Online reconciliation of electricity consumption forecasts

Margaux Brégère

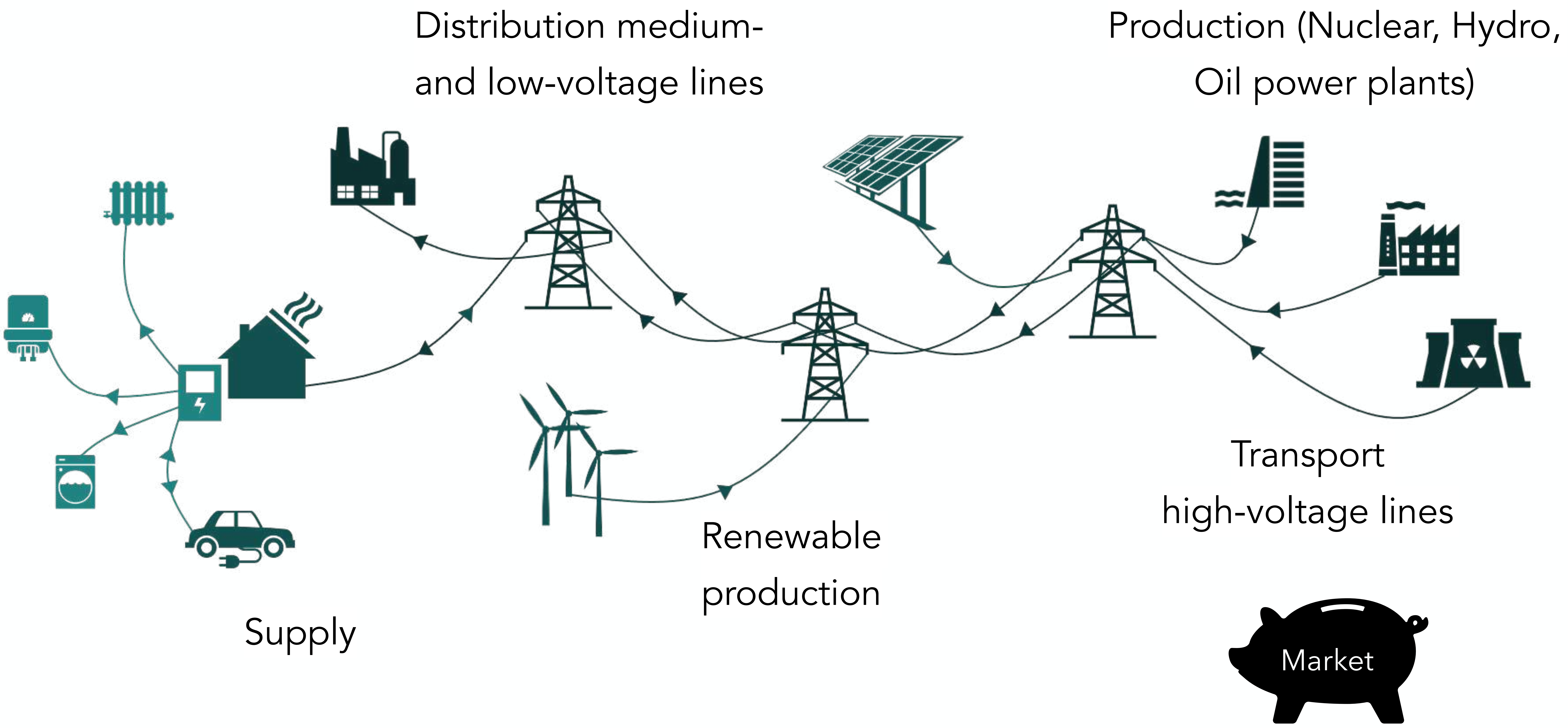
Séminaire Parisien - IHP

April 22nd 2024

LPSM
LABORATOIRE DE PROBABILITÉS STATISTIQUE & MODÉLISATION

Motivations

Electrical system



Electricity consumption forecasting

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

Forecast demand and adapt production accordingly

From dis-aggregated to aggregated level

- Dis-aggregated level
  - Modeling new electrical uses (auto-consumption, electrical vehicles)
  - Designing demand response solutions
  - ⚠ Smart meters data is highly sensitive and erratic ☞ simulation models

- Neighborhood / city level
  - Managing networks locally (Smart Grids)
  - Dispatching electricity at junctions between transport (high-voltage lines) and distribution (medium- and low-voltage lines) networks

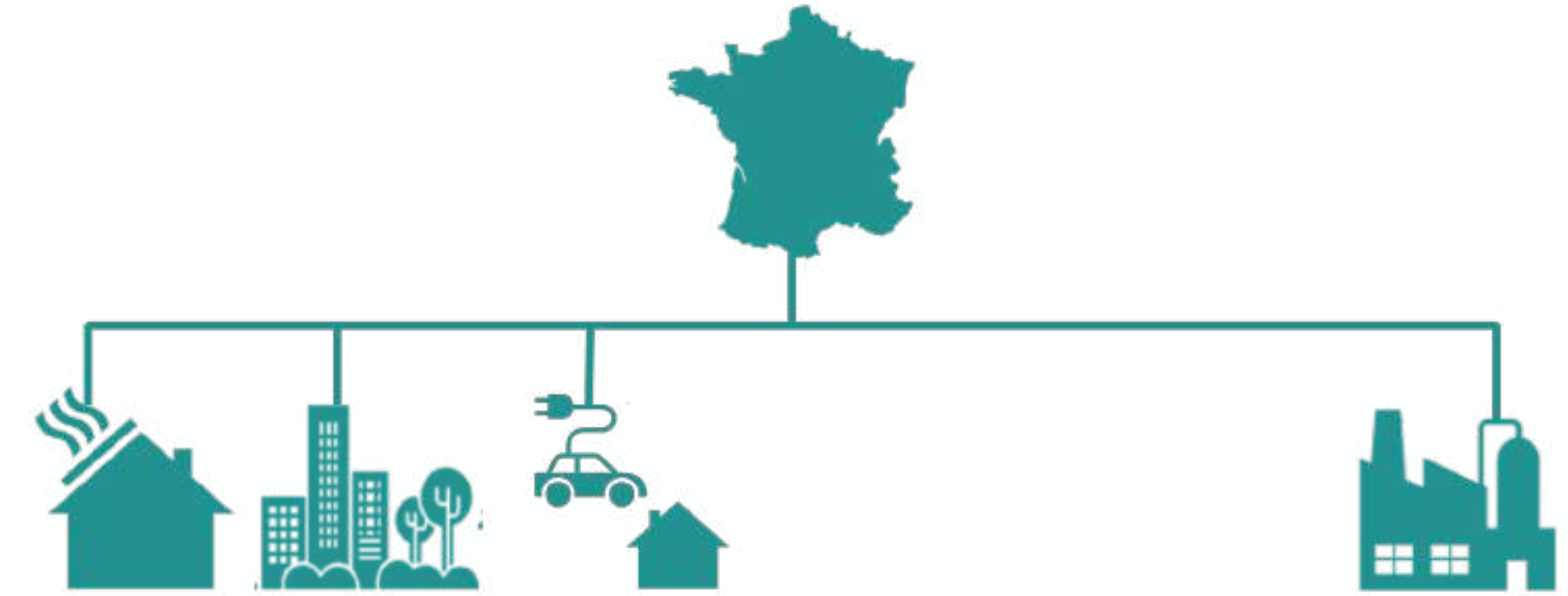
- National level
  - Managing the overall balance
  - Planning cross-border exchanges

