Adaptive Methods for Short-Term Electricity Load Forecasting During COVID-19 Lockdowns in France

from GAM to aggregation of experts, are we still interpretable?

Yannig Goude[1,2], joint work with David Obst[1,3], Joseph de Vilmarest[1,4] , Olivier Wintenberger[4] , Mehdi Raqui[1]

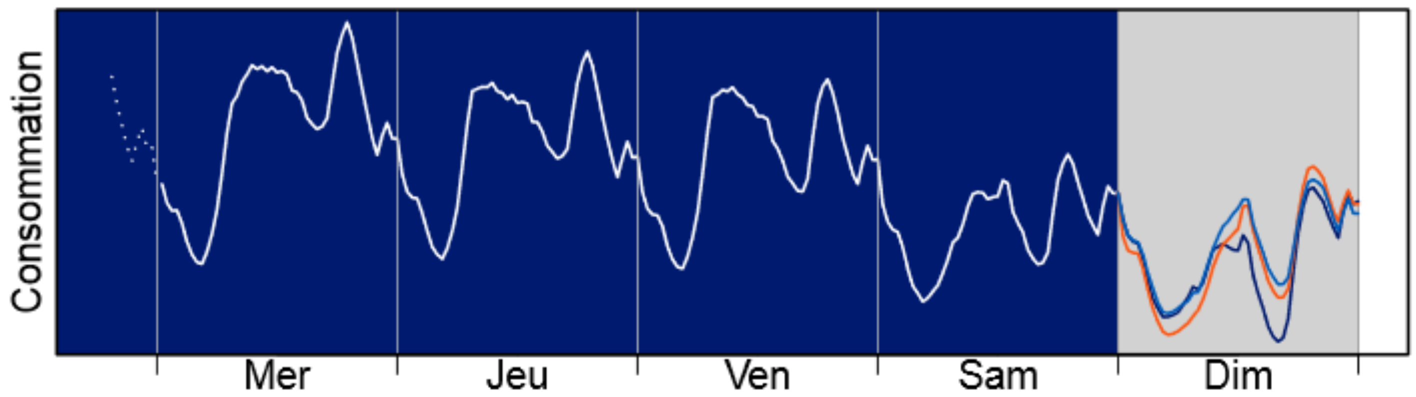
[1]EDF R&D, OSIRIS department, EDF Lab 7 bd Gaspard Monge, Palaiseau, France.
[2]Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay
[3]Institut de Mathématiques de Marseille, Université Aix-Marseille
[4]Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université

Enbis Workshop *Interpretability for Industry 4.0, Pillar 2: Interpretability via additive models*
12-13th July 2021

eDF

Load forecasting is crucial for electrical power system operations

- Generation: optimising production planning
- Trading: buy and sell electricity on the markets
- Grid management: transmission, distribution



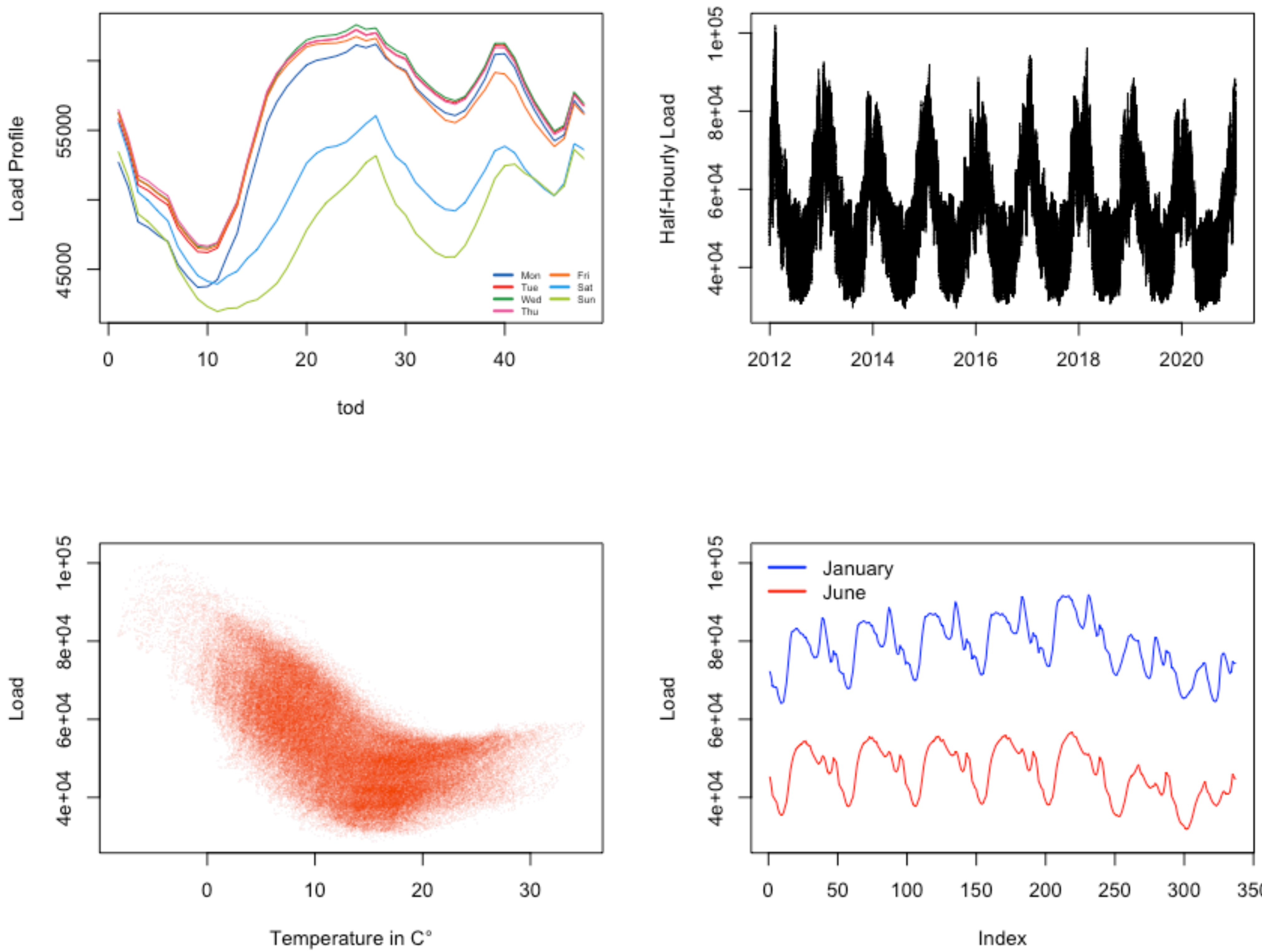
COVID-19 pandemic is impacting our social and economical life

Electricity production and consumption are of course affected

We will present and discuss:

- **Problems**: how does it impact electricity load in the world, in France in particular?
- **Model design**: how the forecasting model could be adapted to maintain good forecasting performances during (and after…) that period, what we did at EDF and other related works? Adaptativity vs interpretability?
- **Data**: what kind of data could be used to improve forecasts?

Electricity Data

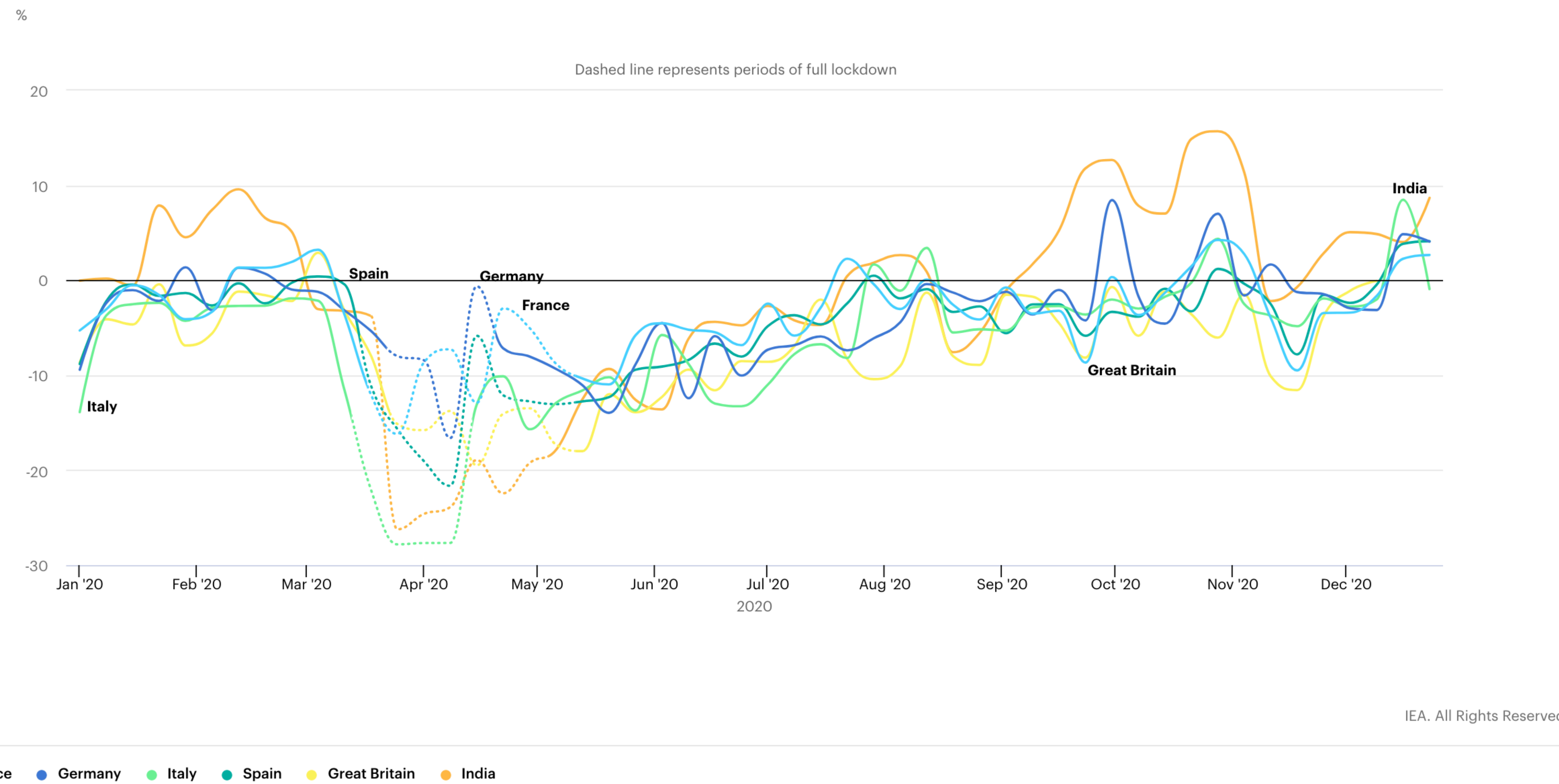

Load Profile
tod
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Half-Hourly Load
Load
Temperature in C°
January
June
Index

Impact on electricity consumption in the world?

Electricity demand dropped quickly with confinement measures.

It steadily recovered as measures were gradually softened; it was still 10% below 2019 levels in EU countries in June.

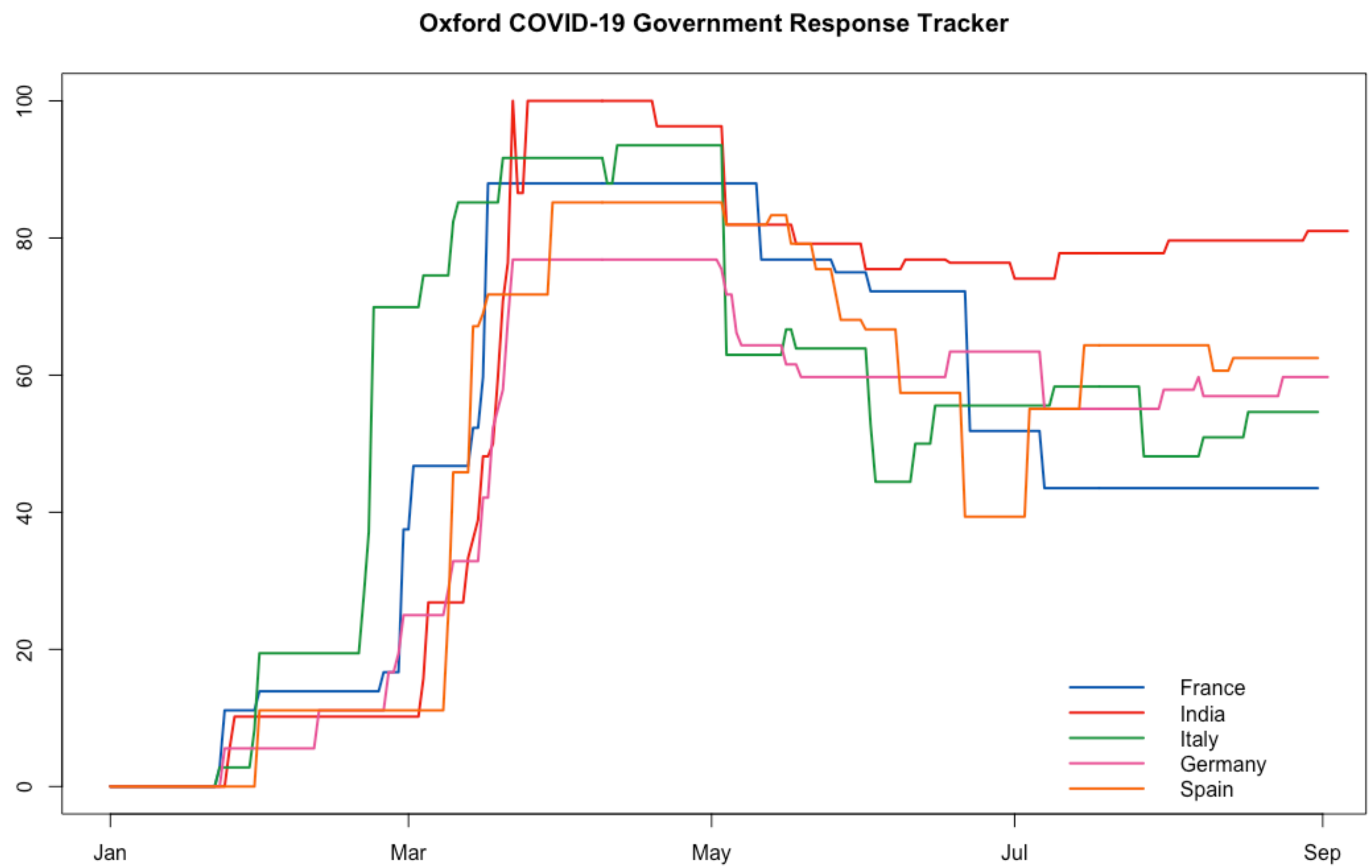
In the last week of July, electricity demand was 5% below 2019 levels in EU countries except Italy. In India, recovery seems faster.



