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

David Obst, Joseph de Vilmarest, and Yannig Goude

***Abstract*—The coronavirus disease 2019 (COVID-19) pandemic has urged many governments in the world to enforce a strict lockdown where all nonessential businesses are closed and citizens are ordered to stay at home. One of the consequences of this policy is a significant change in electricity consumption patterns. Since load forecasting models rely on calendar or meteorological information and are trained on historical data, they fail to capture the significant break caused by the lockdown and have exhibited poor performances since the beginning of the pandemic. In this paper we introduce two methods to adapt generalized additive models, alleviating the aforementioned issue. Using Kalman filters and fine-tuning allows to adapt quickly to new electricity consumption patterns without requiring exogenous information. The proposed methods are applied to forecast the electricity demand during the French lockdown period, where they demonstrate their ability to significantly reduce prediction errors compared to traditional models. Finally, expert aggregation is used to leverage the specificities of each predictions and enhance results even further.**

***Index Terms*—COVID-19, load forecasting, model adaptation, time series.**

I. Introduction

Accurate electricity load forecasting is of paramount importance for the balancing of the electricity grid, since they are the main inputs of the production planning at different horizons [1] and storage capacities are still limited regarding the consumption needs. Load forecasting is performed at different horizons of time, ranging from intra-day (10 minutes to 24 hours ahead) to daily, weekly, monthly or even a few years in advance for industrial needs covering production planning, demand response, grid management, electricity trading, risk management, optimization of production units maintenance and commercialization.

Manuscript received September 24, 2020; revised February 1, 2021; accepted March 6, 2021. Date of publication March 22, 2021; date of current version August 19, 2021. Paper no. TPWRS-01628-2020. *(David Obst and Joseph de Vilmarest contributed equally to this work). (Corresponding author: David Obst.)*

David Obst is with the Électricité de France R&D, 91120, Palaiseau, France and also with the Institut de Mathématiques de Marseille, Aix-Marseille Université, Marseille, France (e-mail: david.obst@edf.fr).

Joseph de Vilmarest is with the Électricité de France R&D, 92140, Clamart, France and also with the Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, 75006, Paris, France (e-mail: joseph.de-vilmarest@edf.fr).

Yannig Goude is with the Électricité de France R&D, 92140, Clamart, France and also with the Laboratoire de Mathématique d'Orsay, Université Paris-Saclay, 91190, Gif-sur-Yvette, France (e-mail: yannig.goude@edf.fr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2021.3067551.

Digital Object Identifier 10.1109/TPWRS.2021.3067551

The field has been thoroughly studied the past decades, especially by the time series, statistics and machine learning communities. Time series approaches are very efficient for very-short term forecasts (typically less than 24 hours ahead). They rely on auto-regressive moving-average (ARIMA) models [2] or functional approaches [3], [4] exploiting daily and weekly patterns in the electricity load data. Machine learning models are usually stronger at incorporating exogenous information for short and mid-term predictions (more than 1 day ahead). They use calendar characteristics (such as the time of the year, day of the week...) as well as meteorological effects (temperature, wind speed) or tariff options as inputs and are then trained on a large set of historical data (usually 3 to 5 years). A good overview of load forecasting practices has been given by the Global Energy Forecasting Competitions (GEFCOM) [5]. Popular algorithms include black box machine learning models such as gradient boosting machines [6] and neural networks [7], [8] or statistical models like Generalized Additive Models (GAM) [9]–[12]. Black box models are attractive due to their good forecasting performances but generally suffer from their lack of interpretability. GAMs are very attractive to electric utilities as they combine the flexibility of fully nonparametric models, the simplicity of multiple regression model and are computationally efficient to scale with big data [13]. The main French electricity provider, EDF (Électricité de France) uses GAM as their lead forecasting tool.