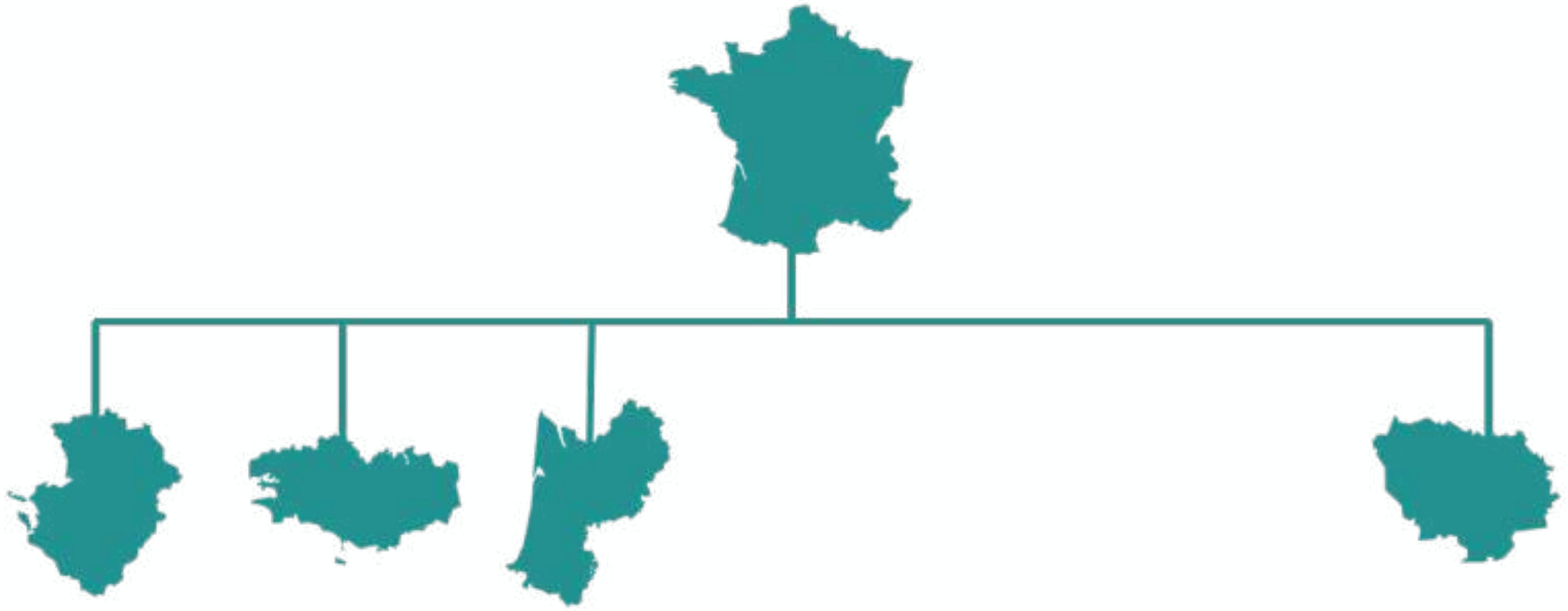
Using various aggregated level forecasts

- Benchmark forecasts at each aggregated levels
  - France: easier to forecast (smoother)
  - Consumer type: same behavior
  - Regions: local weather

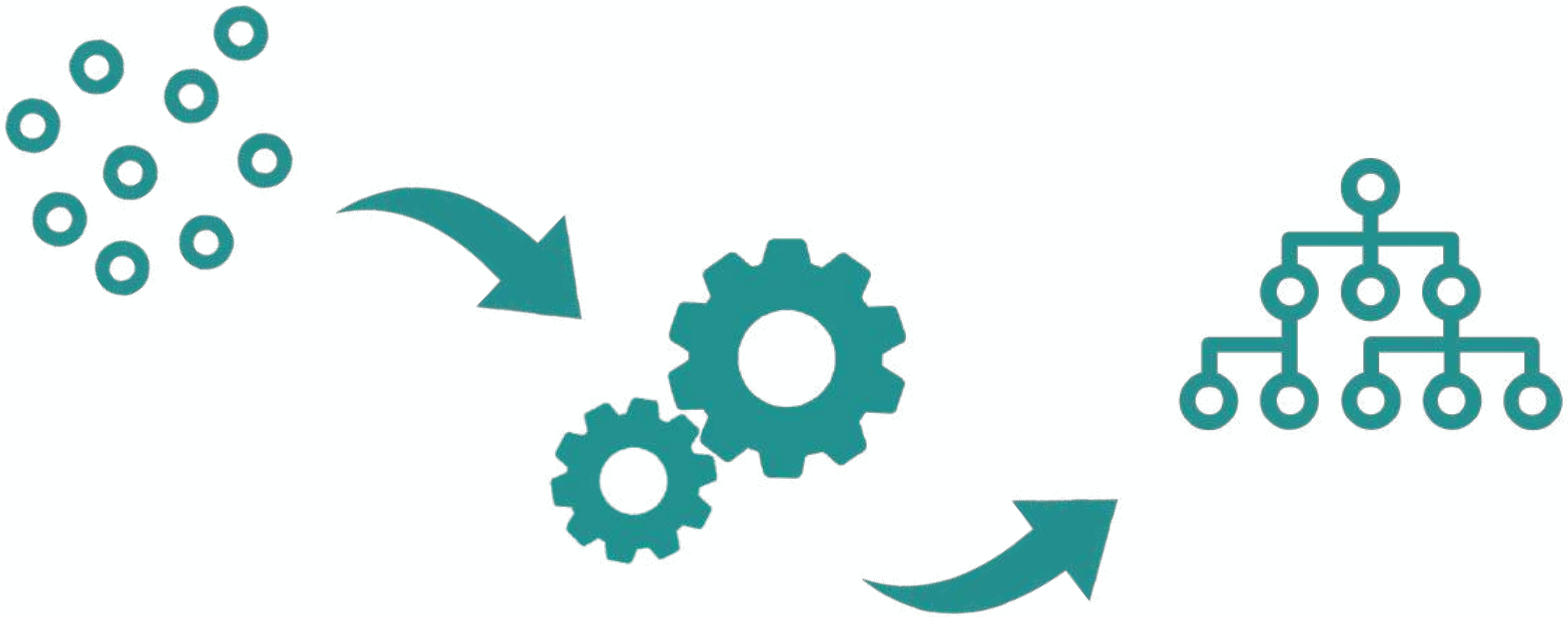
- Correlated and connected times series through summation constraints

→ Reconciliation

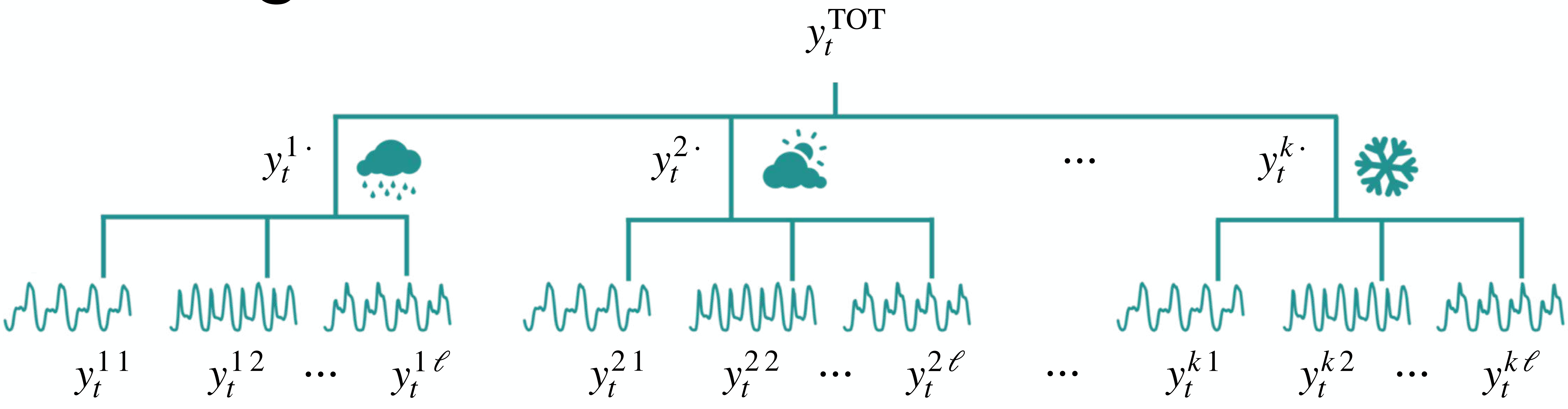




Methods



Modeling



$$y_t = \Big(\underbrace{y_t^{\mathrm{TOT}}}_{\text{All households}},\underbrace{y_t^{1\,\cdot},y_t^{2\,\cdot},\ \ldots,y_t^{k\,\cdot}}_{\text{Households of same region}},\underbrace{y_t^{1\,1},y_t^{1\,2},\ \ldots,y_t^{1\,\ell},y_t^{2\,1},\ \ldots,y_t^{k\,\ell}}_{\text{Households of same behavioral cluster and region}}\Big)^{\mathrm{T}}$$

Modeling: summing matrix

With $m = 1 + k + k\ell$ the number of all times series to forecast and $n = k\ell$ the number of most disaggregated level time series, the summing matrix $S \in \{0,1\}^{m \times n}$ is:

$$
S = \begin{array}{cl}
\begin{array}{ccccccccccccc}
\multicolumn{4}{c}{\overbrace{\qquad\qquad\qquad\qquad}^{\text{Region }1}} & \multicolumn{4}{c}{\overbrace{\qquad\qquad\qquad\qquad}^{\text{Region }2}} & & \multicolumn{4}{c}{\overbrace{\qquad\qquad\qquad\qquad}^{\text{Region }k}}
\end{array} & \\
\left(\begin{array}{ccccccccccccc}
1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 & \dots\dots & 1 & 1 & \dots & 1 \\
1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots\dots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots\dots & 0 & 0 & \dots & 0 \\
 & & & & & & \vdots & & & & & & \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots\dots & 1 & 1 & \dots & 1 \\
 & & & & & & I_n & & & & & &
\end{array}\right) &
\begin{array}{l}
\} \ \text{All households} \\
\left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} \ \text{Households of same region} \\
\} \ \text{Households of same behavioral cluster and region}
\end{array}
\end{array}
$$