IEA, Year-on-year change in weekly electricity demand, weather corrected, in selected countries, January-December 2020, IEA, Paris https://www.iea.org/data-and-statistics/charts/year-on-year-change-in-weekly-electricity-demand-weather-corrected-in-selected-countries-january-december-2020

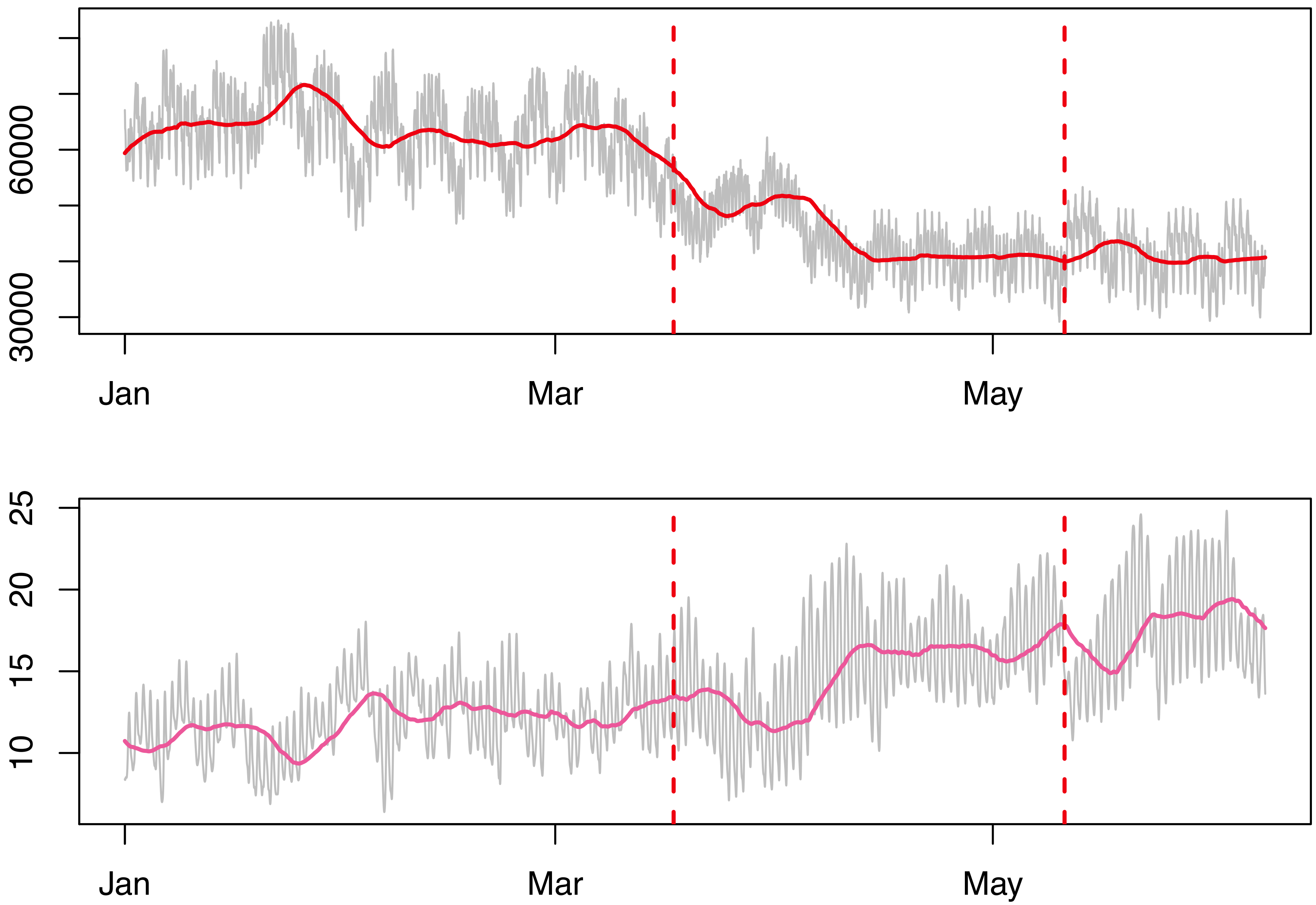
Government responses of different intensities

	First lockdown measures (day 0)	Lockdown strengthened	Lockdown softened
France	March 14	March 17 (day 3)	May 11 (day 55)
Germany	March 15	March 22 (day 7)	April 20 (day 36) and May 4 (day 50)
Italy	March 4	March 13 (day 9)	April 14 (day 41) and May 4 (day 61)
Spain	March 9	March 15 (day 6)	May 11 (day 55)
UK	March 19	March 23 (day 4)	May 11 in England (day 55)
India	March 18	March 25 (day 7)	May 4 (day 47)



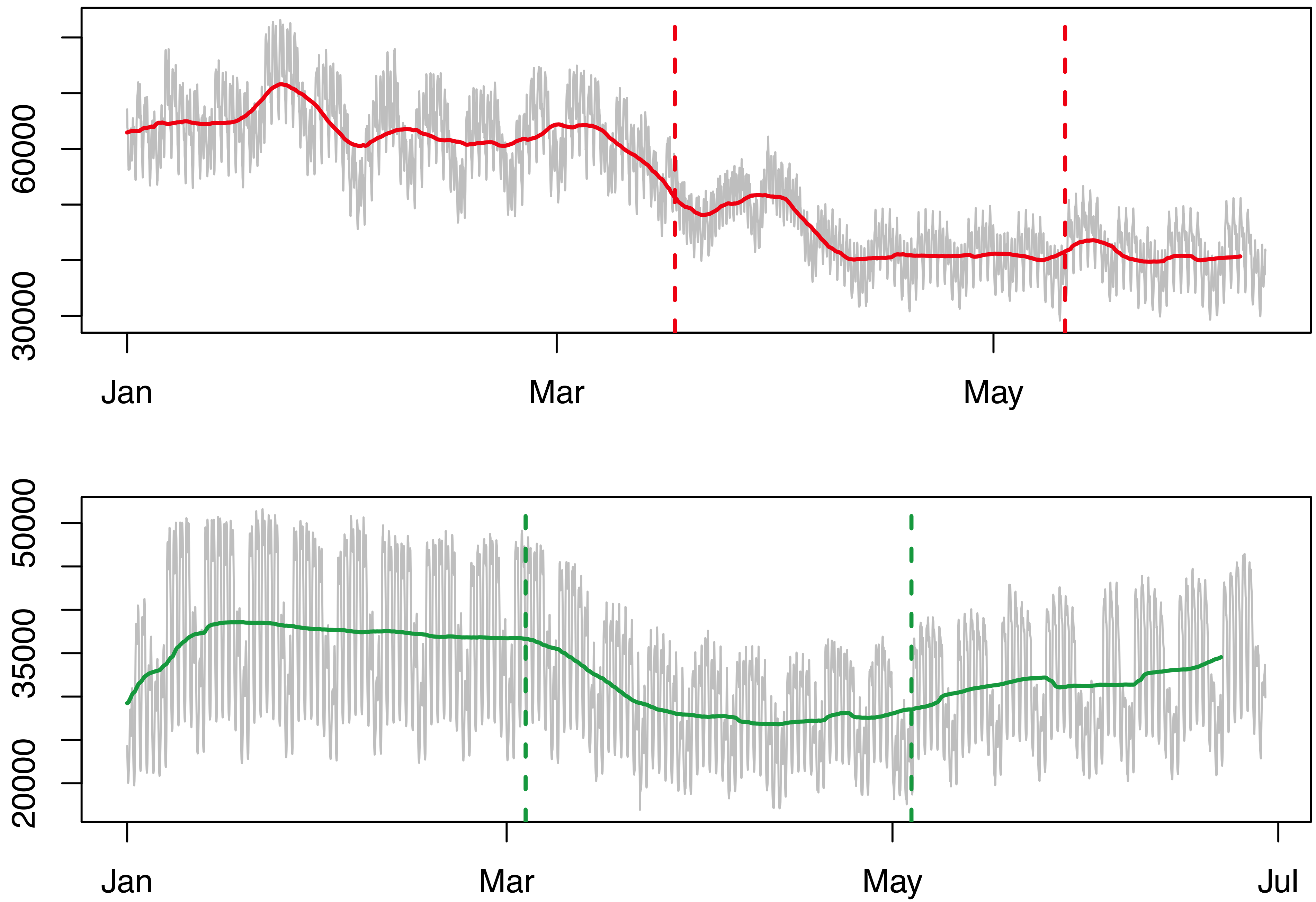
https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker

How does it impact electricity load in France?

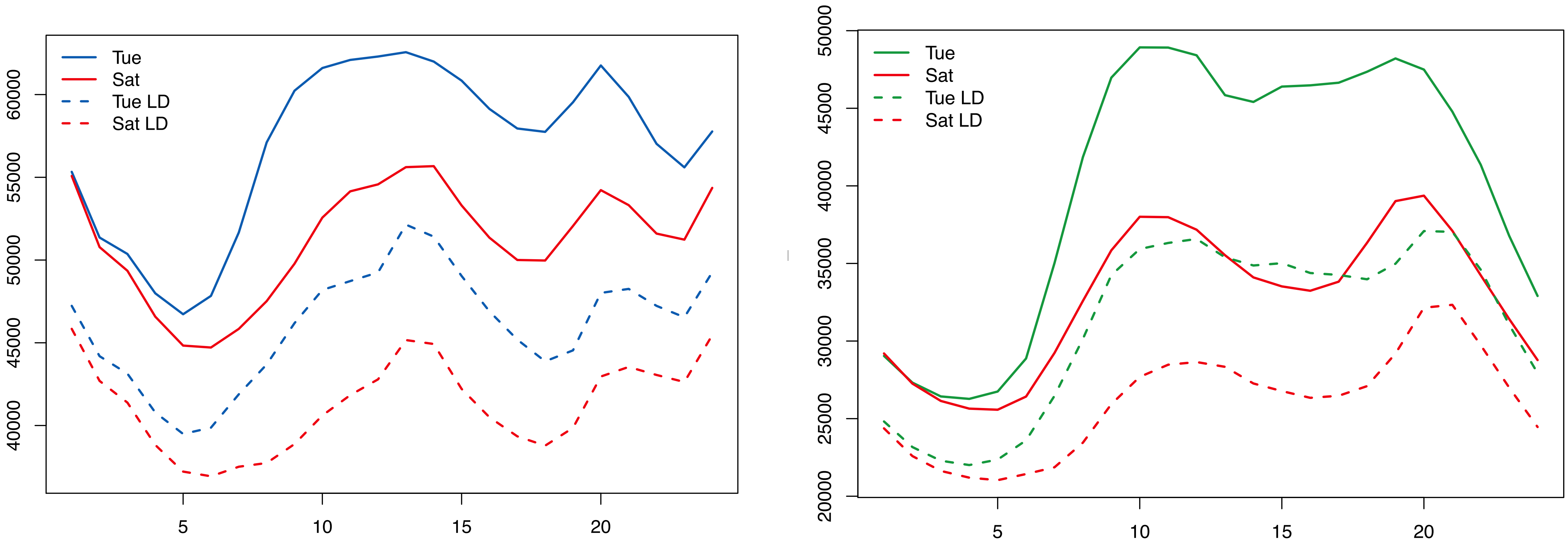


Half-Hourly french electricity load consumption (top) and temperature.

Comparison with Italy

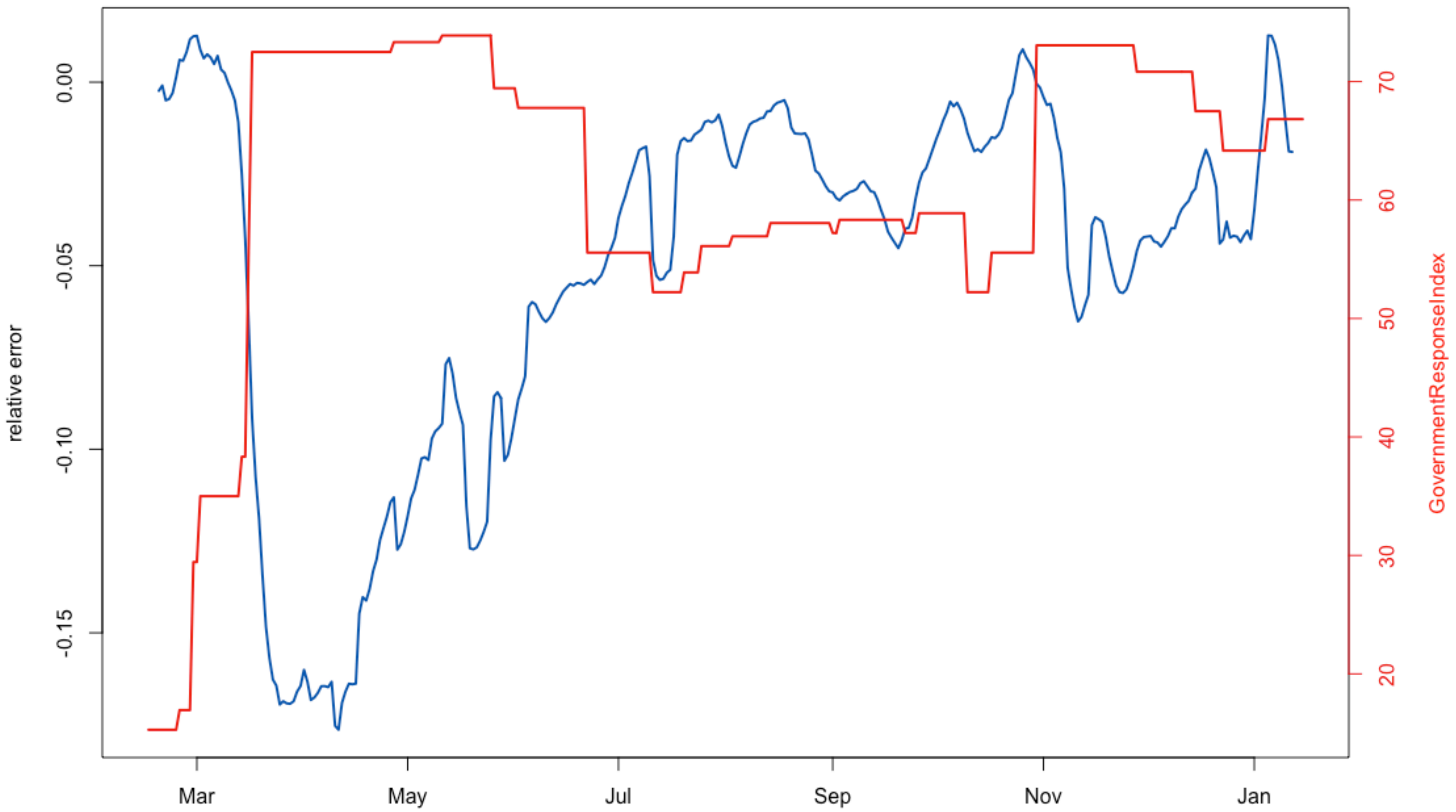


French and Italian electricity load (in MW) at resp. half-hourly and hourly resolution in 2020. Dashed lines are the starting and ending date ofthe lockdown



French and Italian electricity Tuesday and Saturday load profiles before and during the lockdown (Dashed lines)

Government responses of different intensities evolving with time:

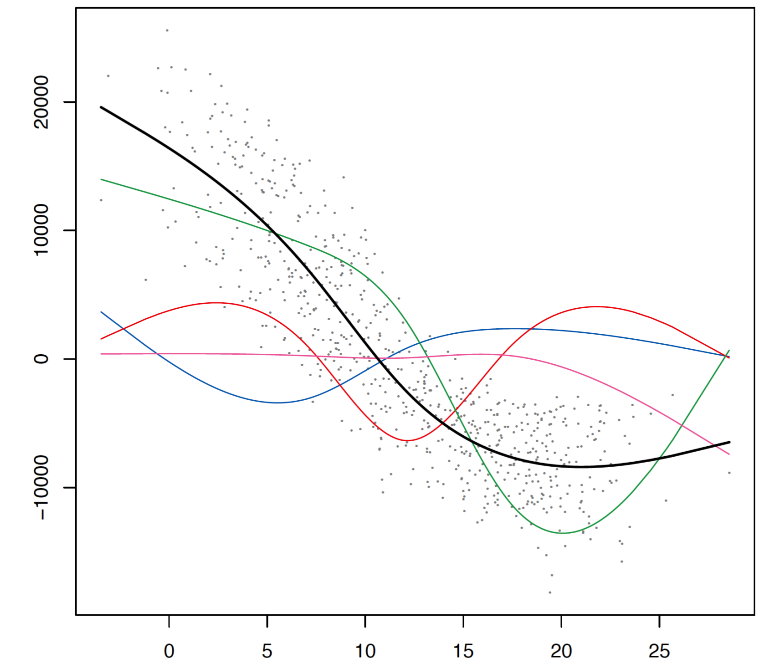


Algorithms and models

Our forecasting model

We model the electricity consumption with GAM, a sum of linear and smooth additive effects (see Hastie & Tibshirani (1990) and Wood (2017))

$$
\begin{aligned}
y_t = & \sum_{i=1}^{7} \sum_{j=0}^{1} \alpha_{i,j} \mathbb{1}_{\text{DayType}_t = i} \mathbb{1}_{\text{DLS}_t = j} \\
& + \sum_{i=1}^{7} \beta_i \text{Load1D}_t \mathbb{1}_{\text{DayType}_t = i} + \gamma \text{Load1W}_t \\
& + f_1(t) + f_2(\text{ToY}_t) + f_3(t, \text{Temp}_t) + f_4(\text{Temp95}_t) \\
& + f_5(\text{Temp99}_t) + f_6(\text{TempMin99}_t, \text{TempMax99}_t) + \varepsilon_t \,,
\end{aligned}
$$



