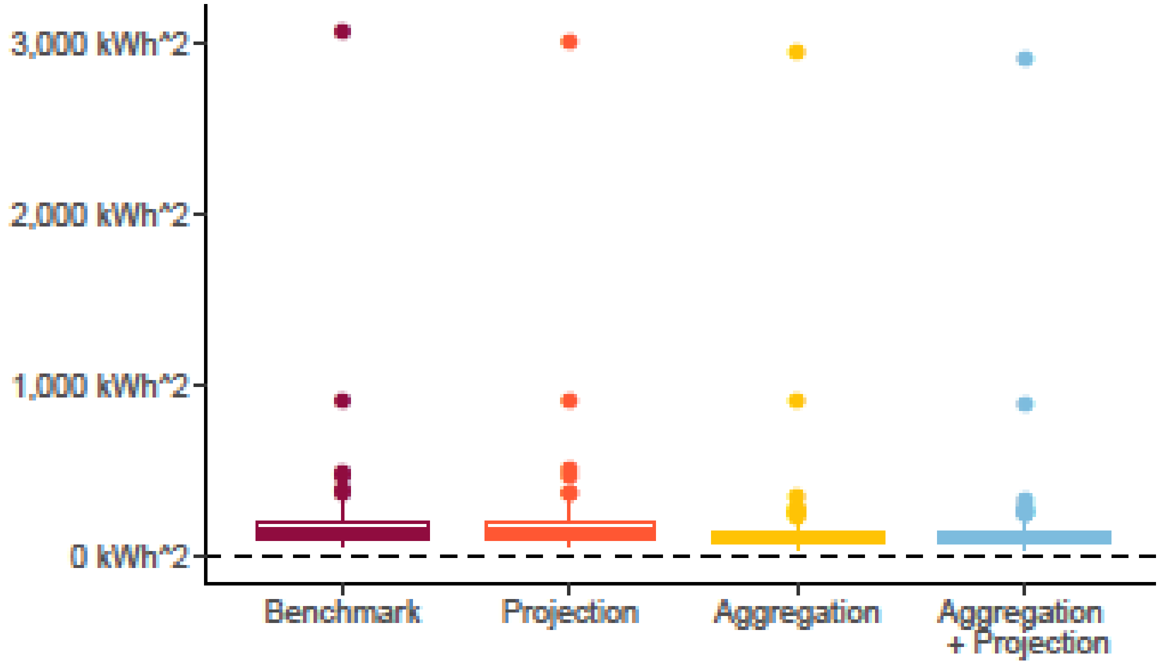
	Start date	End date
Behavioral segmentation		
Benchmark generation model training	April 20, 2009	April 19, 2010
Benchmark and observation standardization		
Initialization of the aggregation	April 20, 2010	April 30, 2010
Model evaluation	May 1, 2010	July 31, 2010

Results – Mean Squared Error (MSE) on test period

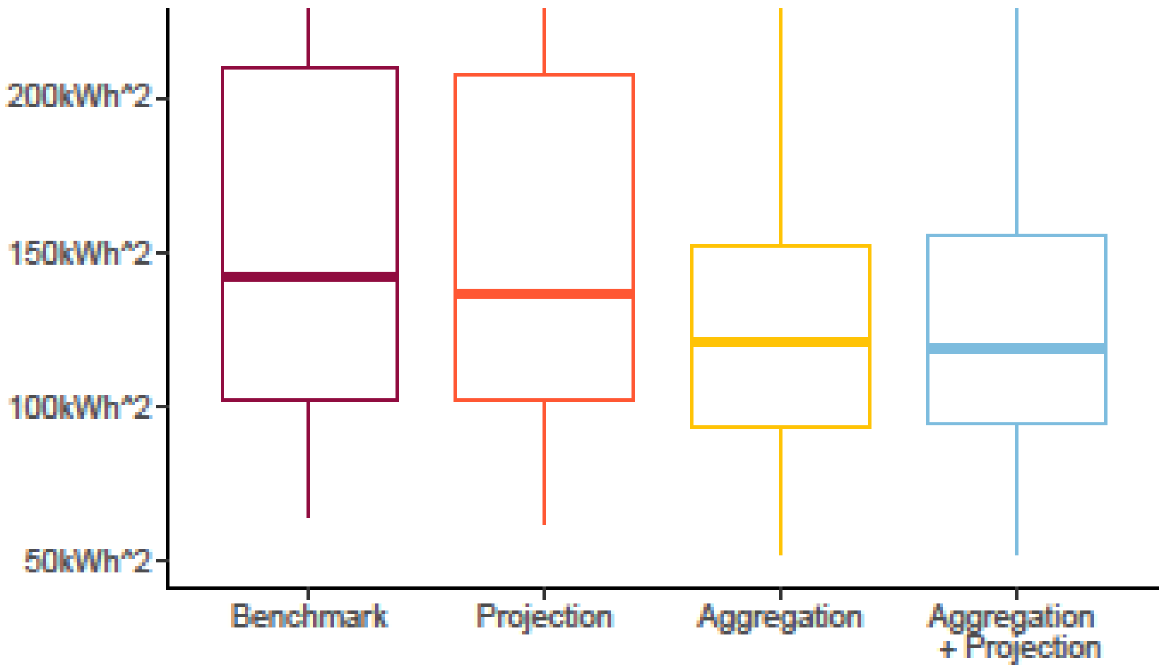
	All aggregated levels	Global	Local
Benchmark	455.5	205.8	66.3
Projection	450.7	200.8	66.3
Aggregation	397.9	172.0	61.2
Aggregation + Projection	396.0	170.3	61.1

Clustering	Benchmark	Bottom-up	Projection	Aggregation	Aggregation + Projection
Region	205.8	189.9	201.3	187.8	186.7
Behavior	-	208.4	205.2	179.3	179.3
Region + Behavior	-	201.0	200.8	172.0	**170.3**

Results – Mean Squared Error (MSE) on test period



Original boxplots



Boxplots trimmed at 220 kWh^2

Distribution over the test period of the daily mean squared error of global consumption for the four strategies "Benchmark", "Projection", "Aggregation", and "Aggregation + Projection"

Conclusion

A three-step approach to forecasting electricity consumption time series at different levels of household aggregation and linked by hierarchical constraints with

- Theoretical result
- Experimental results which suggest that
  - Aggregation and projection improve the forecasts overall (confirming the theoretical result)
  - And forecasts of both global and local consumption
  - Global consumption forecasts improve when using forecasts of groups of households segmented according to their region (with their own weather conditions) and according to their behavior

Thank you for your attention!
Questions?