With the most disaggregated level time series $b_t = \left(y_t^{1\,1}, y_t^{1\,2}, \dots, y_t^{1\,\ell}, y_t^{2\,1}, \dots, y_t^{k\,\ell}\right)^{\mathrm{T}}$, $y_t = Sb_t$

Another modeling: constraint matrix

With c the number of summation constraints

$$K = \begin{pmatrix} -1 & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 & \dots\dots & 1 & 1 & \dots & 1 \\ 0 & -1 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots\dots & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots\dots & 0 & 0 & \dots & 0 \\ & & & & & & \vdots & & & & & & & & & & & \\ 0 & 0 & 0 & \dots & -1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots\dots & 1 & 1 & \dots & 1 \end{pmatrix}$$

$K \in \{-1, 0, 1\}^{m \times c}$

$y_t \in \mathrm{Ker}(K)$

Reconciliation

From base forecasts $\hat{y}_t$, construct reconciled forecasts $\tilde{y}_t$ such that $\tilde{y}_t = S\tilde{b}_t$ or equivalently $\tilde{y}_t K = \mathbf{0}_c$

All linear reconciliation methods can be written as

$$\tilde{y}_t = SP\hat{y}_t\,,$$

for some appropriately selected matrix $P \in \mathbb{R}^{n \times m}$

P is to project the base forecasts into bottom level disaggregated forecasts which are then summed by S

State of the art - Reconciliation

- Bottom-up (Dunn, Williams, and DeChaine, 1976)

$$\hat{y}_t = \left(\square, \square, \square, \ldots, \square, \hat{y}_t^{1\,1}, \hat{y}_t^{1\,2}, \ldots, \hat{y}_t^{1\,\ell}, \hat{y}_t^{2\,1}, \ldots, \hat{y}_t^{k\,\ell}\right)^{\mathrm{T}} \to \tilde{y}_t = \left(\sum_{i=1}^{k}\sum_{j=1}^{\ell}\hat{y}_t^{ij}, \sum_{j=1}^{\ell}\hat{y}_t^{1j}, \sum_{j=1}^{\ell}\hat{y}_t^{2j}, \ldots, \sum_{j=1}^{\ell}\hat{y}_t^{kj}, \hat{y}_t^{1\,1}, \hat{y}_t^{1\,2}, \ldots, \hat{y}_t^{1\,\ell}, \hat{y}_t^{2\,1}, \ldots, \hat{y}_t^{k\,\ell}\right)^{\mathrm{T}}$$

With $P = \left(\mathbf{0}_{n\times(m-n)} \middle| I_n\right)$, $\tilde{y}_t = SP\hat{y}_t$

- Top-down approaches (Gross and Sohl, 1990)

$$\hat{y}_t = \left(\hat{y}_t^{\mathrm{TOT}}, \square, \square, \ldots, \square, \square, \square, \ldots, \square, \square, \ldots, \square\right)^{\mathrm{T}} \to \tilde{y}_t = \hat{y}_t^{\mathrm{TOT}}\left(1, \sum_{j=1}^{\ell}p^{1j}, \sum_{j=1}^{\ell}p^{2j}, \ldots, \sum_{j=1}^{\ell}p^{kj}, p^{1\,1}, p^{1\,2}, \ldots, p^{1\,\ell}, p^{2\,1}, \ldots, p^{k\,\ell}\right)^{\mathrm{T}}$$

With $P = \left(p \middle| \mathbf{0}_{n\times(m-n)}\right)$ where $p = \left(p^{1\,1}, p^{1\,2}, \ldots, p^{1\,\ell}, p^{2\,1}, \ldots, p^{k\,\ell}\right)^{\mathrm{T}}$, $\tilde{y}_t = SP\hat{y}_t$

State of the art - Reconciliation

- Orthogonal ($\Sigma = I_m$) or oblique projection (Wickramasuriya, Athanasopoulos, and Hyndman, 2019)

$$\tilde{y}_t = SP_\Sigma \hat{y}_t \text{ with } P_\Sigma = \left(S^\mathrm{T}\Sigma^\dagger S\right)^{-1} S^\mathrm{T}\Sigma^\dagger$$

Minimum trace (MinT) reconciliation: assuming base forecast errors are stationary conditionally to data observed,

Then the optimal reconciliation (which minimizes the variance of the reconciled forecast errors) is obtained for

$$\Sigma^\star = \mathbb{E}\left[\left(y_t - \hat{y}_t\right)\left(y_t - \hat{y}_t\right)^\mathrm{T} \middle| y_1, \ldots y_{t-d}\right]$$

$\Sigma^\star$ is the variance-covariance matrix of the base forecast errors

Remarks:

- Stationarity implies unbiased base forecast
- Reconciled forecasts will also be unbiased $\Leftrightarrow SPS = S$
- Challenge: estimating the precision matrix $\Sigma^\dagger$
- Orthogonal projection: $\Sigma = I_m \quad \Leftrightarrow \quad \tilde{y}_t = \Pi_K \hat{y}_t$ with $\Pi_K = \left(I_m - K^\mathrm{T}\left(KK^\mathrm{T}\right)^{-1}K\right)$

State of the art - Reconciliation

- Game-theoretically procedure (Van Erven and Cugliari, 2015)

Game-theory set-up: reconciled forecast are obtained by solving the minimax optimization problem

$$\tilde{y}_t \in \arg\min_{\tilde{y} \in \mathscr{A}} \max_{y \in \mathscr{A} \cap \mathscr{B}} \left\{ \ell(y, \tilde{y}) - \ell(y, \hat{y}_t) \right\}$$

with $\mathscr{A} = \{y \,|\,$ sommation constrains hold$\}$ and $\mathscr{B}$ specifies information (confidence intervals around base forecast)

Result: if $\mathscr{B}$ is a closed convex set with $\mathscr{A} \cap \mathscr{B} \neq \varnothing$ and $\ell(y_t, \hat{y}_t) = \|Ay_t - A\hat{y}_t\|_2^2$,

$\tilde{y}_t = \mathrm{Proj}_{\mathscr{A} \cap \mathscr{B}}(\hat{y}_t)$ (L2-projection after scaling according to A) and $\ell(y_t, \tilde{y}_t) \leq \ell(y_t, \hat{y}_t)$ (Pythagorean inequality)

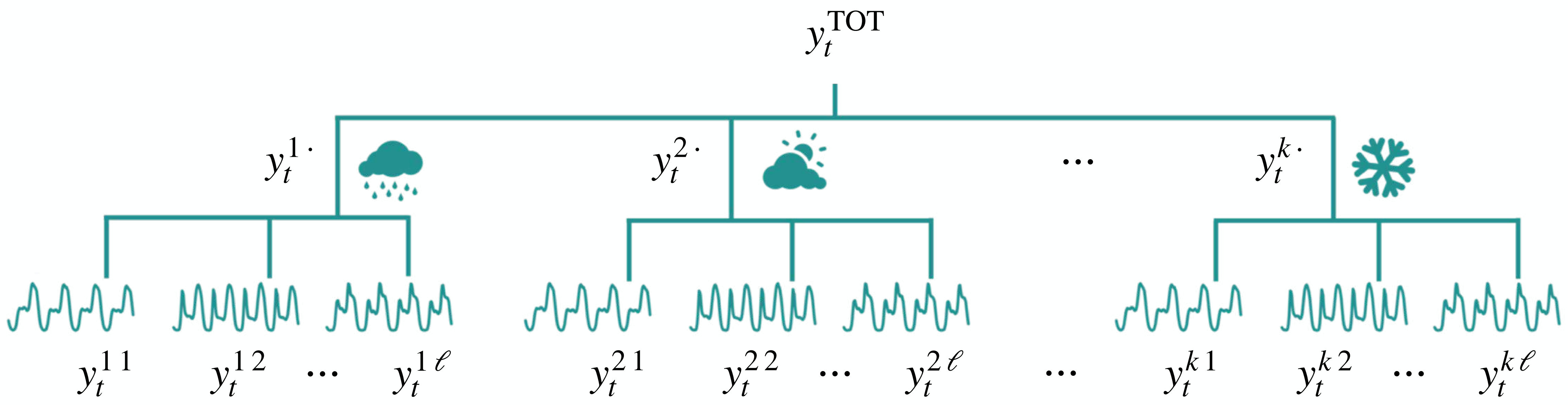
(no assumption on base forecasts)