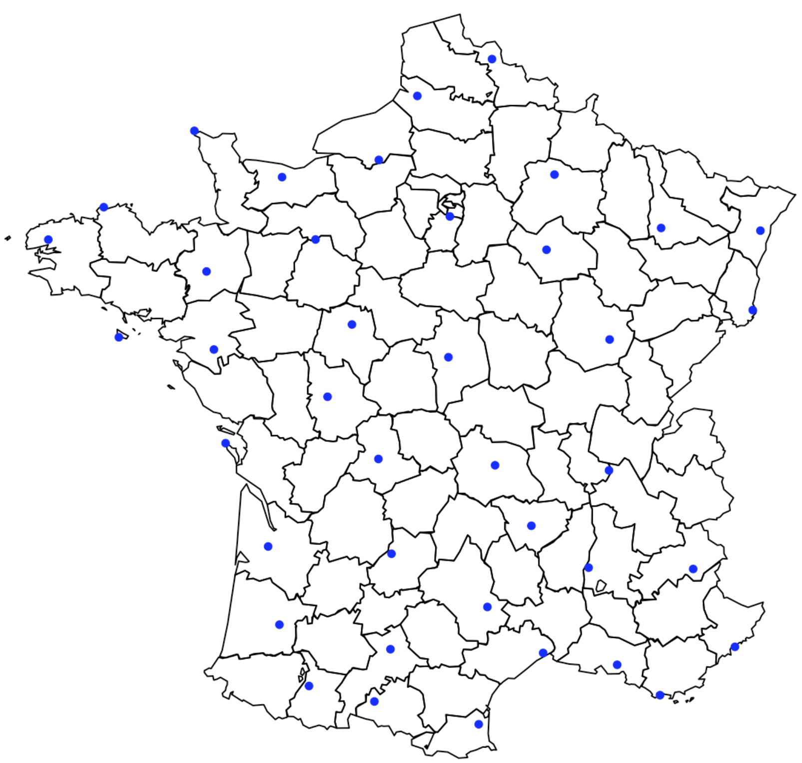
Each effect is obtained by penalised spline regression, minimising a GCV criteria to calibrate the amount of smoothness:

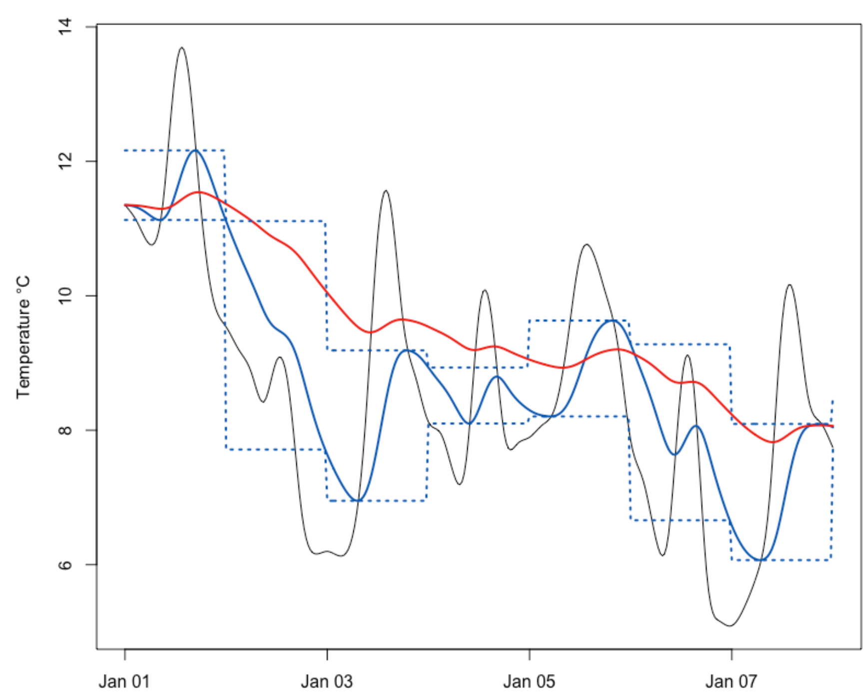
$$
\sum_{i=1}^{n} (y_i - \beta_0 \mathbf{X}_i^0 - \sum_{q=1}^{p} f_q(x_i))^2 + \sum_{q=1}^{p} \lambda_q \int |||f_q^{''}(x)|||^2 dx
$$

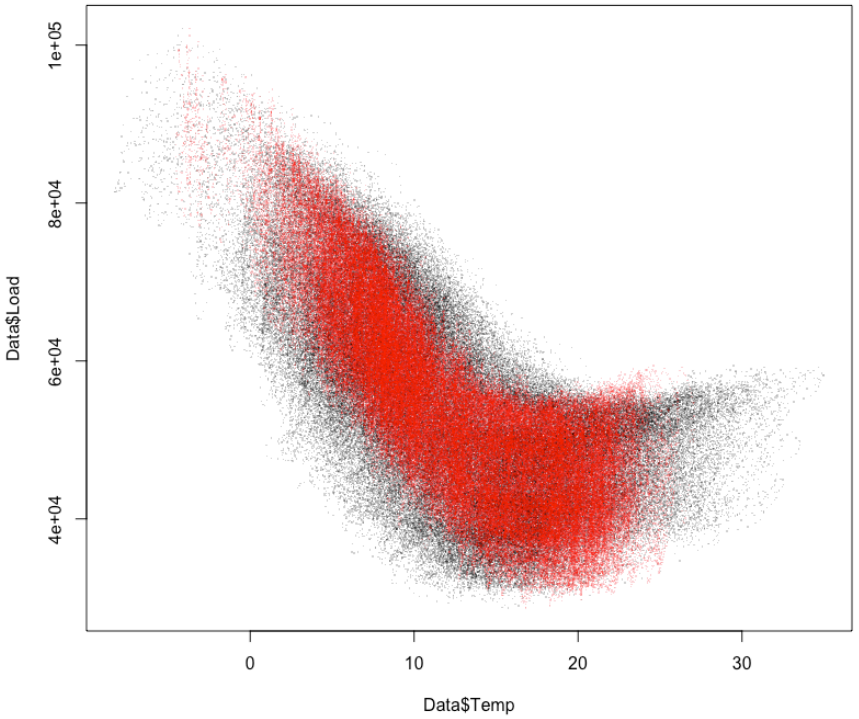
$$
f_j(x_j) = \sum_{i=1}^{k} \beta_{ji} b_{ji}(x_j).
$$

Covariate description:

- y_t is the electricity load for day t.
- $DayType_t$ is the electricity load for day t.
- DLS_t is the electricity load for day t.
- $Load1D_t$ is the electricity load for day t.
- ToY_t is the time of year whose value grows linearly from 0 on the *1^{st} of January 0h00* to 1 on the *31^{st} of December 23h30*.
- $Temp_t$ is the national average temperature.
- $Temp95_t$ and $Temp99_t$ are exponentially smoothed temperatures of factor 0.95 and 0.99.
- $TempMin95_t$ and $TempMax95_t$ are daily min and max of the smoothed temperatures



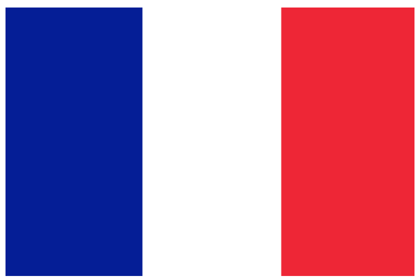




Our forecasting models

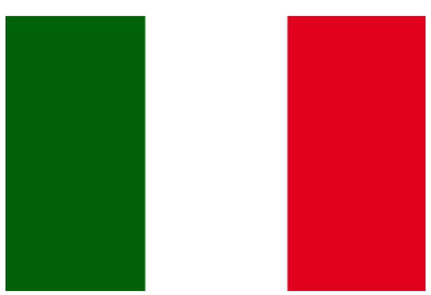
French model

Implemented in R (mgcv package)



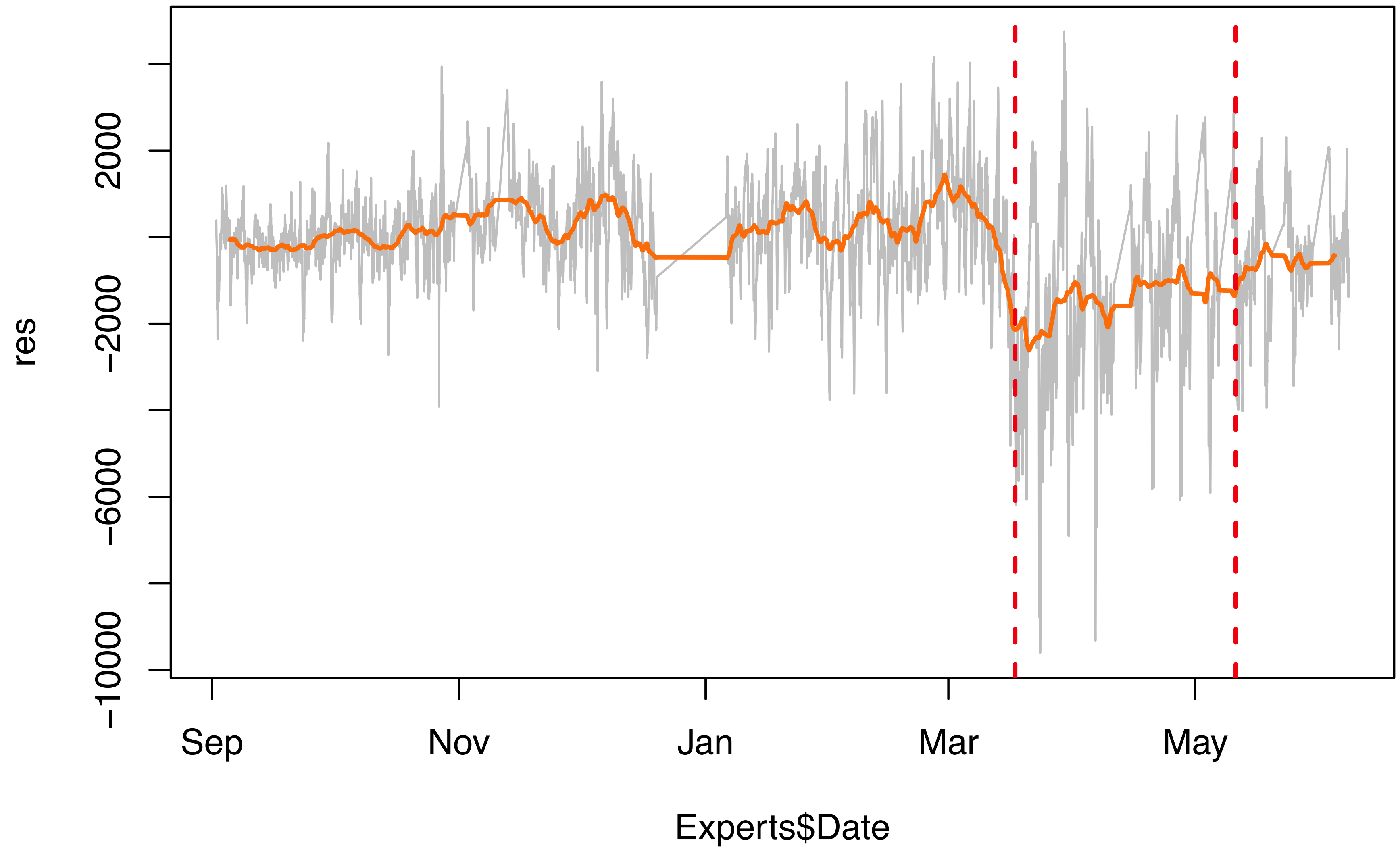
```
gam_france <- Load ~ s(DateN, k=3) + DayType:DLS + s(ToY, k = 20, bs = 'cc') +
s(DateN,Temp, k=c(3,5)) + s(Temp_s95, k=5) + s(Temp_s99, k=5) + s(Temp_s99_min,
Temp_s99_max) + Load.48:DayType +Load.336
```

Italian model



```
gam_france <- Load ~ s(DateN, k=3) + DayType:DLS  + s(ToY, k = 20, bs = ‘cc', by=DayType)
+ s(Temp_s95,k=5) +s(Temp_s99,k=5) + Load.24:DayType +Load.168
```

Of course achieve bad performances after the lockdown



MAPE increases from
1.4% to 3.8%

Online update

- We model the electricity consumption as a sum of time varying additive effects:

$$\mathbb{E}[y_t] = f_t(\boldsymbol{x}_t)$$

- For stability reasons and a good reactivity to changes we restricted to this special case:

$$\mathbb{E}[y_t] = \boldsymbol{\theta}_t^\top f(\boldsymbol{x}_t) \qquad f(\boldsymbol{x}_t) = (1, \overline{f}_1(x_{t,1}), ..., \overline{f}_d(x_{t,d}))^\top$$

- And the time varying coefficients are estimated solving an iterative least square problem with a forgetting factor

$$\hat{\boldsymbol{\theta}}_t = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \sum_{s=1}^{t-1} e^{-\mu(t-s)} \left(y_s - \boldsymbol{\theta}^\top f(\boldsymbol{x}_s)\right)^2$$

Ba, A., Sinn, M., Goude, Y., & Pompey, P. (2012).

Online update

$$y_t = \boldsymbol{\theta}_t^\top f(\boldsymbol{x}_t) + \varepsilon_t \,,$$
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{\eta}_t \,,$$

(ε_t) and $(\boldsymbol{\eta}_t)$ are gaussian white noises

variance / covariance σ^2 and Q

Algorithm 1: Kalman Filter

Initialization: the prior $\boldsymbol{\theta}_1 \sim \mathcal{N}(\hat{\boldsymbol{\theta}}_1, P_1)$ where $P_1 \in \mathbb{R}^{d \times d}$ is positive definite and $\hat{\boldsymbol{\theta}}_1 \in \mathbb{R}^d$.