However the coronavirus pandemic has significantly affected consumption patterns all over the world. As presented in [14], [15], the closure of nonessential businesses as well as the stay-at-home directives have led to a significant drop of the power demand and changes in the daily consumption patterns. Fig. 1 shows the French and Italian electricity load over time in 2020, whose decrease due to the lockdown (which happens before in Italy) is clearly seen. Daily profiles of the French consumption before and after the lockdown are represented in Fig. 2. After lockdown for both countries the daily shapes of the load have converged towards the one of Saturdays.

Since models are trained on historical data and make the underlying assumption that future behavior will be similar to past one, they will fail to produce satisfactory predictions during the lockdown period. For instance in France GAM usually achieve around 1% MAPE (mean absolute percentage error) [9], but were around 5% during the first few weeks of the lockdown thus requiring manual intervention to correct the model forecasts. Not only do these poor forecasts have a high cost for electricity producers and system operators, but they represent a threat to the

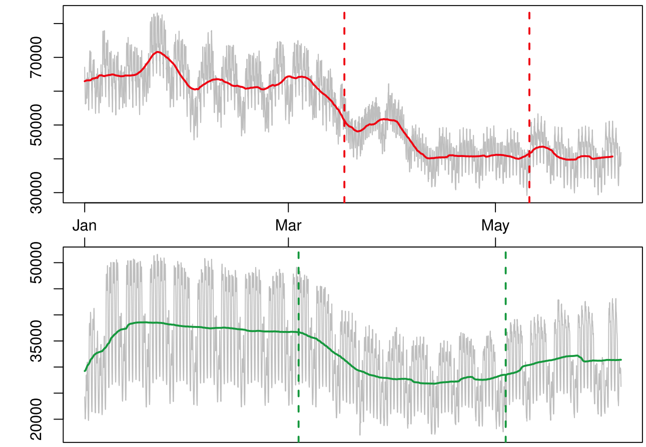


Fig. 1. French and Italian electricity load (in MW) at resp. half-hourly and hourly resolution in 2020. Dashed lines are the starting and ending date of the lockdown.

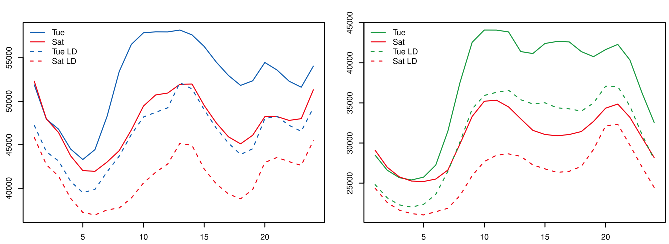


Fig. 2. French and Italian electricity Tuesday and Saturday load profiles before and during the lockdown (Dashed lines).

proper functioning of the electrical network as well, which could have even more consequences than usual during a pandemic.

This is why finding novel approaches to better predict the load demand during these troubled times is of paramount importance. However to our knowledge, up to this date only a few papers have addressed this problem. [16] is among the first to propose an efficient strategy to improve the predictions during the COVID-19 lockdown period in France. Using an adaptive functional state-space model and assimilating the period to non-workable days, the author was able to achieve significantly better performance compared to the French system operator. However these models lack of interpretability, making other approaches preferred in the industrial context. Furthermore the aforementioned work requires to artificially set all days to weekends or holidays, which may be unviable in the long-term. In [17] the integration of mobility data is combined with multi-task learning to improve the forecasting during the lockdown. They show that mobility is indeed a relevant feature that should be integrated in load demand models, and that joint training of a neural network for multiple geographical areas yields additional benefits and compensates for the lack of data. Nonetheless their forecasting errors remain high compared to pre-COVID standards, neural networks lack of interpretability as well and the introduction of exogenous features can be problematic in the future due to the sustainability of such data in operations.