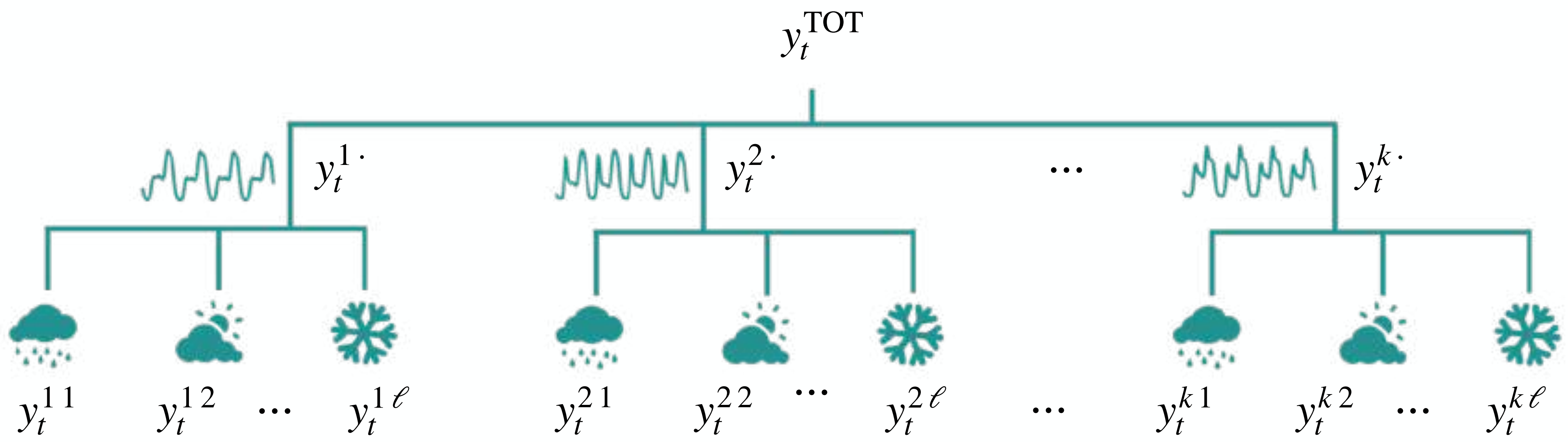
- Deep Learning approach (Leprince et al., 2023)

Train a neural network model with loss function: $\alpha \ell(y_t, \hat{y}) + (1 - \alpha)\|\hat{y} - SP_{\Sigma}\hat{y}\|_2^2$

Online Hierarchical Forecasting for Power Consumption Data

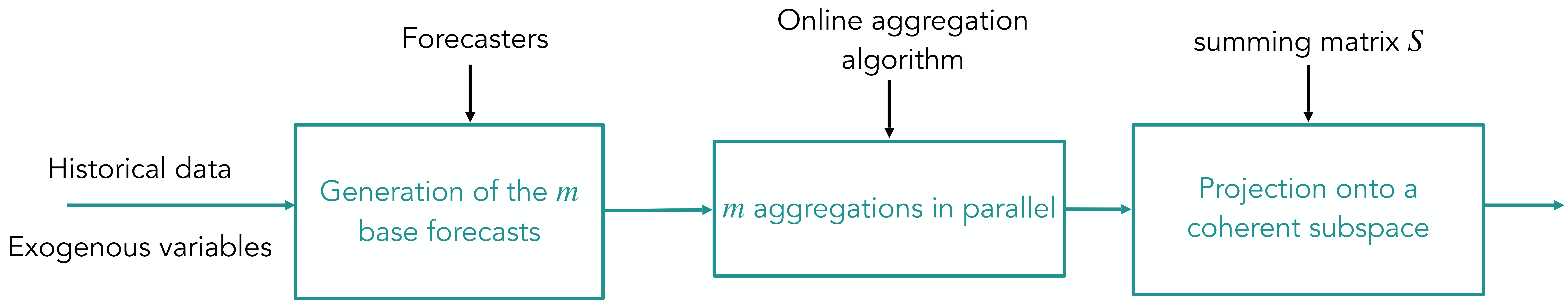
Joint work with Malo Huard





$$y_t = \Big(\underbrace{y_t^{\mathrm{TOT}}}_{\substack{\text{All}\\ \text{households}}}, \underbrace{y_t^{1\,\cdot}, y_t^{2\,\cdot}, \ldots, y_t^{k\,\cdot}}_{\substack{\text{Households of}\\ \text{same region}}}, \underbrace{y_t^{\cdot\,1}, y_t^{\cdot\,2}, \ldots, y_t^{\cdot\,\ell}}_{\substack{\text{Households of}\\ \text{same behavioral}\\ \text{cluster}}}, \underbrace{y_t^{1\,1}, y_t^{1\,2}, \ldots, y_t^{1\,\ell}, y_t^{2\,1}, \ldots, y_t^{k\,\ell}}_{\substack{\text{Households of same}\\ \text{behavioral cluster and region}}}\Big)^{\mathrm{T}}$$

Three-step forecasting approach



Online aggregation: predict y_t from base forecasts (experts) $f_t^1, \ldots f_t^j$ with $p_t \cdot f_t = \sum_{i=1}^{j} p_t^j f_t^j$

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

Algorithm

Input

- Method for base forecast generation
- Online aggregation algorithm Agg
- Summation matrix S (or constraints matrix K)

For each level $\gamma \in \left\{\mathrm{TOT}, 1\ \cdot, 2\ \cdot, \ldots, k\ \cdot, \ \cdot\ 1, \ \cdot\ 2, \ \ldots, \cdot\ \ell, 11, 12, \ \ldots, k\ell\right\}$

- Create a copy of Agg denoted Agg^γ

For $t = 1,\ldots,T$

- Generate base forecast $\hat{y}_t = \left(\hat{y}_t^\gamma\right)_\gamma$
- For each level γ
  - Agg^γ outputs $\bar{y}_t^\gamma = p_t^\gamma \cdot \hat{y}_t$
- Collect forecast $\bar{y}_t = \left(\bar{y}_t^\gamma\right)_\gamma$ and reconcile them with orthogonal projection: $\tilde{y}_t = S\left(S^{\mathrm{T}}S\right)^{-1}S^{\mathrm{T}}\bar{y}_t$ (or $\tilde{y}_t = \Pi_K \bar{y}_t$)
- For each level γ
  - Agg^γ observes y_t^γ and update p_{t+1}^γ

Assessment of the forecasts

To minimize the average prediction error $L_T = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{m}\left\|y_t - \tilde{y}_t\right\|_2^2$ is equivalent to minimize, for a given set $\mathscr{D} \subset \mathbb{R}^m$,

$$R_T(\mathscr{D}) = Tm \times L_T - \underbrace{\min_{U \in C_{\mathscr{D}}} \sum_{t=1}^{T} \Big\| y_t - \overset{\text{Linear combination of base forecasts}}{U^{\mathrm{T}}\hat{y}_t} \Big\|_2^2}_{\text{Approximation error}}$$

with $C_{\mathscr{D}}$ the set of matrices such that the forecasts satisfy the summation constraints and have all their rows in $\mathscr{D}$: $C_{\mathscr{D}} = \left\{ U \in \mathbb{R}^{m\times m} \quad \middle| \quad \forall y \in \mathbb{R}^m, KU^{\mathrm{T}} = 0 \quad \text{and} \quad \forall \gamma, U_{\gamma\cdot} \in \mathscr{D} \right\}$

Theorem

If for any $\mathcal{D}$ such that, for any $\gamma \in \{\mathrm{TOT}, 1\ \cdot, 2\ \cdot, \ldots, k\ \cdot, \cdot\ 1, \cdot\ 2, \ldots, \cdot\ \ell, 11, 12, \ldots, k\ell\}$, for $T > 0$, for any $\hat{y}_1, \ldots, \hat{y}_t$ and any $y_1^\gamma, \ldots, y_t^\gamma$, Algorithm Agg^γ provides a regret bound of the following form:

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

then,

$$R_T(\mathcal{D}) = Tm \times L_T - \min_{U \in C_{\mathcal{D}}} \sum_{t=1}^{T} \left\| y_t - U^{\mathrm{T}} \hat{y}_t \right\|_2^2 \leq B \times m$$

Sketch of the proof: Pythagorean theorem + regret bound of the aggregation algorithm

Example with ML-Poly algorithm

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

$\implies \mathscr{D} = \Delta_m$

Under boundedness assumptions on observations y_t and base forecasts $\hat{y}_t$, the regret satisfies