Recursion: at each time step $t = 1, 2, \ldots$

1) Prediction:

$$\mathbb{E}\left[y_t \mid (\boldsymbol{x}_s, y_s)_{s<t}, \boldsymbol{x}_t\right] = \hat{\boldsymbol{\theta}}_t^\top f(\boldsymbol{x}_t) \,,$$
$$Var\left[y_t \mid (\boldsymbol{x}_s, y_s)_{s<t}, \boldsymbol{x}_t\right] = \sigma^2 + f(\boldsymbol{x}_t)^\top P_t f(\boldsymbol{x}_t) \,.$$

2) Estimation:

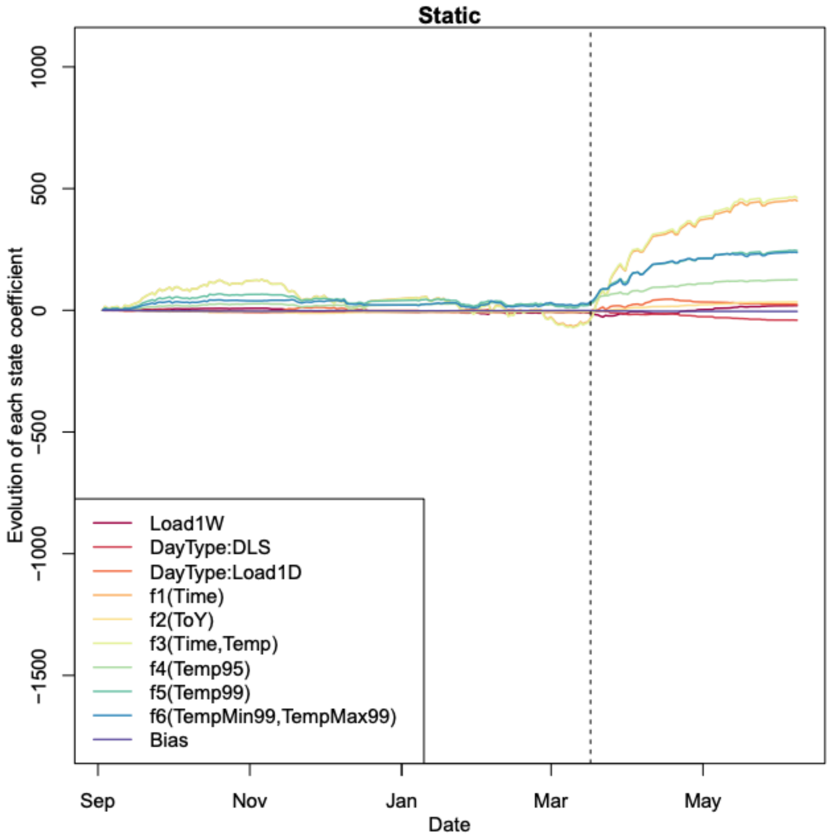
$$\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t + \frac{P_t f(\boldsymbol{x}_t)}{f(\boldsymbol{x}_t)^\top P_t f(\boldsymbol{x}_t) + \sigma^2}\left(y_t - \hat{\boldsymbol{\theta}}_t^\top f(\boldsymbol{x}_t)\right) \,,$$
$$P_{t+1} = P_t - \frac{P_t f(\boldsymbol{x}_t) f(\boldsymbol{x}_t)^\top P_t}{f(\boldsymbol{x}_t)^\top P_t f(\boldsymbol{x}_t) + \sigma^2} + Q \,.$$

Q diagonal

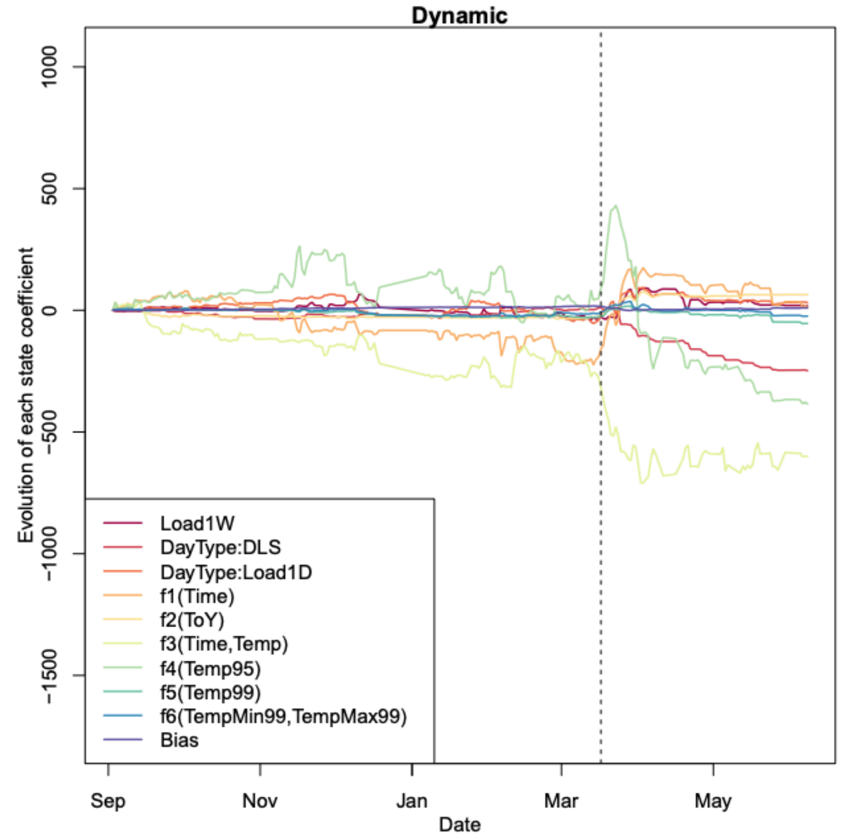
- set to 0: ***Kalman Static*** $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t$ $\quad \hat{\boldsymbol{\theta}}_t = \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^d} \left(\sum_{s=1}^{t-1} (y_s - \boldsymbol{\theta}^\top f(\boldsymbol{x}_s))^2 + \|\boldsymbol{\theta}\|^2 \right)$
- estimated using a greedy algorithm: ***Kalman Dynamic***
- increasing Q at the beginning of the lockdown ***Kalman Dynamic Break***

De Vilmarest, J., & Wintenberger, O. (2020)

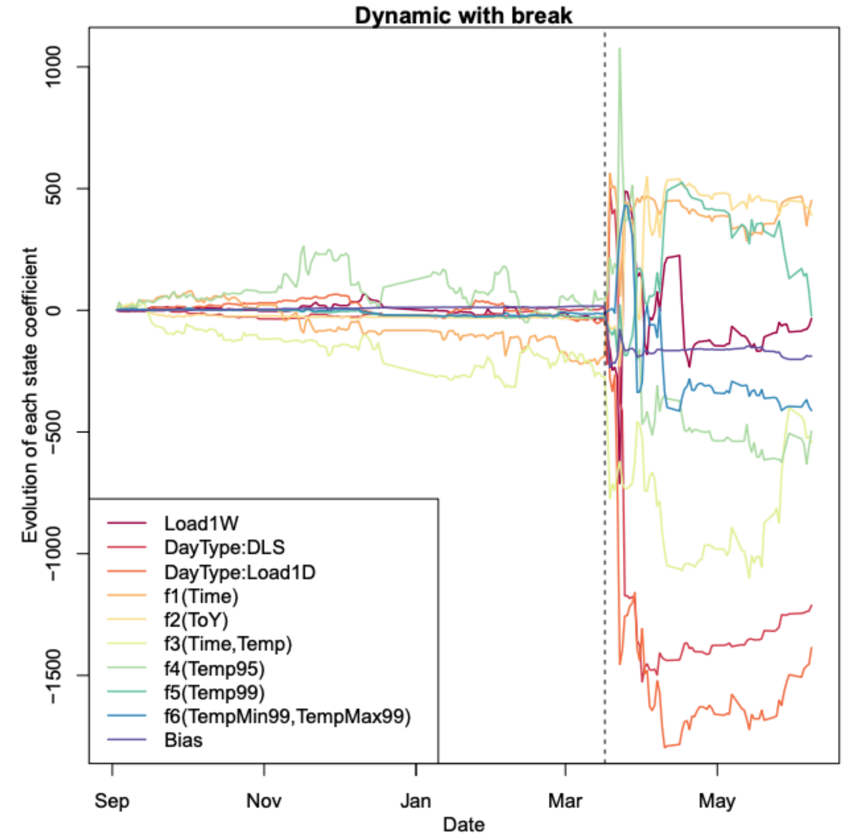
Evolution of state coefficients in function of time for the different Kalman configuration



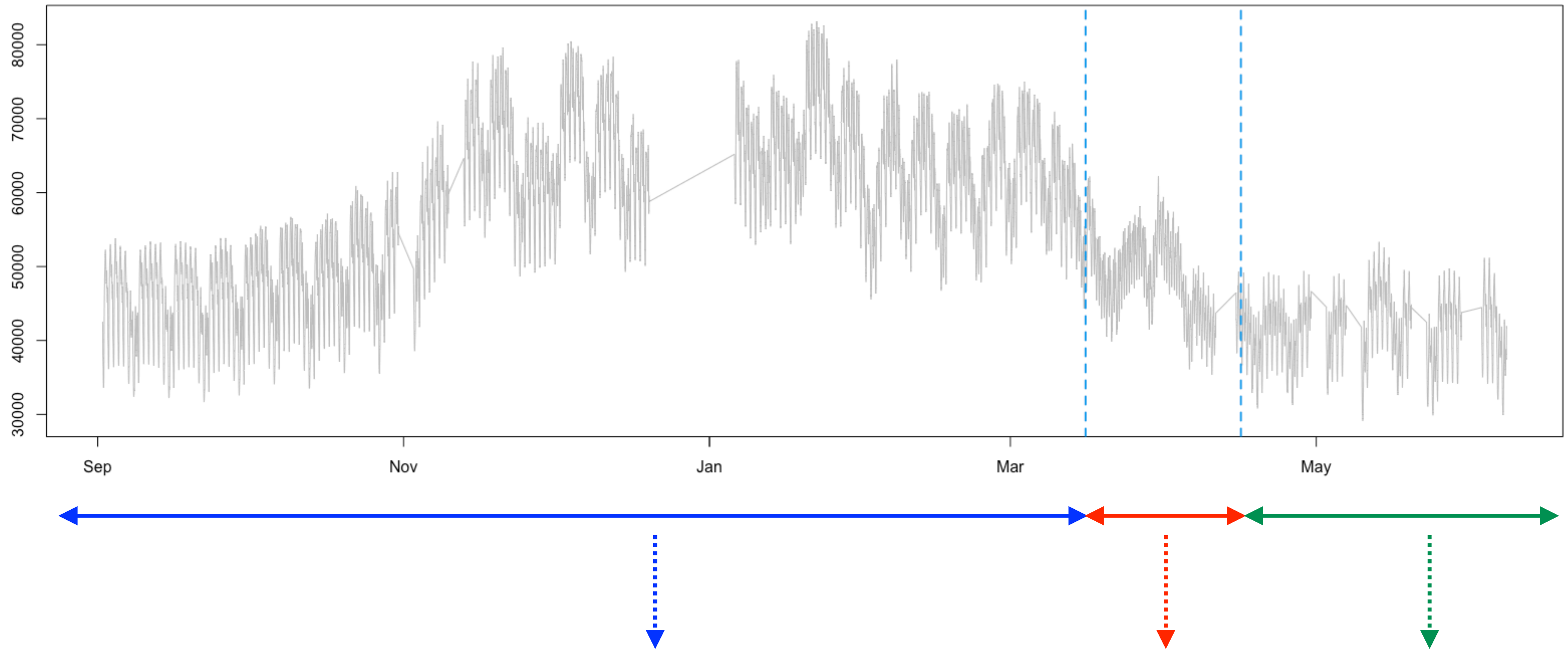
(a) Static



(b) Dynamic



(c) Dynamic with break



Method	2019/09/01 - 2020/03/15	2020/03/16 - 2020/04/15	2020/04/16 - 2020/06/07
ARIMA	4.10 %, 3341 MW	5.44 %, 3248 MW	5.59 %, 3135 MW
GAM	1.39 %, 1085 MW	4.83 %, 2961 MW	3.12 %, 1753 MW
GAM + ARIMA	1.34 %, 1050 MW	4.28 %, 2654 MW	2.65 %, 1464 MW
exp-LS	1.26 %, 982 MW	3.94 %, 2521 MW	1.97 %, 1029 MW
Kalman Static	1.38 %, 1077 MW	4.81 %, 2923 MW	2.85 %, 1588 MW
Kalman Static break	-	2.79 %, 1954 MW	1.59 %, 855 MW
Kalman Dynamic	**1.26 %, 979 MW**	3.66 %, 2351 MW	1.89 %, 1002 MW
Kalman Dynamic break	-	**2.73 %, 1902 MW**	**1.62 %, 854 MW**

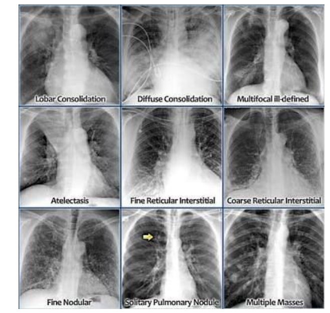
▶ D. Obst, J. De Vilmarest and Y. Goude (2021)

Transfer Learning

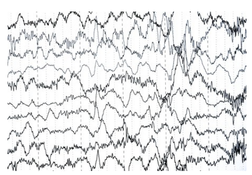
Transfer learning (or learning-to-learn, knowledge transfer, multi-task learning) is a branch of machine learning that aims at reusing knowledge from one source task (usually with a lot of data) on another target one (with few data).

Source

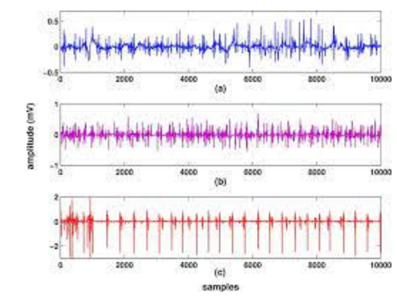
ImageNet



Chestx-ray8



EEG



EMG

Book corpus data

Label	Sentence	Source
*	The more books I ask to whom he will give, the more he reads.	Culicover and Jackendoff (1999)
✓	I said that my father, he was tight as a hoot-owl.	Ross (1967)
✓	The jeweller inscribed the ring with the name.	Levin (1993)
*	many evidence was provided.	Kim and Sells (2008)
✓	They can sing.	Kim and Sells (2008)
✓	The men would have been all working.	Baltin (1982)
*	Who do you think that will question Seamus first?	Carnie (2013)
*	Usually, any lion is majestic.	Dayal (1998)
✓	The gardener planted roses in the garden.	Miller (2002)
✓	I wrote Blair a letter, but I tore it up before I sent it.	Rappaport Hovav and Levin (2008)

Labeled data

Target

- Pan, S. J., & Yang, Q. (2009)
- Bird, J. J., Kobylarz, J., Faria, D. R., Ekárt, A., & Ribeiro, E. P. (2020)
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018)
- Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019).