We consider here the framework of GAM and propose two new adaptive versions of these models. The idea of adaptive models is to take advantage of data observed in an online fashion to update an initial model. This will make them able to adapt to the changes in consumption patterns spontaneously, without exogenous variable or intervention. In every adaptive forecasting method a trade-off has to be found between a good reactivity to a change (whether it is a smooth drift or a break) and a good behavior during stable periods. One of the most popular algorithm for that is the Kalman filter [18] already applied to electricity load forecasting in [19] and [20]. We propose here to couple Kalman filters with GAM to obtain a forecasting procedure which performs well before the lockdown, exploiting the nice properties of GAM but also reacting quickly to the sudden change in the data at the beginning of the lockdown. The second approach we present leverages ideas from transfer learning to fine-tune a GAM on the lockdown period. Transfer learning (also referred as learning-to-learn or knowledge transfer) is a branch of machine learning that aims at reusing knowledge from one source task on another target one [21], [22]. It has shown great success, particularly when the source data is plentily available and the target one scarce. Recently it has even found applications for electricity load forecasting to transfer information from one set of customers to another one [23]. In our case our source data will be the data before the lockdown and the target one the data during the lockdown in the country of interest (France in our study), or even a similar one where the lockdown came before (e.g. Italy here). The contributions of our work are the following:

1) Two mathematical approaches are proposed to efficiently adjust a historical model to consumer behavior change over time, even in the case where data is scarce. Furthermore they do not require the integration of additional features.
2) The two methodologies have been successfully applied on the difficult period of the COVID-19 lockdown in France, achieving forecast accuracy close to the one observed before the pandemic.
3) An empirical strategy is suggested to anticipate the impact of the lockdown on the load using another country's data, thus enabling satisfactory predictions from the very first day of stay-at-home order.

The rest of the paper is organized as following. In Section II we introduce the two model adaptation methods relying on Kalman filtering and fine-tuning. Section III presents the data and the GAM model used for the French load and Section IV summarizes the main results of our experiments. Finally Section V concludes our study and suggests further work.

II. Adaptation of Additive Models

We consider additive models whose assumption is that the response variable y_t is decomposed as

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

where (ε_t) is an independent identically distributed (i.i.d.) random noise, $\boldsymbol{x}_t = (x_{t,1}, \ldots, x_{t,d})$ are the explanatory variables

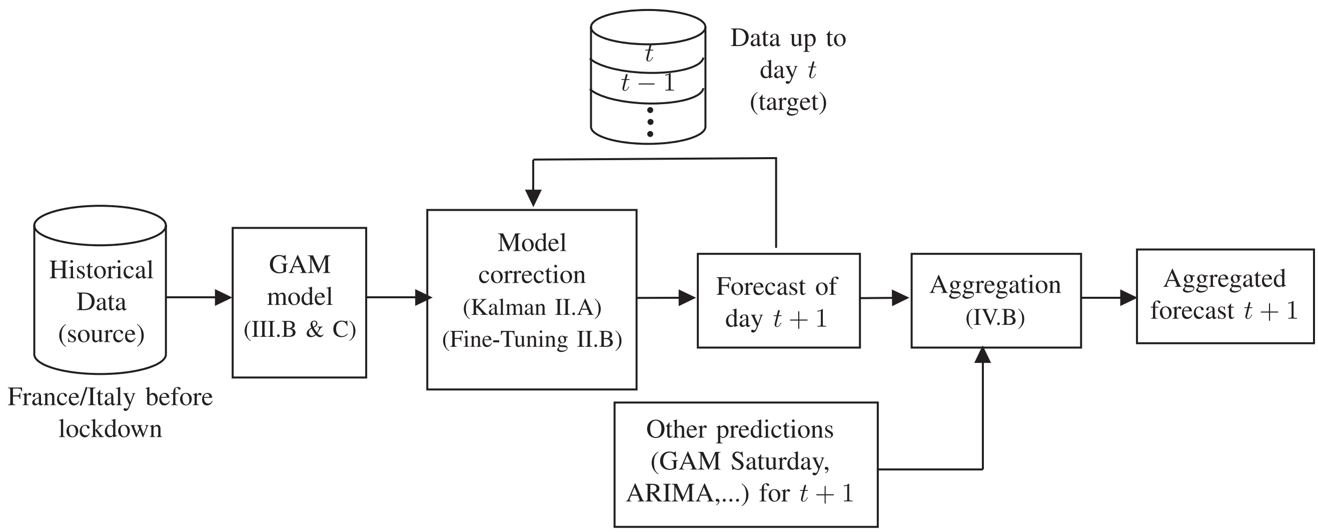


Fig. 3. Flowchart of adaptive methods.