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

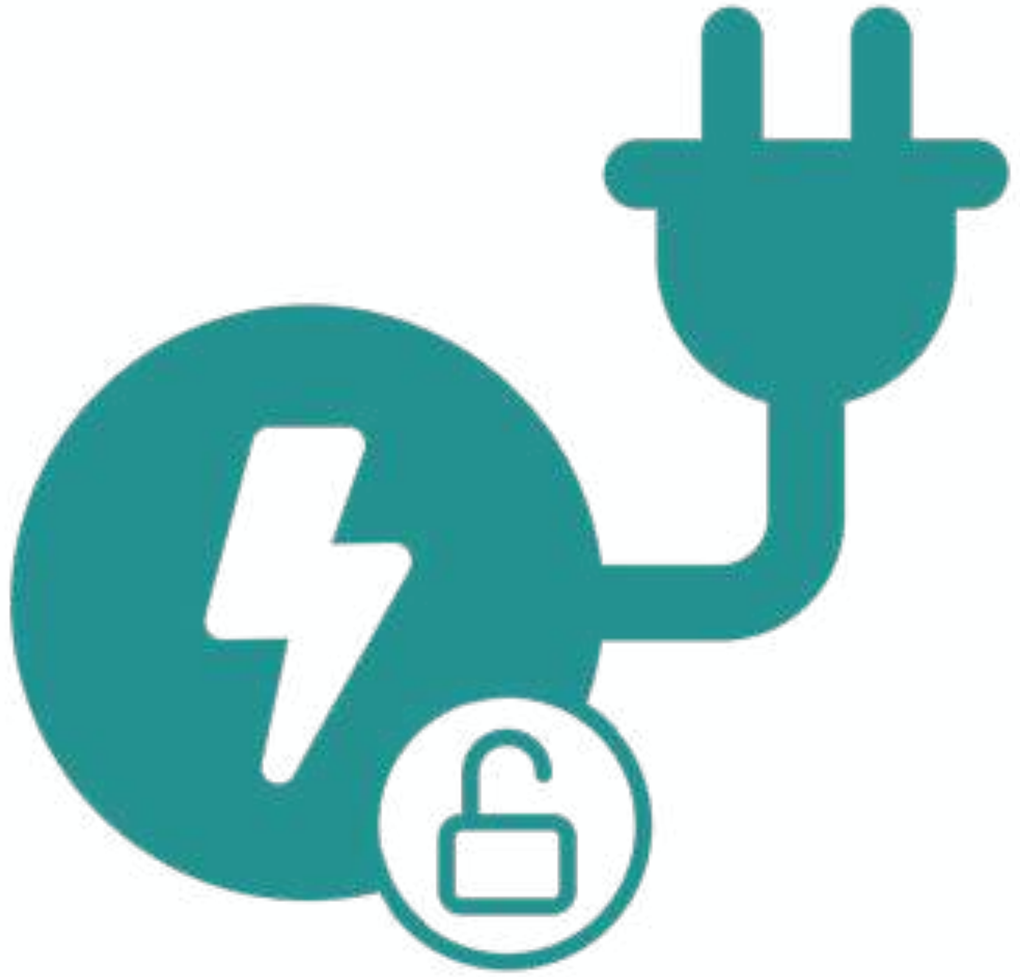
ML-Poly:

Initialization: $p_1^\gamma = \left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$

For $t = 1, \ldots,$ the weight vector $p_{t+1}^\gamma = (p_{t+1}^{\gamma,i})_i$ is defined as

$$p_{t+1}^{\gamma,i} \propto \left(\eta_t^{\gamma,i} \sum_{t=1}^{t} 2(\bar{y}_t^\gamma - y_t^\gamma)(\bar{y}_t^\gamma - \hat{y}_t^{\gamma,i})\right)_+ \text{ with } \eta_t^{\gamma,i} = \left(E + \sum_{t=1}^{t} \left(2(\bar{y}_t^\gamma - y_t^\gamma)(\bar{y}_t^\gamma - \hat{y}_t^{\gamma,i})\right)^2\right)^{-1}$$

Experiments



UK household electricity consumption

Underlying real data set

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

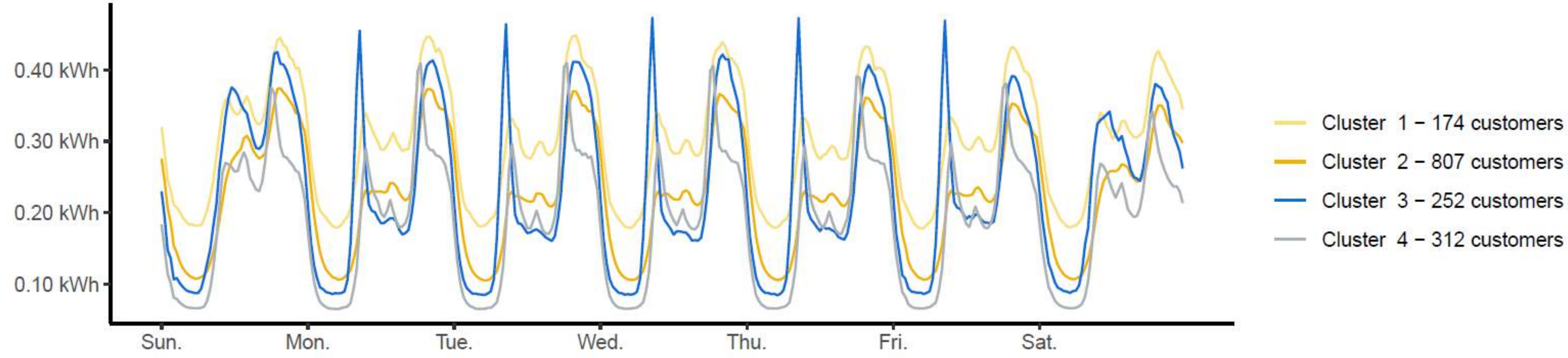
Variable	Description	Range / Value
Date	Current time	From April 20, 2009 to July 31, 2010 (half-hourly)
Consumption	Power consumption	From 0.001 to 900 kWh
Region	UK NUTS of level 3	UK- H23, -J33, -L15, -L16, -L21, -M21, or -M27
Temperature	Air temperature	From -20 °C to 30 °C
Visibility	Air visibility	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 °C to 30 °C

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

From NOAA (National Oceanic and Atmospheric Administration)

Created

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

Base forecasts generation: Generalized Additive Models

Online aggregation algorithm: ML-Poly → standardization of base forecasts and observations

Operational constraint: half-hourly predictions with one-day-delayed observations

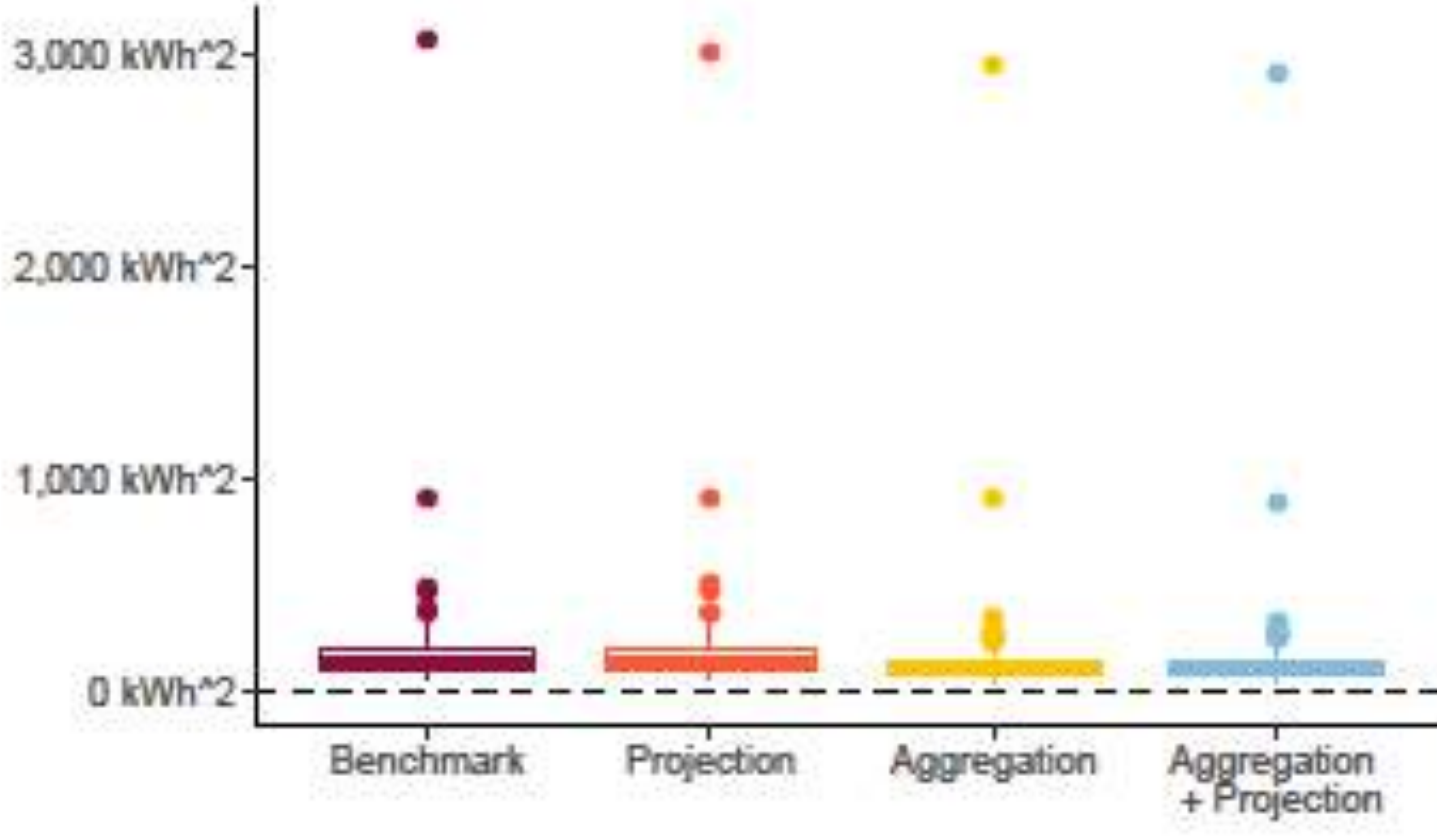
	Start date	End date
Behavioral segmentation		
Benchmark generation model training	April 20, 2009	April 19, 2010
Benchmark and observation standardization		
Initialization of the aggregation	April 20, 2010	April 30, 2010
Model evaluation	May 1, 2010	July 31, 2010