Fine-tuning of GAM

$$y_t = \beta_0 + \sum_{j=1}^{d} f_j(x_{t,j}) + \varepsilon_t$$

$$f_j(x) = \sum_{k=1}^{m_j} \beta_{j,k} B_{j,k}(x)$$

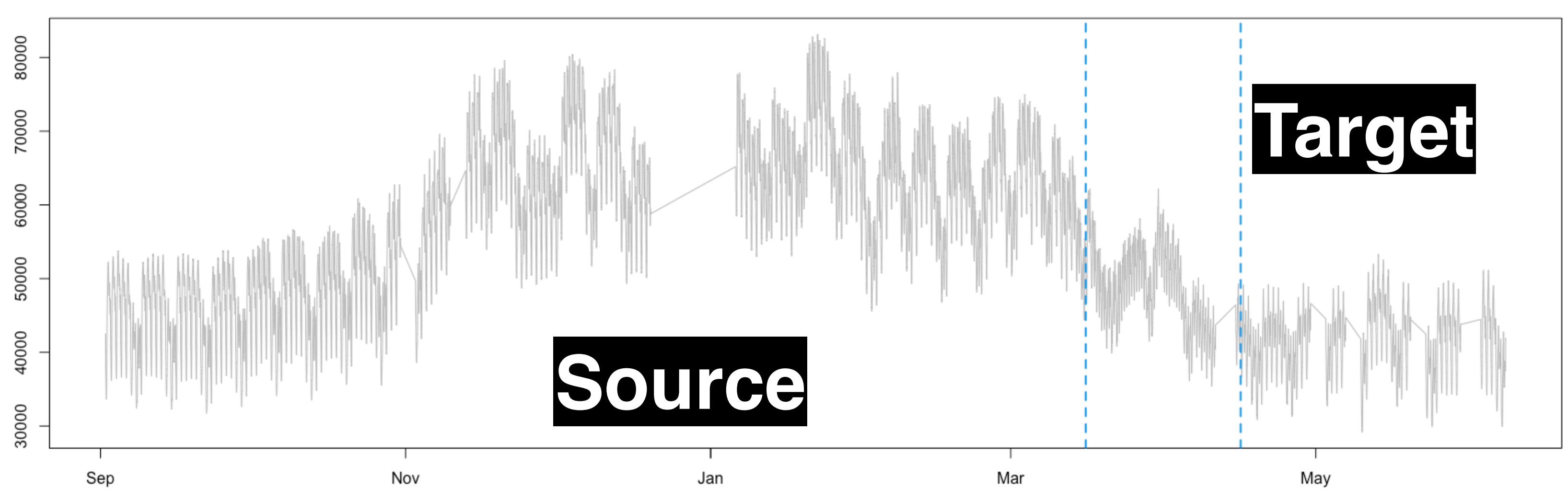
$$\mathcal{L}_t(\boldsymbol{\beta}) = \sum_{s=1}^{t-1} \left(y_s - \sum_{j=1}^{d} \sum_{k=1}^{m_j} \beta_{j,k} B_{j,k}(x_{s,j}) \right)^2$$

Algorithm 2: Transfer learning at time step t: GAM fine-tuned

Inputs: Step size α, number of iterations K, French source parameters $\hat{\beta}_S^{FR}$

- Initialize $\hat{\beta}_t \leftarrow \hat{\beta}_S^{FR}$.
- Repeat K times:
  $\hat{\beta}_t \leftarrow \hat{\beta}_t - \alpha \nabla \mathcal{L}_{t-1}^{FR}(\hat{\beta}_t)$.
- Predict $\hat{y}_t = \hat{\beta}_t^\top B(x_t)$.

for each time step we perform K iterations of batch gradient descent with fixed step size



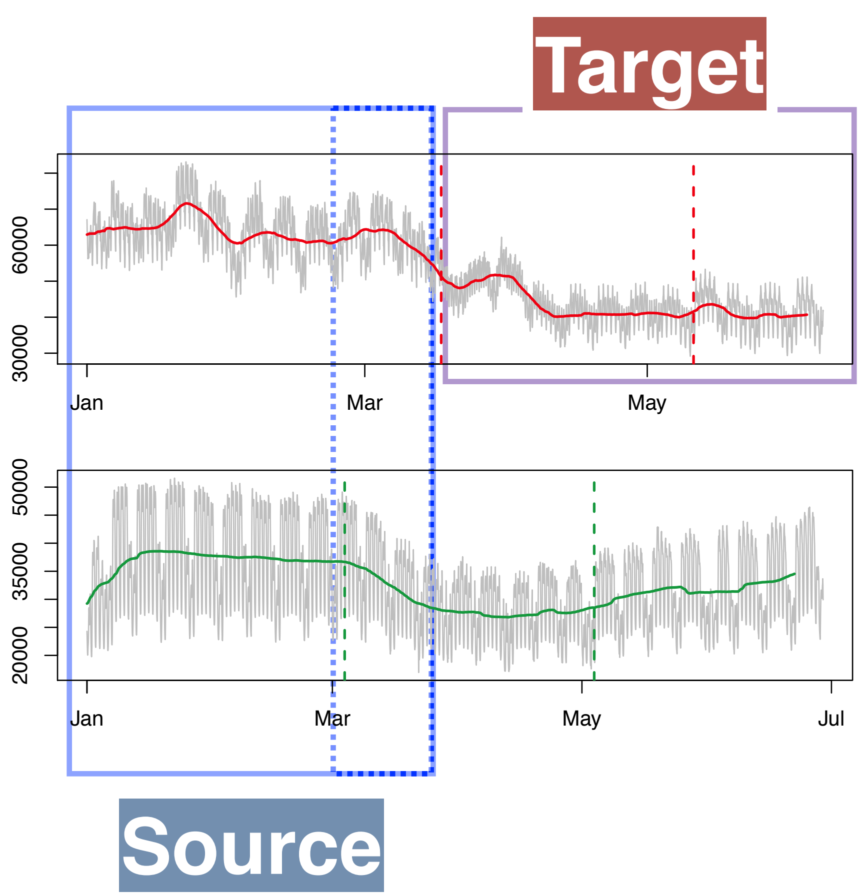
Transfer Learning from Italian data

Italy was the first country to be massively affected by the COVID 19 in Europe.
The Italian government decreed a total lockdown 7 days before the French one.
The idea is to use this one week head-start to adjust our GAM model for France accordingly to the changes observed in Italy.

We fit a similar GAM on Italian Data, we suppose the variation before/during the lockdown is similar in Italy and France:

Let $\hat{\boldsymbol{\delta}}_t$ be the adjustment done on the Italian GAM coefficients when fine-tuned version on the beginning of march (before the 15th) and ρ a scaling factor between France and Italian data.

$$\tilde{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}_S^{FR} + \rho\, \hat{\boldsymbol{\delta}}_t \qquad \hat{\rho} = \sum_t y_t^{FR} / \sum_t y_t^{IT}$$



Algorithm 3: Transfer learning at time step t: GAM-δ

- Initialize $\hat{\beta}_t^{IT} \leftarrow \hat{\beta}_S^{IT}$.
- Repeat K times:
  $\hat{\beta}_t^{IT} \leftarrow \hat{\beta}_t^{IT} - \alpha \nabla \mathcal{L}_{t-1}^{IT}(\hat{\beta}_t^{IT})$.
- Set $\hat{\delta}_t = \hat{\beta}_t^{IT} - \hat{\beta}_S^{IT}$, $\tilde{\beta}_t = \hat{\beta}_S^{FR} + \rho\, \hat{\delta}_t$.
- Predict $\hat{y}_t = \tilde{\beta}_t^{\top} B(x_t)$.

Transfer Learning from Italian data+fine-tuning

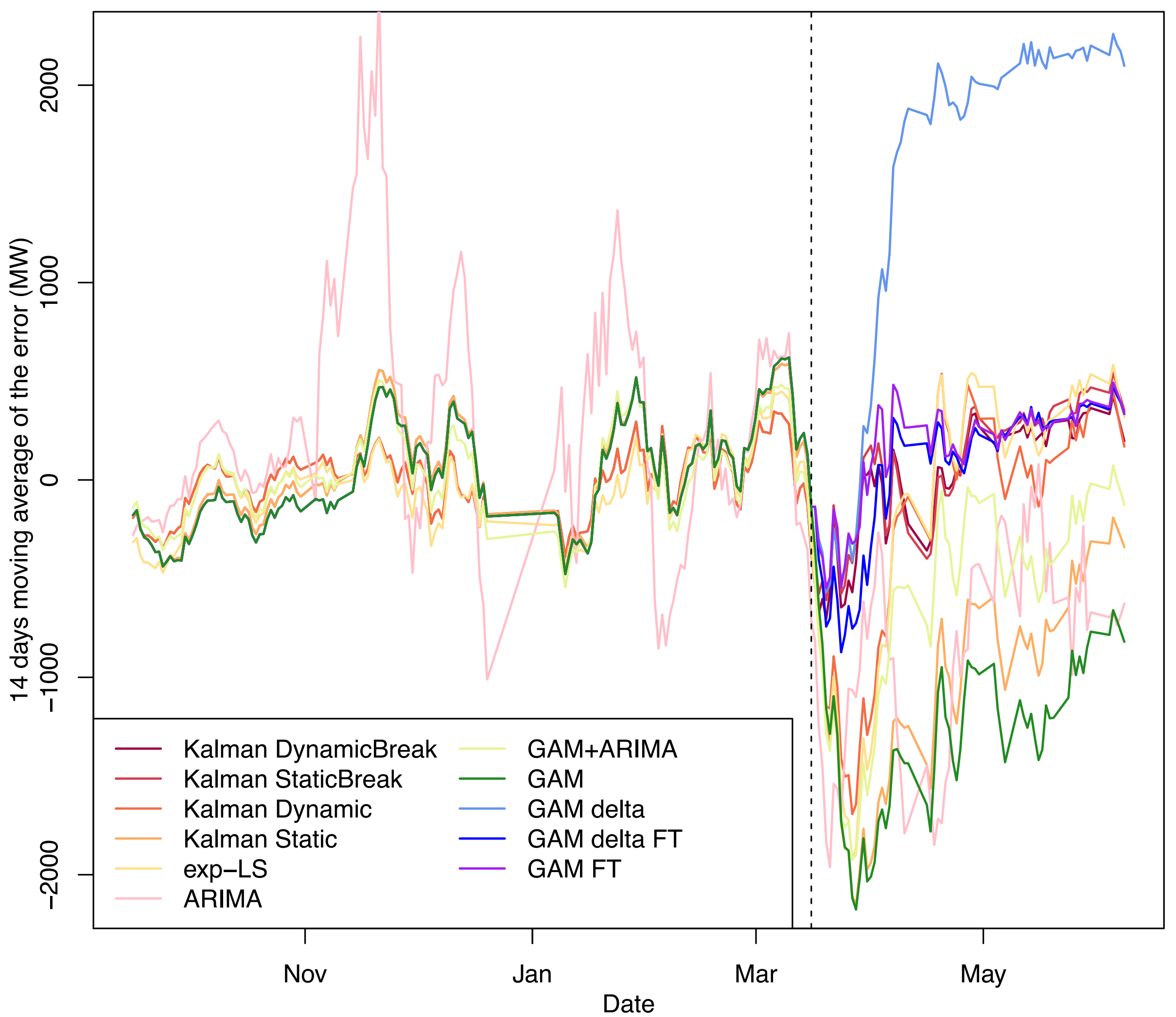
The advantage of GAM-δ is that it can be applied to reduce the prediction error starting at the very first day of lockdown.

One can afterwards combine this procedure with fine-tuning on the eventually available French data.

Algorithm 4: Transfer learning at time step t: GAM-δ fine-tuned

---

- Do fine-tuning on Italian data: $\tilde{\beta}_t = \hat{\beta}_S^{FR} + \rho\,\hat{\delta}_t$.
- Repeat K times:
  $$\tilde{\beta}_t \leftarrow \tilde{\beta}_t - \alpha \nabla \mathcal{L}_{t-1}^{FR}(\tilde{\beta}_t).$$
- Predict $\hat{y}_t = \tilde{\beta}_t^\top B(x_t)$.

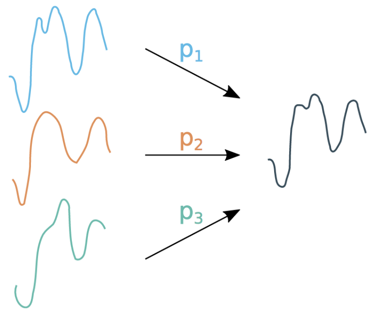


Method	2019/09/01 - 2020/03/15	2020/03/16 - 2020/04/15	2020/04/16 - 2020/06/07
ARIMA	4.10 %, 3341 MW	5.44 %, 3248 MW	5.59 %, 3135 MW
GAM	1.39 %, 1085 MW	4.83 %, 2961 MW	3.12 %, 1753 MW
GAM + ARIMA	1.34 %, 1050 MW	4.28 %, 2654 MW	2.65 %, 1464 MW
exp-LS	1.26 %, 982 MW	3.94 %, 2521 MW	1.97 %, 1029 MW
Kalman Static	1.38 %, 1077 MW	4.81 %, 2923 MW	2.85 %, 1588 MW
Kalman StaticBreak	-	2.79 %, 1954 MW	1.59 %, 855 MW
Kalman Dynamic	**1.26 %, 979 MW**	3.66 %, 2351 MW	1.89 %, 1002 MW
Kalman DynamicBreak	-	**2.73 %, 1902 MW**	**1.62 %, 854 MW**
Fine-tuned	-	2.78 %, 1917 MW	1.80 %, 938 MW
GAM δ	-	4.11 %, 2364 MW	6.09 %, 2713 MW
GAM δ - Fine-tuned	-	2.81%, 1912 MW	1.72 %, 905 MW

Online expert aggregation

- We sequentially observe a bounded sequence of observations $y_1, \dots, y_T$
- We forecast it step by step and have access at each time *t* to a set of ***experts,*** $x_{1,t}, \dots, x_{K,t}$ this experts could be any ML/physical model, human forecasts…
- We then build an aggregation forecast :

$$\widehat{y}_t = \sum_{j=1}^{K} p_{j,t} x_{j,t}$$



- Evaluation of the performances of the individual forecasts and the aggregation is measured with any convex loss e.g. $l_t(x) = (y_t - x)^2$
- The experts and the aggregation are then updated

Cesa-Bianchi, N., & Lugosi, G. (2006)

Online expert aggregation

- To fix the mind let's consider the EWA algorithm (Exponentially Weighted Aggregation)
- It depends on a single parameter (learning rate) η and the weights are updated this way:

$$p_{k,t} = \frac{\exp(-\eta \sum_{s=1}^{t-1}(y_s - x_{k,s})^2)}{\sum_{k=1}^{K} \exp(-\eta \sum_{s=1}^{t-1}(y_s - x_{k,s})^2)}$$

▶ Vovk, V. G. (1990)
▶ Warmuth & Littlestone (1994)

- Oracle bounds of this form can then be obtained (loss in [0,B])

$$\sum_{t=1}^{T}(y_t - \widehat{y}_t)^2 - \min_k \sum_{t=1}^{T}(y_t - x_{k,t})^2 \leq \frac{\log(K)}{\eta} + \eta\frac{B^2}{8}T \leq B\sqrt{\frac{T\log(K)}{2}}$$

$$\eta = \frac{1}{B}\sqrt{\frac{8log(K)}{T}}$$

- A priori information can be added by using sleeping experts: activation or not of an expert at time t

CRAN - Package opera

opera: Online Prediction by Expert Aggregation

▶ Wintenberger(2017)
▶ Gaillard, Stoltz & Van Erven (2014)
▶ Devaine, M., Gaillard, P., Goude, Y., & Stoltz, G. (2013)

How to choose the experts?