at time t, and each nonlinear effect f_j is decomposed on a spline basis $(B_{j,k})$ with coefficients $\boldsymbol{\beta}_j$:

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

where m_j depends on the dimension of the spline basis. The f_j's are centered to ensure the identifiability of the model, and more details concerning the basis are given in Subsection III-B. The coefficients $\beta_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d$ then are estimated by penalized least-squares. The penalty term involves the second derivatives of the functions f_j, forcing the effects to be smooth (see [24]).

The random residuals ε_t are supposed to be Gaussian i.i.d. in the first place. Later in the numerical experiments we will introduce another variant of this model, where the residuals are supposed to be an ARIMA model optimised with classical time series methods. We focus here on structural adaptation of the GAM over time. We present two different levels of adaptation. Firstly, we consider the reduced problem of adapting a linear combination of the frozen effects $f_1, \ldots, f_d$. Secondly we try to adapt the whole model by fine-tuning. A flowchart of the forecasting pipeline is given in Fig. 3.

A. Multiplicative Correction of the Effects

In order to reduce the dimension of the adaptation problem, a strategy is to freeze the nonlinear effects, and to correct these effects by a multiplicative factor. Precisely, we define $f(\boldsymbol{x}_t) = (1, \overline{f}_1(x_{t,1}), \ldots, \overline{f}_d(x_{t,d}))^\top$ where $\overline{f}_j$ is a normalized version of f_j obtained by subtracting the mean on the train set and dividing by the standard deviation. Then we adaptively estimate a vector $\boldsymbol{\theta}_t$ such that

$$\mathbb{E}[y_t \mid \boldsymbol{x}_t] = \boldsymbol{\theta}_t^\top f(\boldsymbol{x}_t)\,.$$

The estimator at time t will be denoted as $\hat{\boldsymbol{\theta}}_t$ in both Section II-A1 and Section II-A2. Thus we reduce the number of coefficients from $1 + \sum_{j=1}^d m_j$ to $1 + d$. This is a good trade-off to obtain a simple model which will react quickly to a break in the data generation process but also complex enough to fit well with the nonlinear properties of the load.

1) Exponential Least-Squares (exp-LS): An empirical method consists in solving at each step a least-squares problem where we specify a weight decreasing exponentially with the time difference. Precisely we define

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

and we predict $\hat{y}_t = \hat{\boldsymbol{\theta}}_t^\top f(\boldsymbol{x}_t)$. This formalisation leads to a single parameter, the exponential forgetting factor μ. The advantage of this type of adaptation lies in its simplicity. The forgetting factor μ is determined by minimizing the RMSE on a validation set composed of the last year of the train set for a GAM trained on the beginning of the train set, then we keep the same μ for the GAM trained on the whole train set. Previous work has been done on estimating this parameter online, but leads to computational issues and potential instability of the model (see [25]).

2) Kalman Filter: We present also a state-space model approach. We assume the following equations:

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

where (ε_t) and $(\boldsymbol{\eta}_t)$ are gaussian white noises of respective variance/covariance σ^2 and Q. This is the setting of Kalman filtering [18], thus we use the recursive formulae of Kalman providing the expectation and covariance of the state $\boldsymbol{\theta}_t$ given the past observations, and these estimators yield the mean and variance of y_t given the past. This is described in Algorithm 1. Note that the exp-LS method has a very similar recursive form starting from t_0 such that $P_{t_0} = (\sum_{s=1}^{t_0-1} e^{-\mu(t_0-s)} f(\boldsymbol{x}_s) f(\boldsymbol{x}_s)^\top)^{-1}$ exists. Indeed, the same update rule stands for $\hat{\boldsymbol{\theta}}_t$ (with $\sigma = 1$) and the update on P_t is the following:

$$P_{t+1} = e^{\mu}\left(P_t - \frac{P_t f(\boldsymbol{x}_t) f(\boldsymbol{x}_t)^\top P_t}{f(\boldsymbol{x}_t)^\top P_t f(\boldsymbol{x}_t) + 1}\right)\,.$$

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:

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

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

The simplicity stands in a single scalar parameter e^μ as multiplicative factor for the update of P_t, whereas Kalman Filter needs a matrix parameter Q added in the recursion.