Results – Mean Squared Error (MSE) on test period

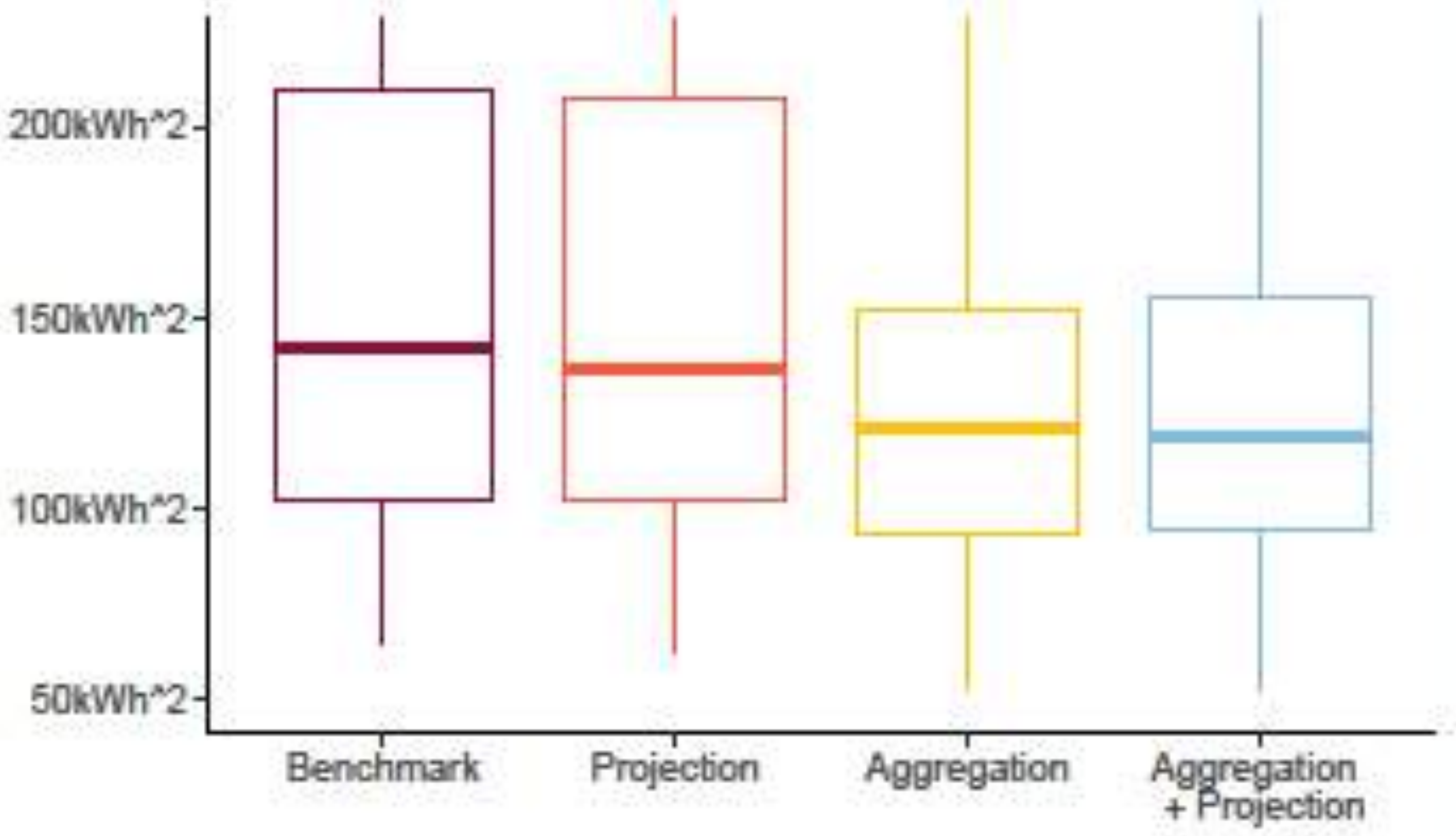
	All aggregated levels	Global	Local
Benchmark	455.5	205.8	66.3
Projection	450.7	200.8	66.3
Aggregation	397.9	172.0	61.2
Aggregation + Projection	**396.0**	**170.3**	**61.1**