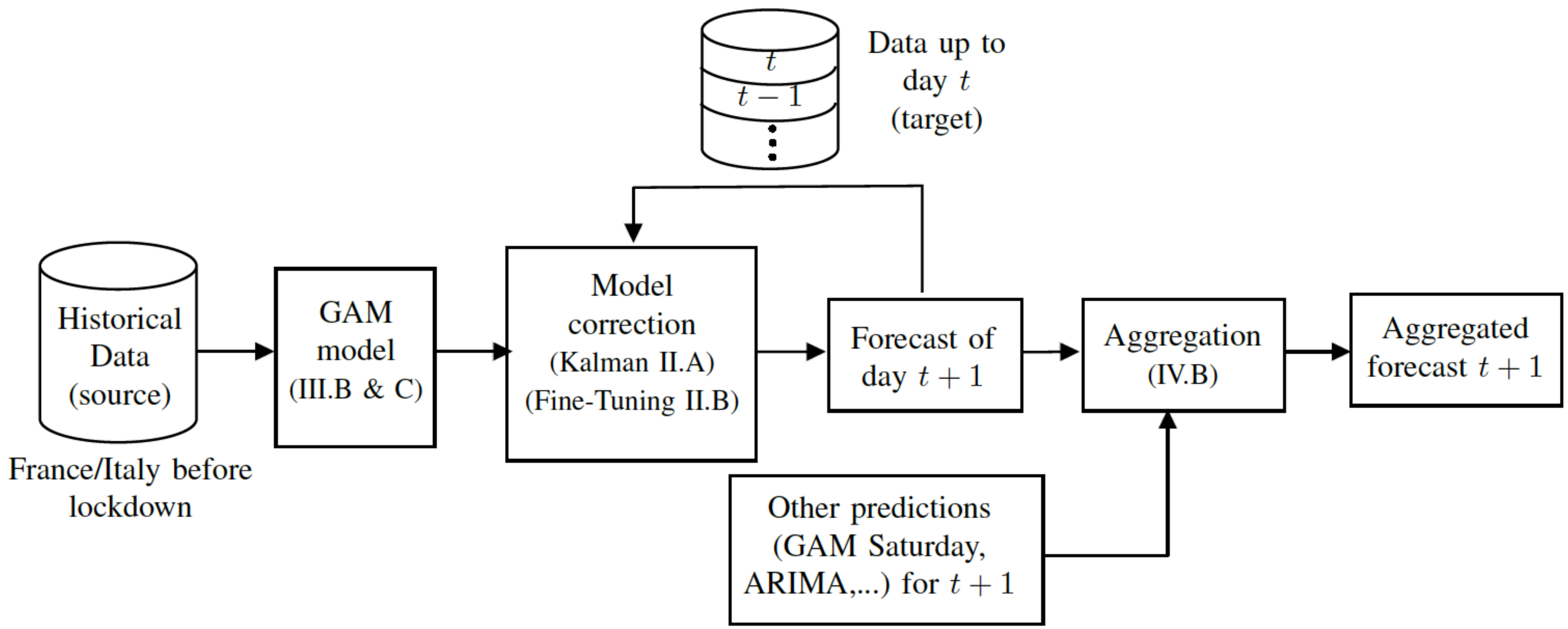
Trade-off between **diversity among the experts** / **performance of each expert**

In practice diversity can be obtained by different ways:

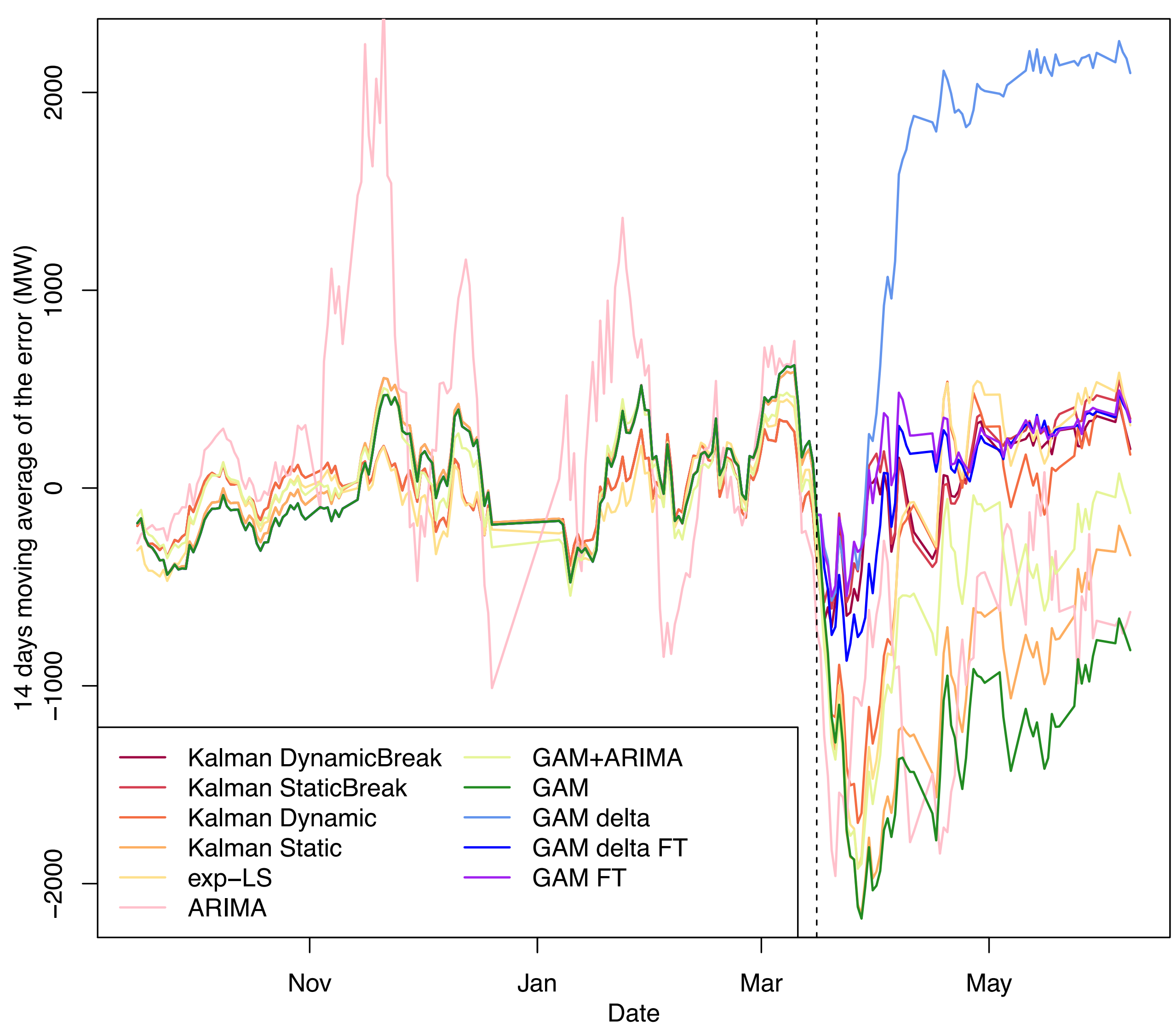
- **DATA**: using different data sets, manipulating data distribution (bagging, boosting…), feature set manipulation (RF), spatial/temporal resolution, time subsets.
- **METHODS**: linear models, additive models, non-parametric models, functional data regression, time series, RF, boosting,… Among a single method playing with parameters/objective functions.

Here we consider:

- DATA: french data, Italian data, time split according to different lockdown periods
- METHODS: GAM, ARIMA, Kalman (different Q), transfer, an « expert » correction of GAM: *GAM saturday*



Data up to day t (target)
t
$t-1$
Historical Data (source)
France/Italy before lockdown
GAM model (III.B & C)
Model correction (Kalman II.A) (Fine-Tuning II.B)
Forecast of day $t+1$
Aggregation (IV.B)
Aggregated forecast $t+1$
Other predictions (GAM Saturday, ARIMA,...) for $t+1$



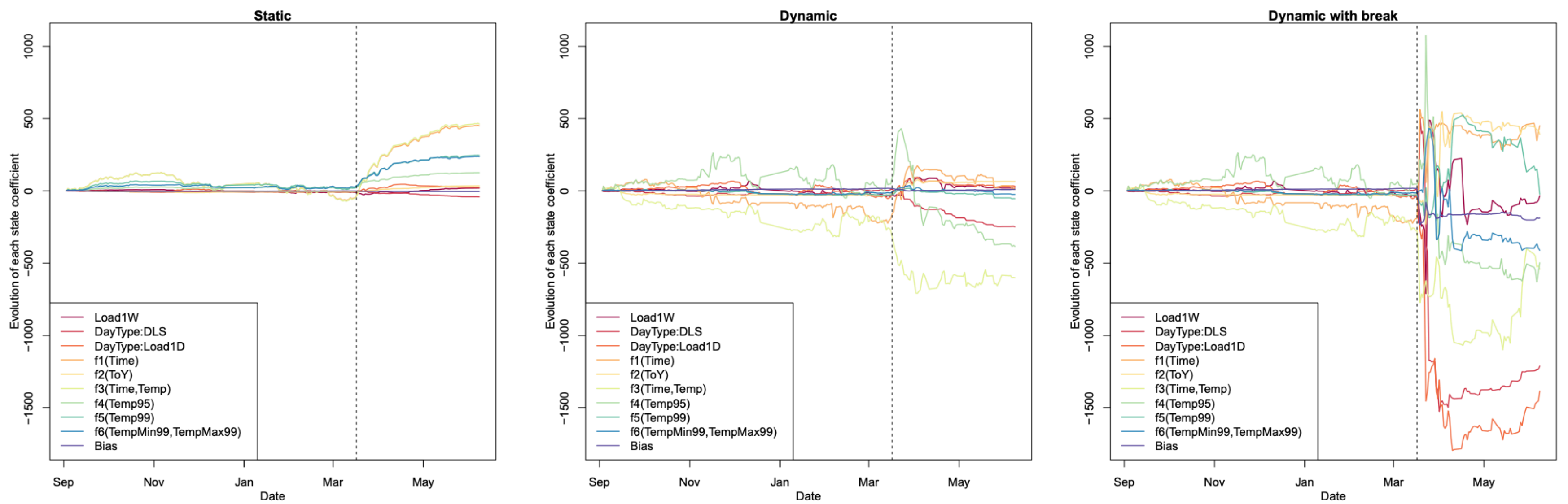
14 days moving average of the error (MW)
2000
1000
0
−1000
−2000
Nov
Jan
Mar
May
Date
Kalman DynamicBreak
Kalman StaticBreak
Kalman Dynamic
Kalman Static
exp−LS
ARIMA
GAM+ARIMA
GAM
GAM delta
GAM delta FT
GAM FT

Method	2019/09/01 - 2020/03/15	2020/03/16 - 2020/04/15	2020/04/16 - 2020/06/07
ARIMA	4.10 %, 3341 MW	5.44 %, 3248 MW	5.59 %, 3135 MW
GAM	1.39 %, 1085 MW	4.83 %, 2961 MW	3.12 %, 1753 MW
GAM + ARIMA	1.34 %, 1050 MW	4.28 %, 2654 MW	2.65 %, 1464 MW
exp-LS	1.26 %, 982 MW	3.94 %, 2521 MW	1.97 %, 1029 MW
Kalman Static	1.38 %, 1077 MW	4.81 %, 2923 MW	2.85 %, 1588 MW
Kalman StaticBreak	-	2.79 %, 1954 MW	1.59 %, 855 MW
Kalman Dynamic	**1.26 %, 979 MW**	3.66 %, 2351 MW	1.89 %, 1002 MW
Kalman DynamicBreak	-	**2.73 %, 1902 MW**	**1.62 %, 854 MW**
Fine-tuned	-	2.78 %, 1917 MW	1.80 %, 938 MW
GAM δ	-	4.11 %, 2364 MW	6.09 %, 2713 MW
GAM δ - Fine-tuned	-	2.81%, 1912 MW	1.72 %, 905 MW
GAM Saturday	8.33 %, 6425 MW	6.09 %, 3970 MW	8.40 %, 4616 MW
Aggregation without GAM Saturday	1.28 %, 1005 MW	3.01 %, 2014 MW	**1.44 %, 745 MW**
Aggregation with GAM Saturday	1.28 %, 1005 MW	**2.54 %, 1636 MW**	**1.49 %, 766 MW**

significant improvement (Diebold-Mariano test) between exp-LS and kalman/fine-tuning approaches
also for aggregation over kalman/fine-tuning approaches

Interpretability

- GAM and adaptive GAMS are interpetable by design but can suffer from identifiability issues, specially locally in time



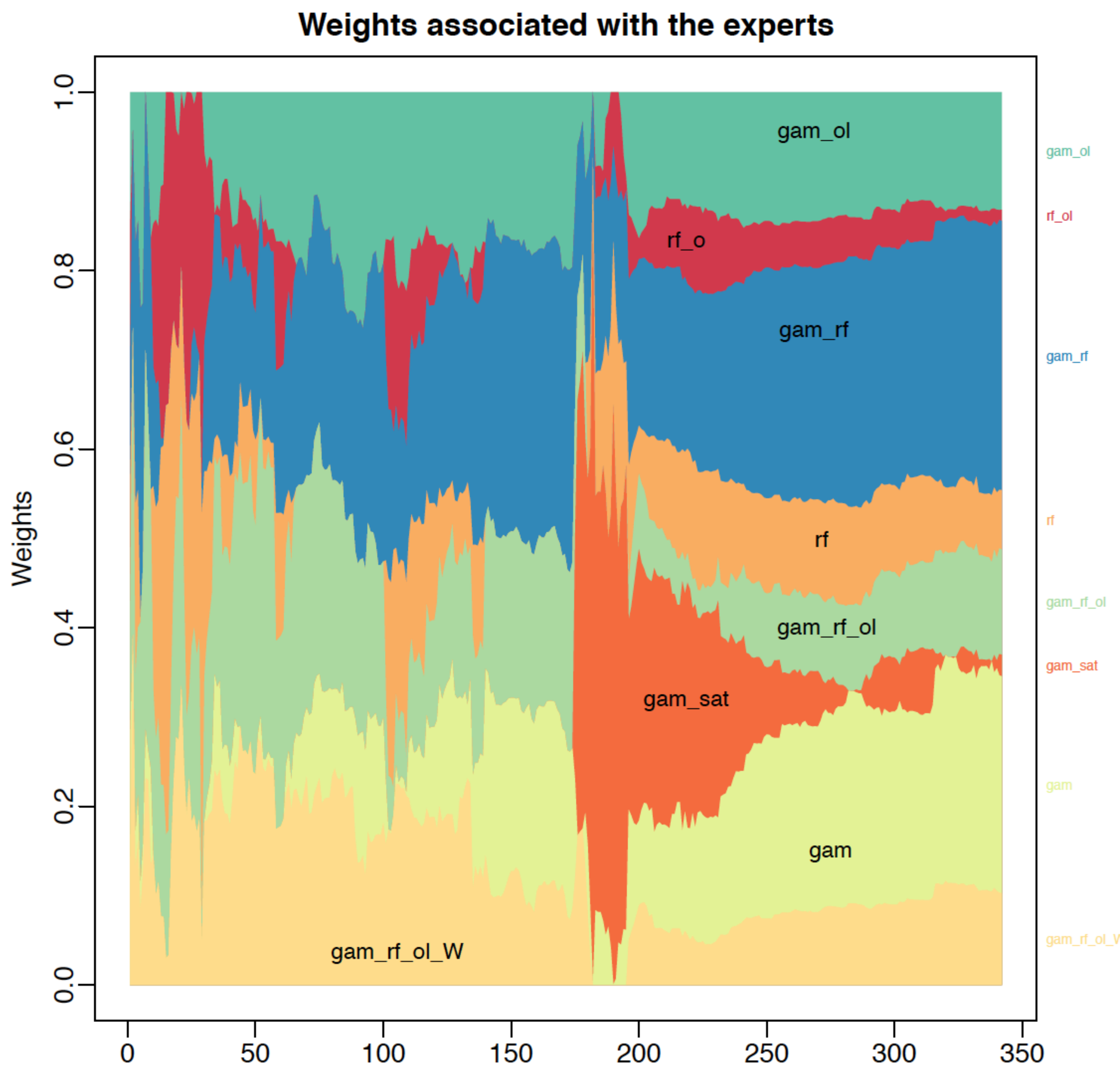
(a) Static (b) Dynamic (c) Dynamic with break