There is a wide literature concerning the setting of the hyper-parameters $\hat{\boldsymbol{\theta}}_1, P_1, \sigma^2, Q$ on which the Kalman Filter crucially relies, see for instance [26]–[28]. We observe that the iterates of $\hat{\boldsymbol{\theta}}_t$ depend only on $\hat{\boldsymbol{\theta}}_1, P_1^* = P_1/\sigma^2$ and $Q^* = Q/\sigma^2$, reducing the set of hyper-parameters as in [26].

An interesting degenerate covariance matrix is the static setting $Q^* = 0$ (the state equation becomes $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t$). Defining $\hat{\boldsymbol{\theta}}_1 = 0,\ P_1^* = I$, the estimate $\hat{\boldsymbol{\theta}}_t$ becomes a Ridge Forecaster:

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

To obtain a dynamic setting we maximize the likelihood on the training set. The Expectation-Maximization algorithm is a renowned algorithm allowing to find a local optimum. However the lack of global guarantee makes it inefficient in our case, and we applied instead some kind of grid search. Precisely we decided to set $P_1^* = I$ as in the static setting, and for a given Q^* the optimal $\hat{\boldsymbol{\theta}}_1$ for the likelihood has a closed-form solution. Q^* is of dimension 10×10 and we chose to restrict ourselves to diagonal matrices whose coefficients are in the set $\{2^j, -30 \leq j \leq 0\}$. This is still a set of around $8 \cdot 10^{14}$ elements, thus we used an iterative greedy procedure: we start from $Q^{*(0)} = 0$ and at each step, having $Q^{*(k)}$ in hand, we compute the likelihood of each matrix where only one coefficient differ from $Q^{*(k)}$, and we define $Q^{*(k+1)}$ as the one maximizing the likelihood among those tested. This algorithm yielded less than 10^4 evaluations of the likelihood.

In order to take the lockdown into account in the state-space representation, it is natural to consider a varying state noise covariance Q_t. Indeed, we expect the model to change much faster during and after the lockdown than before. It motivates a dynamic estimation of Q_t, however due to the amplitude of the crisis we modelled a break in the data at the lockdown beginning. We chose to change only the state noise covariance at the break time T, and for $t \neq T$ we use $Q_t^* = 0$ in the static setting or $Q_t^* = Q^*$ in the dynamic setting. We don't want to put any *a priori* on the break, therefore we defined $Q_T^* = P_1^* = I \gg Q^*$.

B. Correction of the Full Model

In the previous methods the nonlinear effects $f_j(\cdot)$ were frozen and adjusted with a multiplicative factor. However it may be insufficient on certain new types of behavior. Since learning a new model from scratch is inadvisable considering the few samples of target data available, we would like to start from the model trained on historical data and adapt it on the few instances available. This is a particular case of the framework of transfer learning, more specifically of model fine-tuning (FT). It consists in reusing a part of the parameters learned on the source set (typically neural network layers) and adjust them with a few gradient iterations on the target one for instance. Model fine-tuning has been successful in different fields such as computer vision [29] or even time series forecasting [30].