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

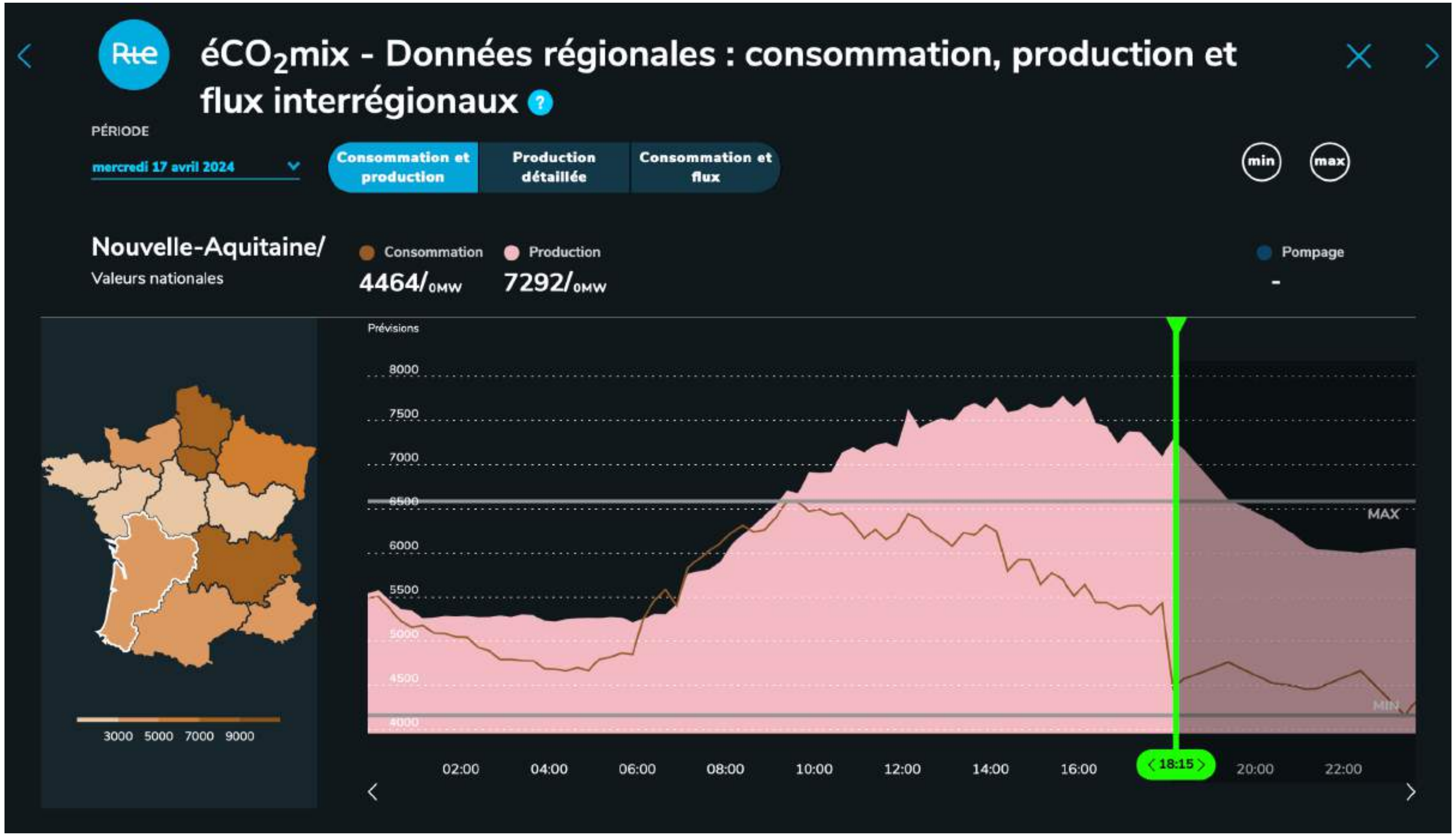


Box-plots trimmed at 220 kWh^2

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

France electricity consumption

The underlying real data set



Prospect: use cities consumption data

Online MinT

Input

- d: delay in data reception
- τ: window for the variance-covariance matrix of the base forecast errors estimation

For $t = 1,\ldots,T$

- For each level $\gamma \in \{$France, Auvergne-Rhône-Alpes, $\ldots$, Provence Alpes Côte d'Azur$\}$

  Generate online base forecast $\hat{y}_t = \left(\hat{y}_t^\gamma\right)_\gamma$

  Compute online empirical variance $\hat{\sigma}_t^\gamma = \dfrac{1}{\tau} \displaystyle\sum_{s=t-d-\tau}^{t-d} \left(e_s^\gamma - \bar{e}^\gamma\right)^2$ with $e_s^\gamma = y_s^\gamma - \hat{y}_s^\gamma$ and $\bar{e}^\gamma = \dfrac{1}{\tau} \displaystyle\sum_{s=t-d-\tau}^{t-d} e_s^\gamma$

- Reconcile base forecasts

  $\tilde{y}_t = SP_t\hat{y}_t$ with $P_t = \left(S^{\mathrm{T}}\widehat{\Sigma}_t^{\dagger}S\right)^{-1}S^{\mathrm{T}}\widehat{\Sigma}_t^{\dagger}$ with $\widehat{\Sigma}_t^{\dagger} = \mathrm{diag}\left(1/\sigma_t^\gamma\right)_\gamma$

Underlying assumption: base forecast errors of two different levels are independent ($\widehat{\Sigma}_t$ diagonal)

→ Perform faster and better

Results - 01.01.24 - 18.04.24

- Gam Kalman: generalized additive model + Kalman filter on model effects
- Gam Regions: Bottom up approaches based on 13 (one for each french region) generalized additive model + online linear regression on models effects
- Gam MinT: Online MinT on using Gam Kalman (for France) and the 13 models (of the regions) of the bottom up approach as base forecasts
- Best model: online aggregation of many models

Model	RMSE (MW)	MAPE (%)	Mean bias (MW)
Gam Kalman	1 381	1.75	-11
Gam Regions	1 499	2.07	-112
Gam MinT	1 288	1.79	-42
Best model	1 191	1.60	49
RTE D-2	1 549	2.19	409

Results

Performances prévision J+1
RMSE (MW)
Biais moyen (MW)
4000
3000
2000
1000
0
-2000
-4000
Modèle
R&D Gam Kalman
R&D Gam MiNT
R&D Gam Regions
2023-12-31
2024-01-01
2024-01-02
2024-01-03
2024-01-04
2024-01-05
2024-01-06
2024-01-07
2024-01-08
2024-01-09
2024-01-10
2024-01-11
2024-01-12
2024-01-13
2024-01-14
2024-01-15
2024-01-16
2024-01-17
2024-01-18
2024-01-19
2024-01-20
2024-01-21
2024-01-22
2024-01-23
2024-01-24
2024-01-25
2024-01-26
2024-01-27
2024-01-28
2024-01-29
2024-01-30
2024-01-31
2024-02-01
2024-02-02
2024-02-03

Alberta areas electricity consumption

joint work (in progress!) with Raffaele Mattera (Sapienza University of Rome)

Underlying real data set and methodology

- Alberta Electric System Operator (AESO)

→ Hourly power consumption data of the 42 Alberta areas from January 1, 2011 to October 31, 2023

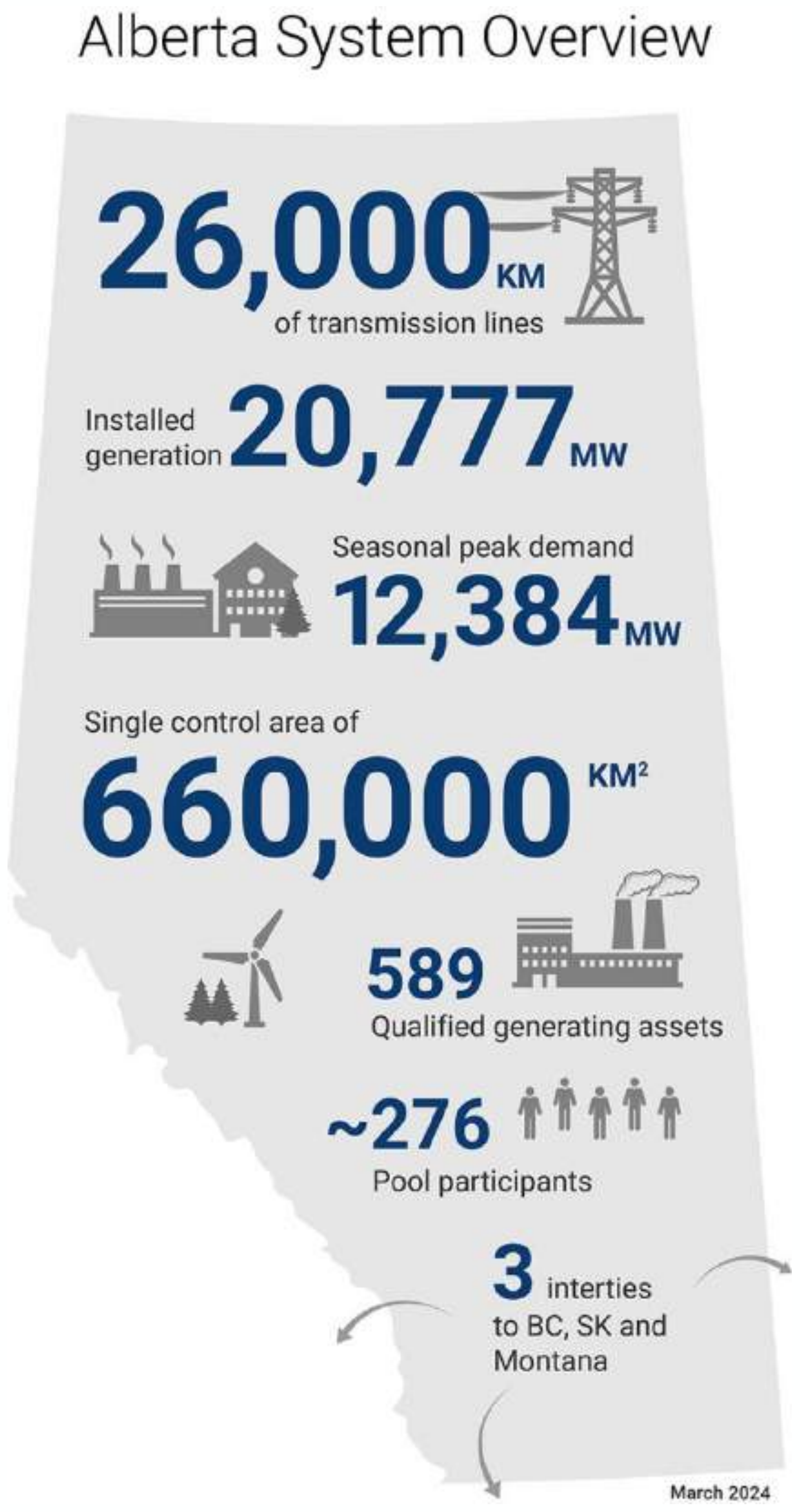
- National Oceanic and Atmospheric Administration (NOAA)