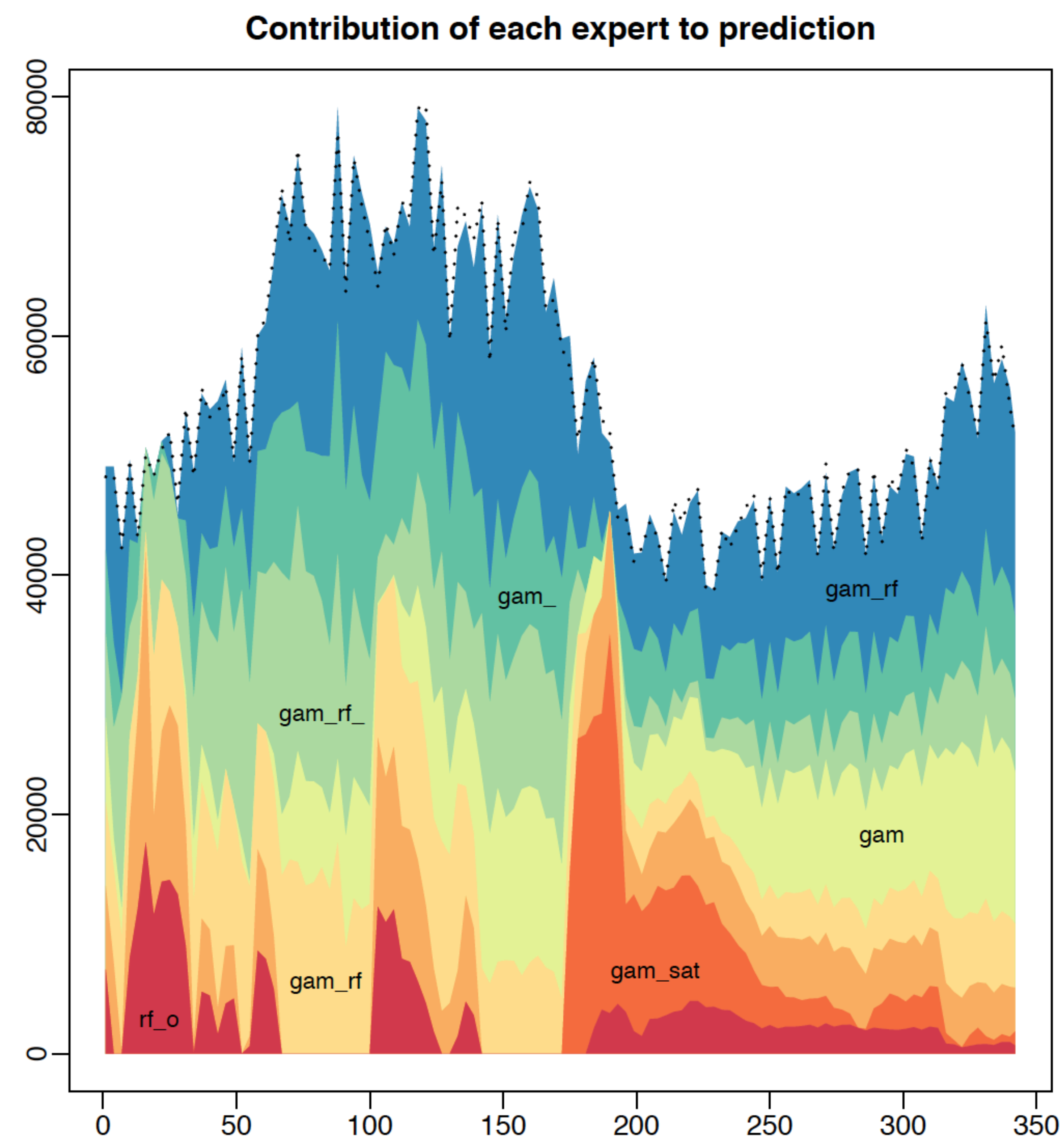
- Aggregation of experts can be seen as a forecasting (semi) blackbox that can be interpreted
  - using the time varying weights and the associated contribution to the forecasts

$$\widehat{y}_t = \sum_{j=1}^{K} p_{j,t} x_{j,t}$$

  - see it as a time varying blackbox and use marginal and/or ALE plots to interpretative it

Opera outputs

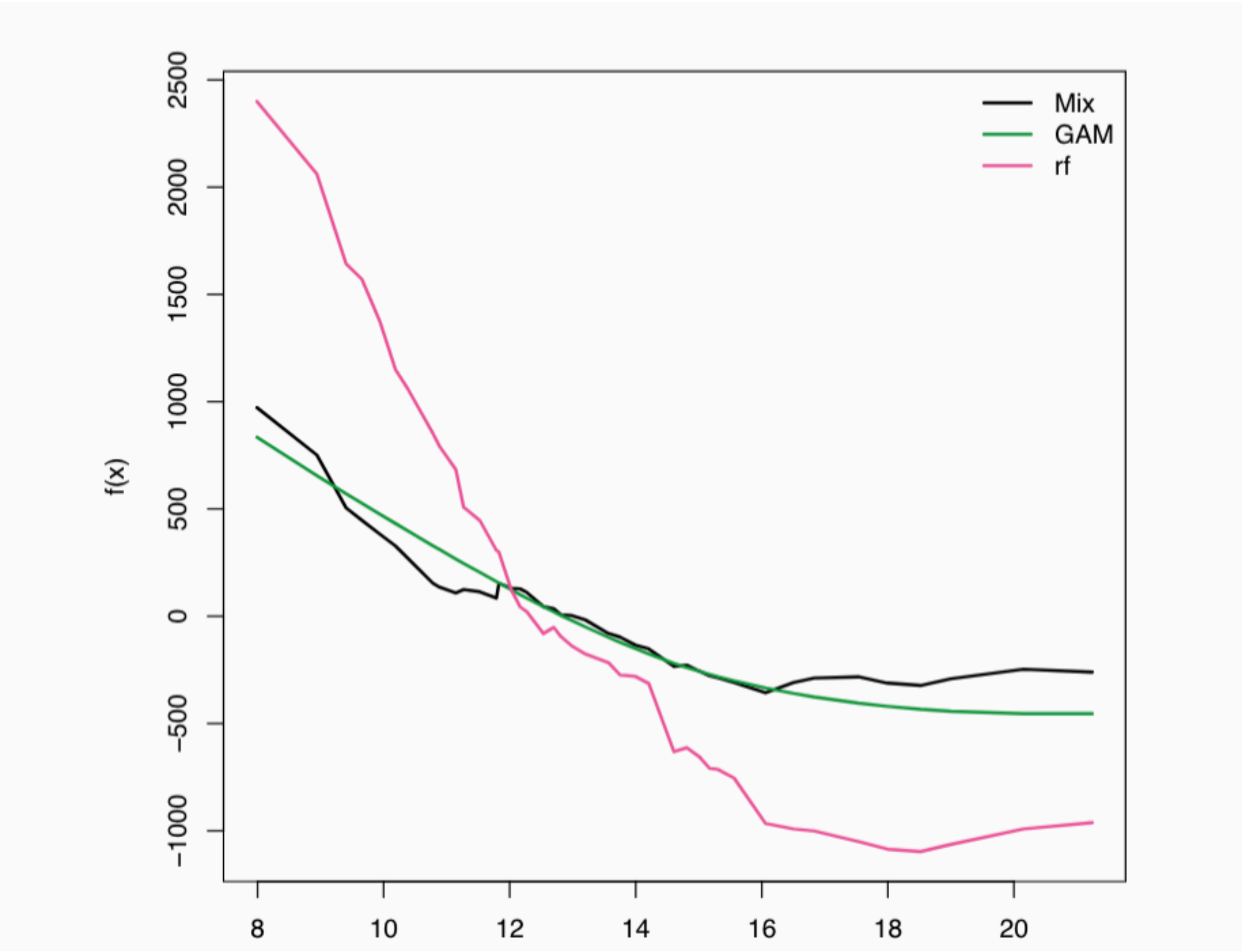




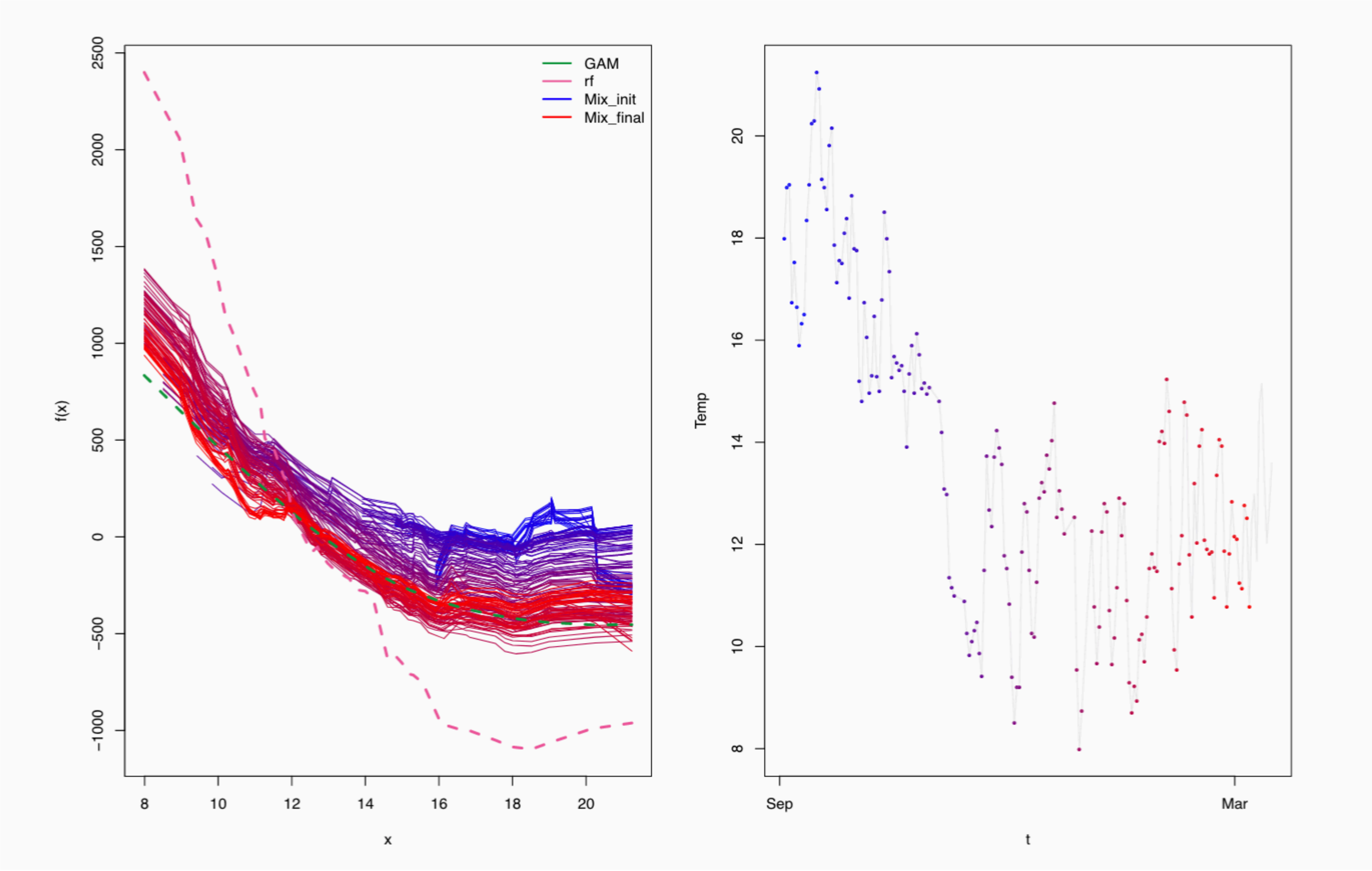
ALE plot Apley, D. W. and J. Zhu (2016) and adaptive ALE plots, exemple of a simple aggregation of 2 models (RF + GAM):

```
eq1_gam <- Load ~ s(DateN, k=3) + WeekDays:DLS + s(toy, k=20, bs='cc') +
                  ti(DateN, Temp, k=c(3,5)) +s(Temp_s95, k=5) +
                  s(Temp_s99, k=5) + s(Temp_s99_min, Temp_s99_max) + Load.48:WeekDays + Load.336


eq_rf <- Load ~ DateN  + WeekDays + DLS + toy + Temp+ Temp_s95 + Temp_s99 +
                Temp_s99_min + Temp_s99_max + Load.48 + Load.336
```



Winter 2019-2020 T° effects

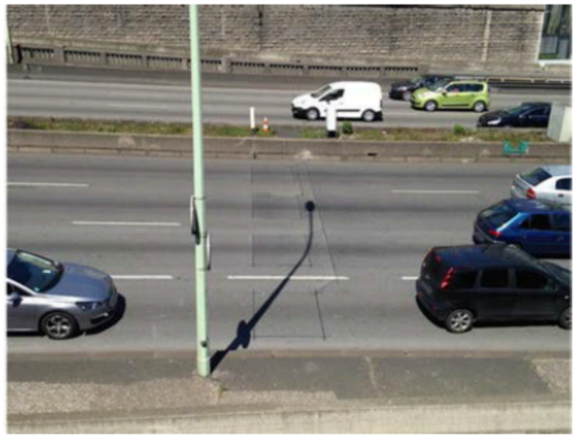


Time evolution of winter 2019-2020 T° effects

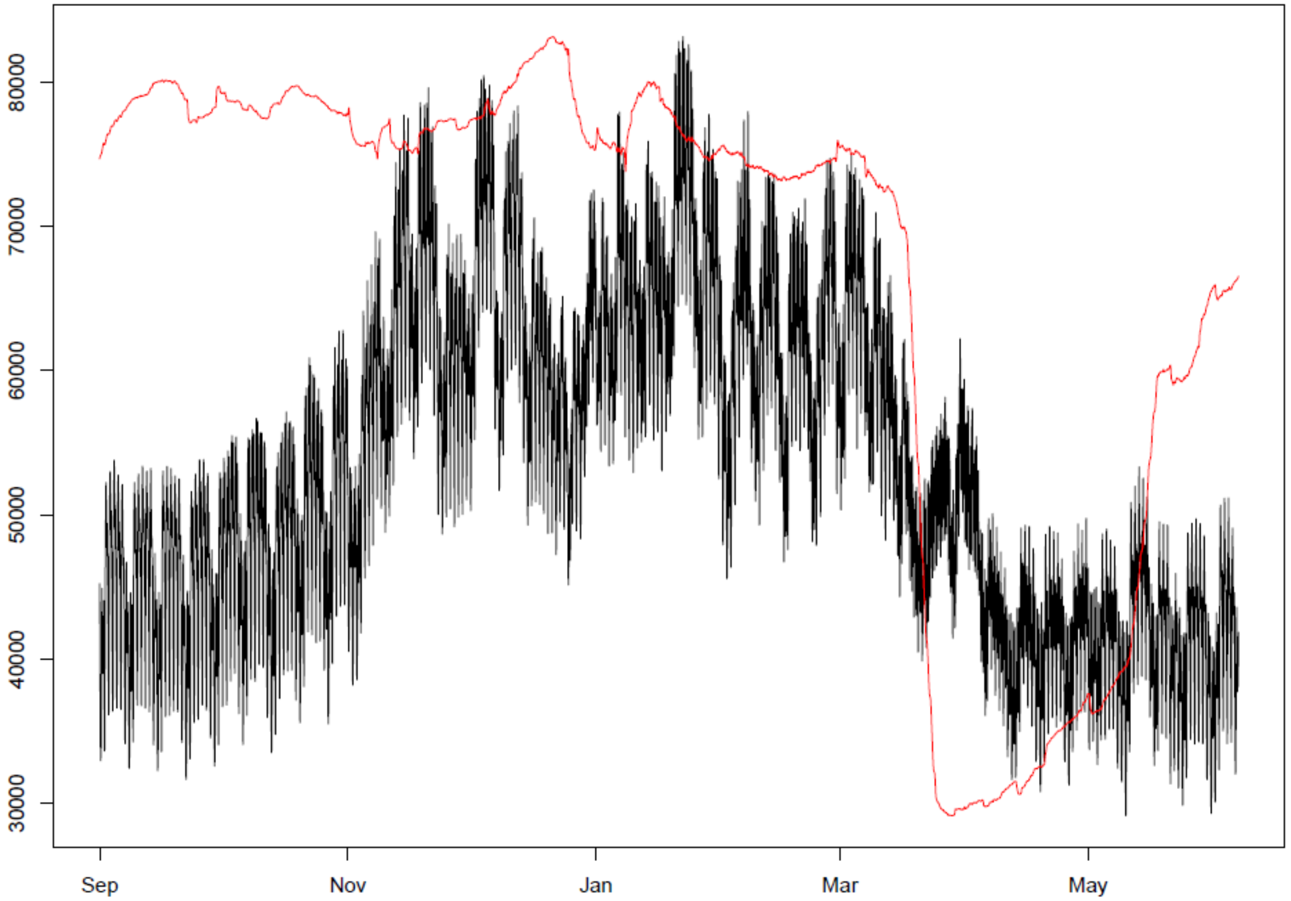
Complementary Data
Ongoing work

Mobility Data

Traffic Data: https://dataviz.cerema.fr/trafic-routier/



- Occupancy rate
- Traffic flow



GAM with smoothed traffic data improves by 1% (online update) error during lockdown after lockdown the effect of traffic is not consistent.
See also the work of Chen, Y., Yang, W., & Zhang, B. (2020) using mobility data from location app:
https://www.google.com/covid19/mobility/
https://www.apple.com/covid19/mobility