In our case we will fine-tune the parameters of our GAM. Since it boils down to a penalized linear regression problem, it consists in fine-tuning a linear model. This framework was elaborated in [31]. Starting from the coefficients $\hat{\boldsymbol{\beta}}_0$ learned on the historical source data, for each time step we perform K iterations of batch gradient descent with fixed step size α on the following objective function to yield an adjusted parameter vector $\hat{\boldsymbol{\beta}}_t$:

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

Let $B(\boldsymbol{x}_s)$ be the vector of the $B_{j,k}(x_{s,j})$ and $B(X_t)$ denote the matrix made by the concatenation (by row) of the $B(\boldsymbol{x}_s)$ for $s = 1, \ldots, t-1$. Again more details concerning the basis $(B_{j,k})$ is found in Section III-B. As discussed by the aforementioned paper, the choice of the step size α is not crucial, as long as it is small enough. In practice a good step size is $\alpha = \alpha^*/5$ where $\alpha^* = 2/(\lambda_{\max}(B(X_t)^\top B(X_t)) + \lambda_{\min}(B(X_t)^\top B(X_t)))$ with $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$ being respectively the maximum and minimum eigenvalue of M. Ergo the major hyperparameter to tune is K the number of gradient iterations to perform. Theoretical methods are currently being investigated in the aforementioned paper and have been used to guide our choice here, but it was also observed empirically that K between 50 and 100 yields good results. Therefore a number of iterations in that range is always considered, and this choice usually coincides with the suggested theoretical guidelines.

III. Data and Model Presentation

In this section we detail the GAM model that has been used to forecast the French electricity consumption, as well as the data on which is has been applied.

A. Presentation of the Data

The French electricity consumption is freely available on the website of the system operator RTE (Réseau et Transport d'Électricité).[1] Our dataset ranges from the 1st of January 2012 to the 7th of June 2020 with a 30 minutes temporal resolution.

As explanatory variables we obtained national averaged temperature on the website of the French weather forecaster Météo-France.[2] We took observed temperatures instead of forecasts in order to use only open data and make the results reproducible. As our goal is to compare different forecasting strategies on the same data this choice is relevant and allows a more precise comparison as we don't include in the score the uncertainty due to physical meteorological forecast.

We train the models on historical data from the beginning of 2012 to the end of August 2019. In this paper we are interested in predicting the load during and after the COVID-19 lockdown period in France. Since the consumer behavior changed abruptly during the first month and stabilized during the second one, we divide the crisis test data in two periods. The first one ranges from March 16th to April 15th and the second one from April 16th to June 7th. Note that although the lockdown officially begun Tuesday the 17th of March 2020 at midday in France, we consider March 16th as the first day of our lockdown period as the behavior had already changed. Finally, in order to assess the suitability of the offline methods and of the ones that do not model the break we consider the pre-lockdown period between September 1st 2019 and March 15th 2020.

B. The Additive Model

The time of day is crucial for load forecasting. It doesn't appear in the following definition of the additive model because we build one model for each instant of day, i.e. we treat the 48 half-hour time series independently:

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

where at each day t,

- y_t is the electricity load for the considered instant,
- $\mathrm{DayType}_t$ is a categorical variable indicating the type of the day of the week,
- DLS_t is a binary variable indicating whether t is in summer hour or winter hour,
- Load1D and Load1W are the load of the day before and the load of the week before,
- ToY_t is the time of year whose value grows linearly from 0 on the 1st of January 00h00 to 1 on the 31st of December 23h30,
- Temp_t is the national average temperature,
- $\mathrm{Temp95}_t$ and $\mathrm{Temp99}_t$ are exponentially smoothed temperatures of factor $\alpha = 0.95$ and 0.99. E.g. for $\alpha = 0.95$ at a given instant i, $\mathrm{Temp95}_i = \alpha \mathrm{Temp95}_{i-1} + (1-\alpha)\mathrm{Temp}_i$,
- $\mathrm{TempMin99}_t$ and $\mathrm{TempMax99}_t$ are the minimal and maximal value of $\mathrm{Temp99}_t$ at the current day.

The models are trained in R using the library mgcv [32]. We use the default thin plate spline basis to represent the f_j's, except for the time of year effect f_2 for which we choose cyclic cubic splines (see [24] for a complete description of spline basis). The dimensions of the bases are usually below 5, excluding f_2 which uses a basis of dimension 20.