→ Hourly temperature in 27 weather stations

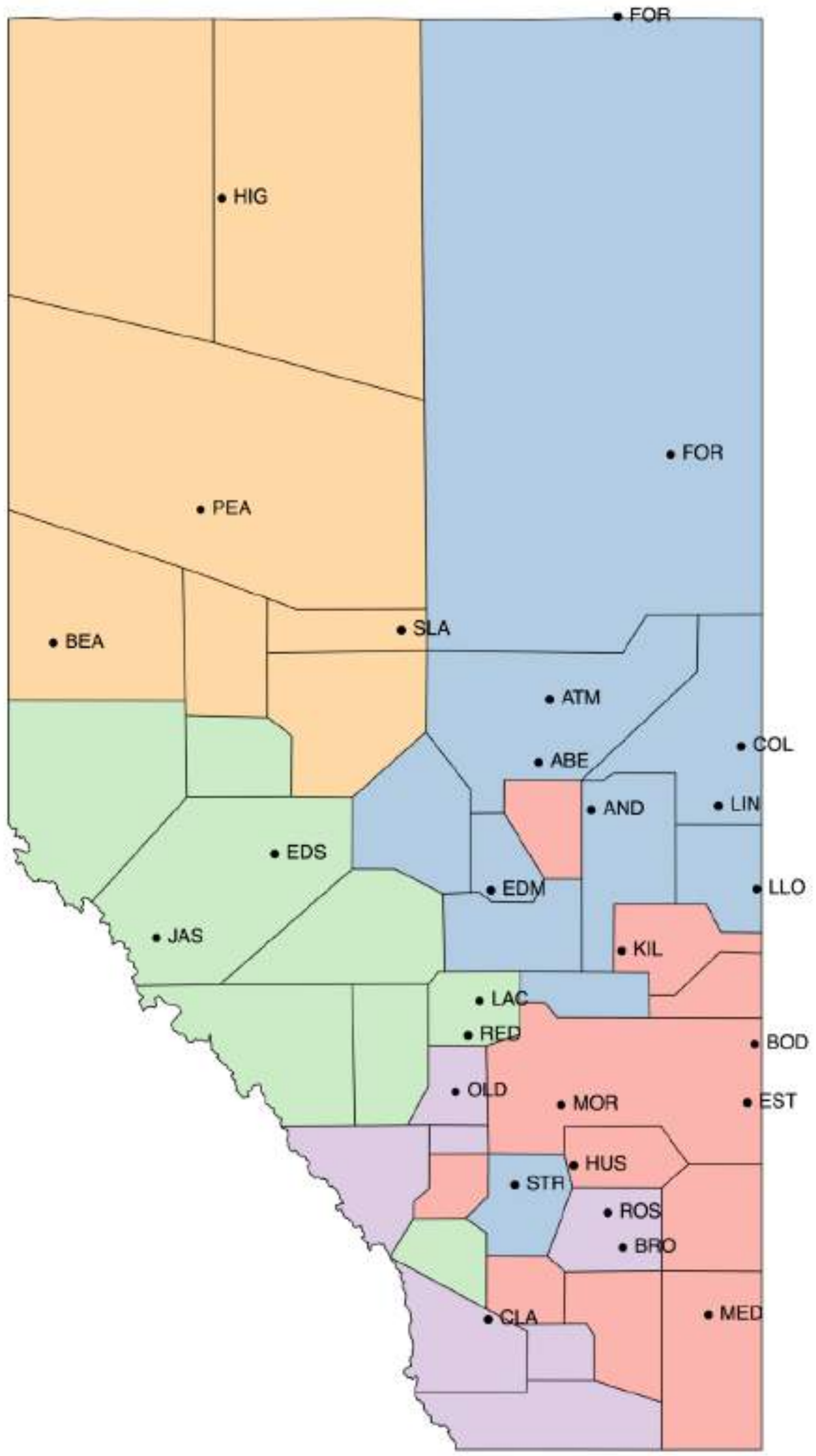
Objective: predict the power consumption of each area

1. Cluster areas and create a hierarchy
2. Forecast power at all levels
3. Reconcile forecasts (MinT)

Are area consumption reconciled forecasts better than original ones?



Results

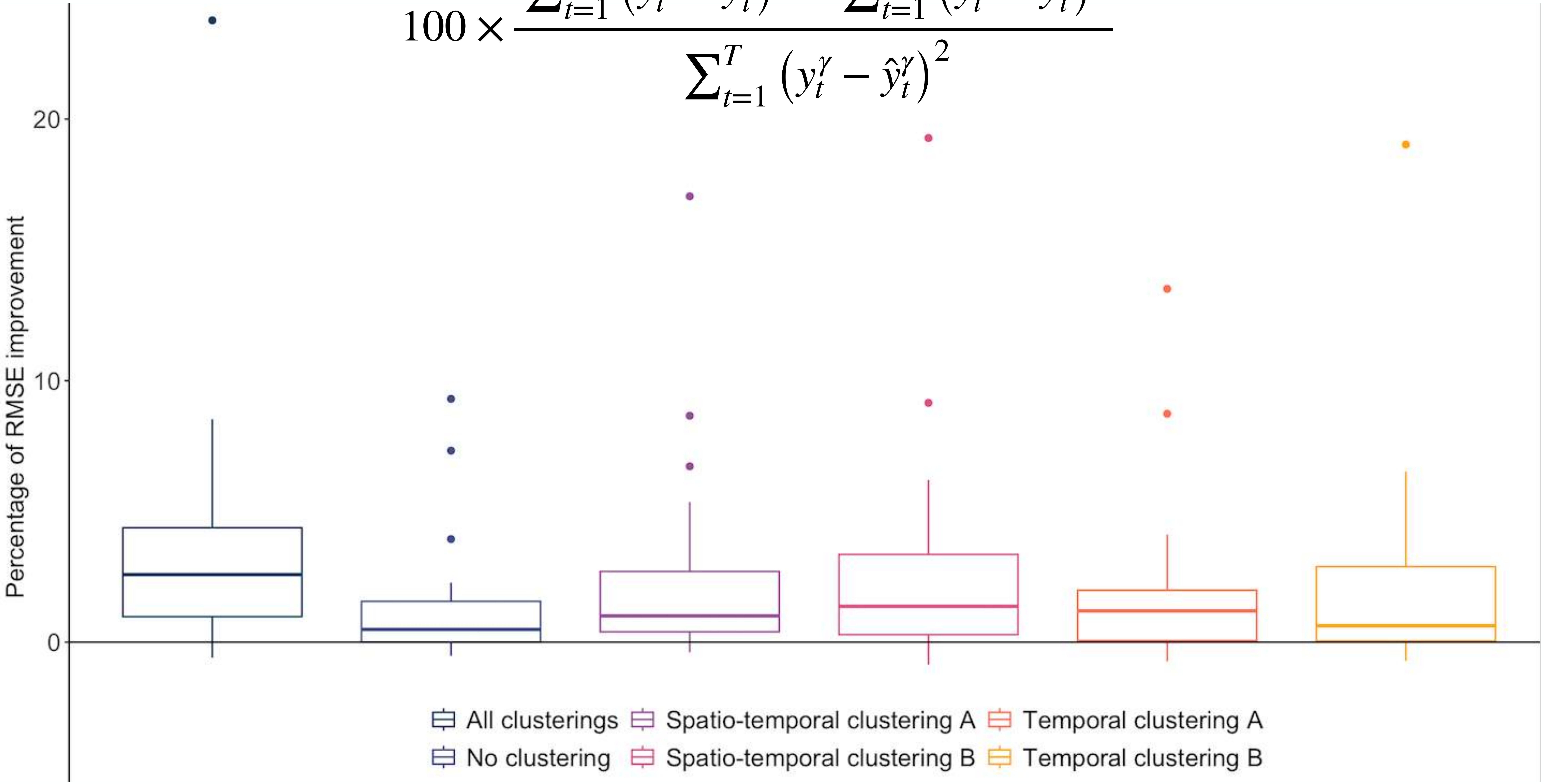


- (Space-)time clusterings
- Base forecasts using
  - Generalized additive models with and without lags
  - Persistent models (linear regression on lags)
  - Random forests
- Online and Offline reconciliation with

$\widehat{\Sigma}$ = empirical variance/covariance matrix

Results

$$100 \times \frac{\sum_{t=1}^{T} \left(y_t^\gamma - \hat{y}_t^\gamma\right)^2 - \sum_{t=1}^{T} \left(y_t^\gamma - \tilde{y}_t^\gamma\right)^2}{\sum_{t=1}^{T} \left(y_t^\gamma - \hat{y}_t^\gamma\right)^2}$$



That's all folks!