- Charansonney, L. (2018)
- Chen, Y., Yang, W., & Zhang, B. (2020)

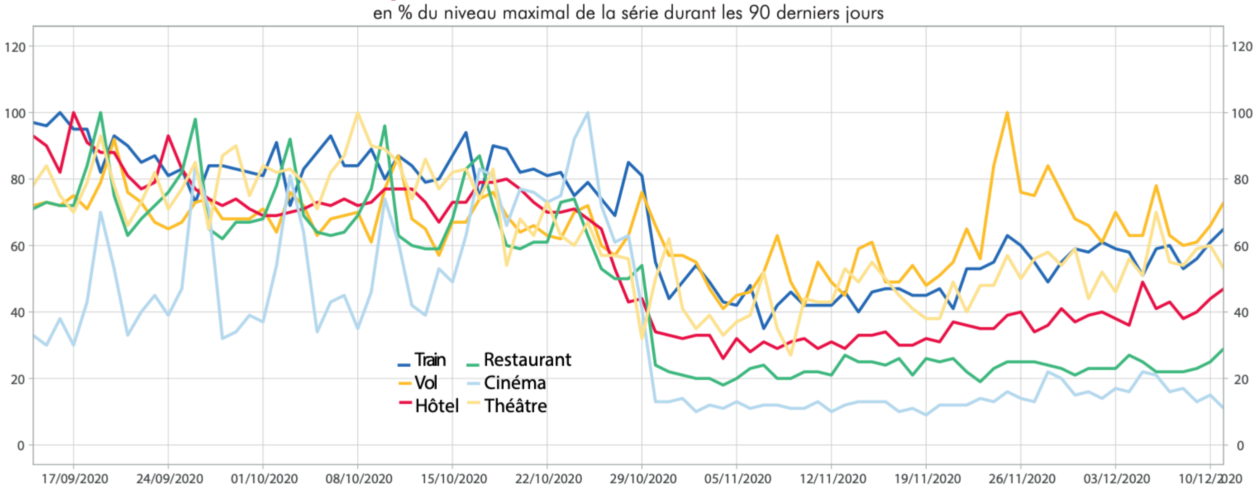
High frequency activity data

High frequency activity data is currently used by INSEE in his « note de conjoncture » to measure socio-economical impact of the COVID in France:

- -traffic data
- -google mobility data
- -google trend (keywords: Train, Vol, Hôtel, Restaurant, Cinéma, Théâtre…)
- -electricity load data

4 - Fréquence de recherche de mots-clés sur internet

en % du niveau maximal de la série durant les 90 derniers jours



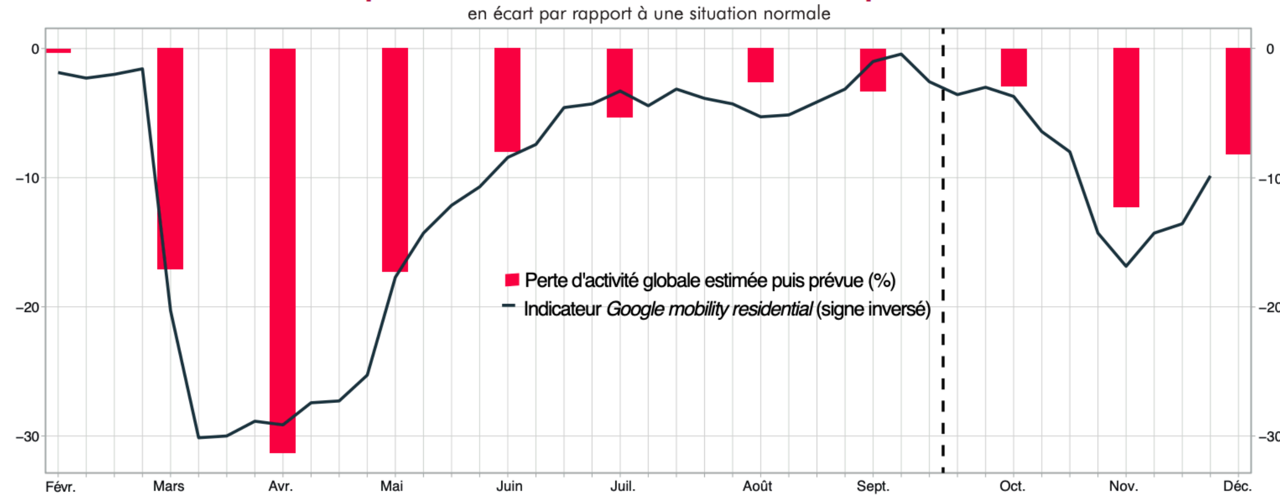
Lecture : le 11 décembre, la fréquence des requêtes du terme « vol » sur *Google* s'élevait à 73 % de son niveau maximal depuis mi-septembre, niveau atteint le 25 novembre.

Note : pour chaque série, l'indice est fixé à 100 au maximum de fréquence observé durant les 90 derniers jours.

Sources : Google Trends. Calculs Insee

3 - Indicateur de temps total hebdomadaire passé chez soi et pertes d'activité mensuelle estimées et prévues

en écart par rapport à une situation normale



Lecture : durant la première semaine de décembre, le temps passé chez soi a été supérieur de 10 % par rapport à une situation normale.

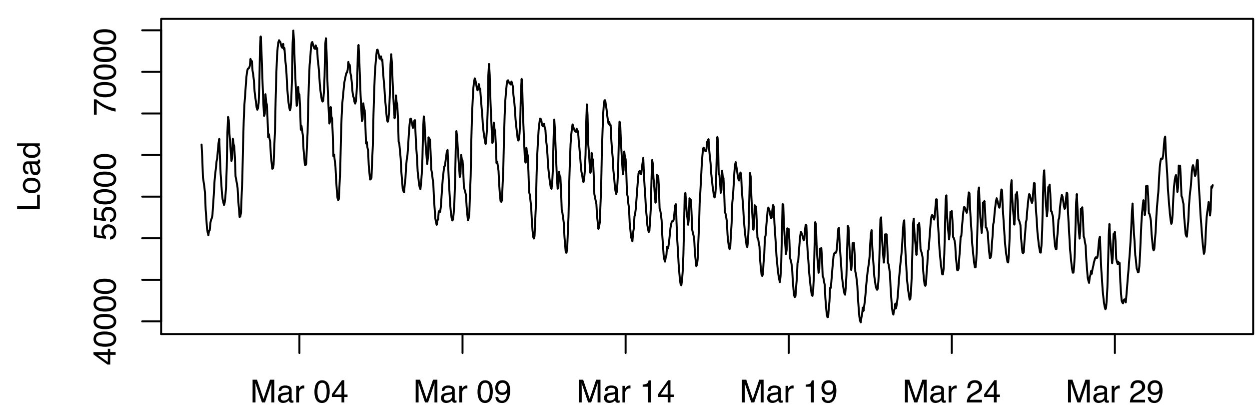
Note : les données de l'indicateur sont actuellement disponibles jusqu'au 6 décembre. Les valeurs hebdomadaires sont la moyenne des valeurs quotidiennes de l'indicateur.

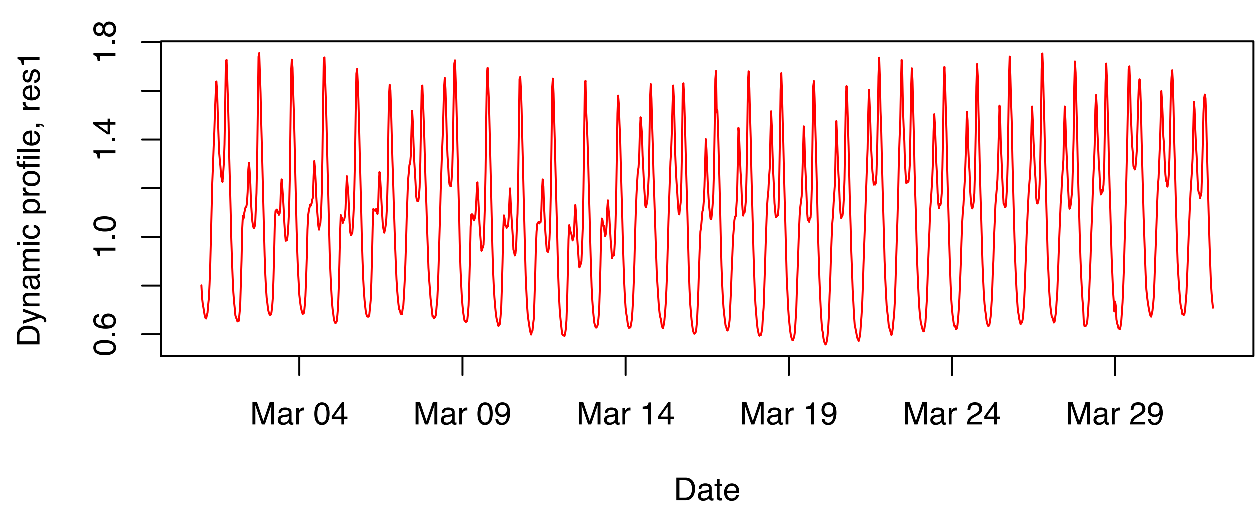
Source : Google Maps Mobility

NOTE DE CONJONCTURE Décembre 2020 © Insee 2020

Dynamic Panel from Smart Meters

Data published by Enedis (French DSO) since June 2018 https://www.enedis.fr/coefficients-des-profils





Conclusions / Perspectives

We exhibit the consequences of the lockdown on electricity time series forecasting in France

- Sudden change in level and shape of electricity load.
- Similarities in Italy/France, time shifted

Related statistical methodes/pbs:

- Online update: Kalman, aggregation of experts
- Transfer learning

Open problems/perspectives

- Enrich with new data from mobility, local/panel data, socio-economic data
- Spatio-temporal models to reflect local impact of the pandemic (regional level)
- Kalman and transfer learning with more black box models: RF, deep learning (see IEEE post-coved competition win)
- VIKING: Variational Bayesian Tracking to automatize the Kalman updates
- Other data: 1st place at IEEE competition https://ieee-dataport.org/competitions/day-ahead-electricity-demand-forecasting-post-covid-paradigm with similar methods

Open problems/perspectives regarding Interpretbility

- Error propagation among the different times updates (lag effects, kalman, aggregation)

- Adjakossa, E, Goude, Y. & Wintenberger, O. (2020) Kalman Recursions Aggregated Online (KAO) https://arxiv.org/abs/2002.12173, submitted to annals of applied statistics.
- Amato, U., Antoniadis, A., De Feis, I., & Goude, Y. (2017). Estimation and group variable selection for additive partial linear models with wavelets and splines. *South African Statistical Journal*, *51*(2), 235-272.
- Amato, U., Antoniadis, A., De Feis, I., Goude, Y. & Pichavant, A. (2020) Forecasting high resolution electricity demand data with additive models including smooth and jagged components, to appear in *International Journal of Forecasting,* 2020.
- Ba, A., Sinn, M., Goude, Y., & Pompey, P. (2012). Adaptive learning of smoothing functions: Application to electricity load forecasting. In *Advances in neural information processing systems* (pp. 2510-2518).
- Brown, G., Wyatt, J. L., & Tiňo, P. (2005). Managing diversity in regression ensembles. *Journal of machine learning research*, *6*(Sep), 1621-1650.
- Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge university press.
- De Vilmarest, J., & Wintenberger, O. (2020). Stochastic online optimization using kalman recursion. arXiv preprint arXiv: 2002.03636.
- D. Obst, J. De Vilmarest and Y. Goude (2021), Adaptive Methods for Short-Term Electricity Load Forecasting During COVID-19 Lockdown in France, in IEEE Transactions on Power Systems.
- Devaine, M., Gaillard, P., Goude, Y., & Stoltz, G. (2013). Forecasting electricity consumption by aggregating specialized experts. Machine Learning, 90(2), 231-260.
- Fasiolo, M., Wood, S.N., Zaffran; M., Nédellec, R. & Goude, Y. (2020). Fast calibrated additive quantile regression, to appear in Journal of American Statistical Association.
- Gaillard, P., Stoltz, G., & Van Erven, T. (2014, May). A second-order bound with excess losses. In *Conference on Learning Theory* (pp. 176-196).
- Gaillard, P., & Goude, Y. (2015). Forecasting electricity consumption by aggregating experts; how to design a good set of experts. In Modeling and stochastic learning for forecasting in high dimensions (pp. 95-115). Springer, Cham.
- Goehry, B., Goude, Y., Massart, P., & Poggi, J. M. (2019). Aggregation of Multi-scale Experts for Bottom-up Load Forecasting. *IEEE Transactions on Smart Grid.*
- VIKING: Variational Bayesian Variance Tracking Winning a Post-Covid Day-Ahead Electricity Load Forecasting Competition, Joseph de Vilmarest, Yannig Goude, Olivier Wintenberger, accepted at ICML 2021 Time Series Workshop.
- Goude, Y., Nedellec, R., & Kong, N. (2013). Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE transactions on smart grid*, *5*(1), 440-446.
- Hastie, T. J., & Tibshirani, R. J. (1990). Generalized Additive Models, volume 43 CRC Press.
- Launay, T., Philippe, A., & Lamarche, S. (2015). Construction of an informative hierarchical prior for a small sample with the help of historical data and application to electricity load forecasting. Test, 24(2), 361-385.
- Pierrot, A., & Goude, Y. (2011). Short-term electricity load forecasting with generalized additive models. *Proceedings of ISAP power*, *2011*.
- Thouvenot, V., Pichavant, A., Goude, Y., Antoniadis, A., & Poggi, J. M. (2015). Electricity forecasting using multi-stage estimators of nonlinear additive models. *IEEE Transactions on Power Systems*, *31*(5), 3665-3673.
- Vovk, V. G. (1990). Aggregating strategies. *Proc. of Computational Learning Theory, 1990*.
- Warmuth, M., & Littlestone, N. (1994). The weighted majority algorithm. *Information and Computation*, *108*(2), 212-261.
- Wintenberger, O. (2017). Optimal learning with Bernstein online aggregation. *Machine Learning*, *106*(1), 119-141.
- Wood, S. N. (2017). Generalized additive models: an introduction with R. Chapman and Hall/CRC.
- Bird, J. J., Kobylarz, J., Faria, D. R., Ekárt, A., & Ribeiro, E. P. (2020). Cross-Domain MLP and CNN Transfer Learning for Biological Signal Processing: EEG and EMG. IEEE Access, 8, 54789-54801.
- Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
- Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. In Advances in neural information processing systems (pp. 3347-3357).
- Charansonney, L. (2018). " Nouvelles données de trafic": quels usages pour les gestionnaires de voirie dans l'évaluation des politiques de circulation? L'exemple parisien de la Voie Express Rive Droite..
- Chen, Y., Yang, W., & Zhang, B. (2020). Using Mobility for Electrical Load Forecasting During the COVID-19 Pandemic. arXiv preprint arXiv:2006.08826.
- Apley, D. W. and J. Zhu (2016). Visualizing the effects of predictor variables in black box supervised learning models, arXiv preprint arXiv :1612.08468