As previously mentioned in Section II, we suppose that ε_t is a Gaussian noise with 0 mean and constant variance. However this hypothesis is rarely true in practice and we observe an auto-correlation structure in the error. We thus propose to model it with an ARIMA model by selecting the best model with AIC criteria [33] in the family of ARIMA(p,d,q) where $p, q \leq 100$ and $d \leq 1$ (we use the R function auto.arima of R. Hyndman). In that case the forecast are performed adding GAM forecasts and the short term correction of the ARIMA models exploiting recent observations.

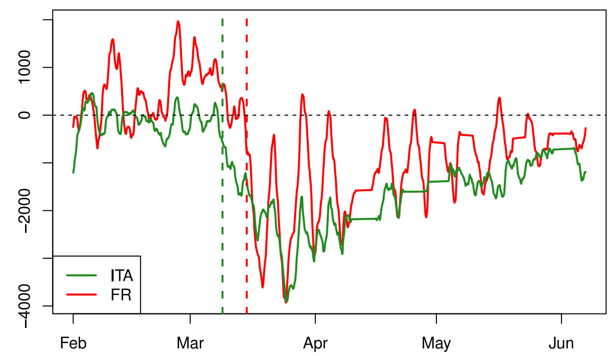


Fig. 4. Comparison of the smoothed residuals of the French and Italian GAMs in 2020. The dashed lines represent the start of the respective lockdowns.

C. Knowledge Transfer From Italy

Italy was the first country to be massively affected by the novel coronavirus in Europe. The Italian government decreed a total lockdown from the 9th of March 2020, hence 7 days before the French one. Since GAM models for both countries usually exhibit similar behavior (see Fig. 4 for a comparison of residuals) and indices such as the Oxford COVID-19 Government Response Tracker [34] show that both countries took comparable measures during the lockdown, our idea is to use this one week head-start and to adjust our GAM model for France accordingly to the changes observed in Italy. We have at our disposal data from the Italian system operator Terna[3] and meteorological data gathered through the R package Riem available from the 1st of January 2015 to the 28th of June 2020 with a 1 hour temporal resolution. For each instant, a model similar to (1) is constructed on the data on the range 2015-2019, with the main differences

[1] https://opendata.rte-france.com
[2] https://donneespubliques.meteofrance.fr/
[3] https://www.terna.it

Algorithm 2: Transfer Learning At Time Step t.

Inputs: Step size α, number of iterations K, French and Italian historical source parameters $\hat{\boldsymbol{\beta}}_0^{FR}, \hat{\boldsymbol{\beta}}_0^{IT}$, scale parameter ρ.

If GAM fine-tuned:
1) Initialize $\hat{\boldsymbol{\beta}}_t^{FR} \leftarrow \hat{\boldsymbol{\beta}}_0^{FR}$.
2) Repeat K times:
$\hat{\boldsymbol{\beta}}_t^{FR} \leftarrow \hat{\boldsymbol{\beta}}_t^{FR} - \alpha \nabla \mathcal{L}_{t-1}^{FR}(\hat{\boldsymbol{\beta}}_t^{FR})$.
3) Predict $\hat{y}_t = \hat{\boldsymbol{\beta}}_t^{FR\top} B(\boldsymbol{x}_t)$.

If GAM-δ:
1) Initialize $\hat{\boldsymbol{\beta}}_t^{IT} \leftarrow \hat{\boldsymbol{\beta}}_0^{IT}$.
2) Repeat K times:
$\hat{\boldsymbol{\beta}}_t^{IT} \leftarrow \hat{\boldsymbol{\beta}}_t^{IT} - \alpha \nabla \mathcal{L}_{t-1}^{IT}(\hat{\boldsymbol{\beta}}_t^{IT})$.
3) Set $\hat{\boldsymbol{\delta}}_t = \hat{\boldsymbol{\beta}}_t^{IT} - \hat{\boldsymbol{\beta}}_0^{IT}$, $\hat{\boldsymbol{\beta}}_t^{FR} = \hat{\boldsymbol{\beta}}_0^{FR} + \rho\, \hat{\boldsymbol{\delta}}_t$.
4) Predict $\hat{y}_t = \hat{\boldsymbol{\beta}}_t^{FR\top} B(\boldsymbol{x}_t)$.

If GAM-δ fine-tuned:
1) Perform steps 1) to 3) of **GAM-δ**, obtaining
$\hat{\boldsymbol{\beta}}_t^{FR} = \hat{\boldsymbol{\beta}}_0^{FR} + \rho\, \hat{\boldsymbol{\delta}}_t$.
2) Repeat K times:
$\hat{\boldsymbol{\beta}}_t^{FR} \leftarrow \hat{\boldsymbol{\beta}}_t^{FR} - \alpha \nabla \mathcal{L}_{t-1}^{FR}(\hat{\boldsymbol{\beta}}_t^{FR})$.
3) Predict $\hat{y}_t = \hat{\boldsymbol{\beta}}_t^{FR\top} B(\boldsymbol{x}_t)$.

being that the effects $f_3(\cdot)$ and $f_6(\cdot)$ are removed, and that $f_2(\cdot)$ is replaced by a sum of 7 effects, one for each day of the week. Then the same procedure as described in Section II-B is applied. Let $\hat{\boldsymbol{\beta}}_0^{IT}$ be the coefficient learned on the Italian source data, and $\hat{\boldsymbol{\beta}}_t^{IT}$ be the coefficients obtained by performing the aforementioned fine-tuning on Italian data ranging from the 28$^{\text{th}}$ of February up to date t (typically t could correspond to the 15$^{\text{th}}$ of March, the day before the stay-at-home order begun in France). We thus obtain $\hat{\boldsymbol{\delta}}_t = \hat{\boldsymbol{\beta}}_t^{IT} - \hat{\boldsymbol{\beta}}_0^{IT}$ the adjustment of the model on the beginning of the lockdown period. We then use $\hat{\boldsymbol{\beta}}_t^{FR} = \hat{\boldsymbol{\beta}}_0^{FR} + \rho\, \hat{\boldsymbol{\delta}}_t$ to perform the predictions for France, where ρ is a scale parameter accounting for the difference of load levels between the two countries. We refer to this model as GAM-δ. Since the ToY effect is modelized differently for the Italian model (one function per day of the week), we will not adjust the corresponding coefficients in the French model. This is further justified by the fact that in general the ToY effect is very specific to a country, and it should be learned on a whole year at least. As for the choice of ρ, making the assumption that the consumption in France and Italy are proportional with a factor ρ allows us to use the simple estimate $\hat{\rho} = \sum_t y_t^{FR} / \sum_t y_t^{IT}$ summed over a year for instance. 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. The procedures for both regular fine-tuning and GAM-δ are summarized in Algorithm 2.

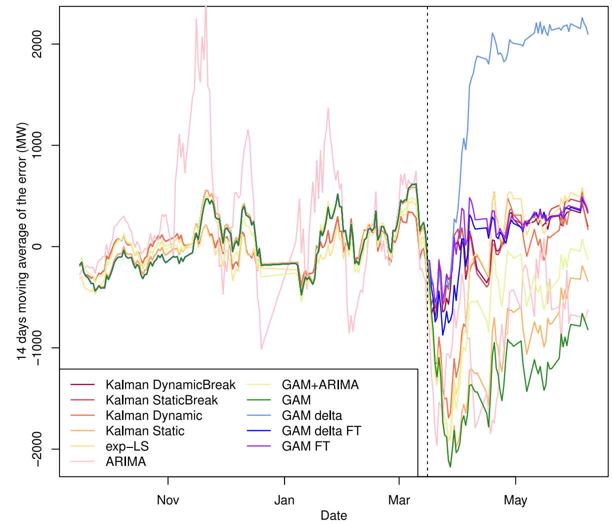


Fig. 5. Moving average of the error of the different models at 8-8:30 PM.

IV. Experiments

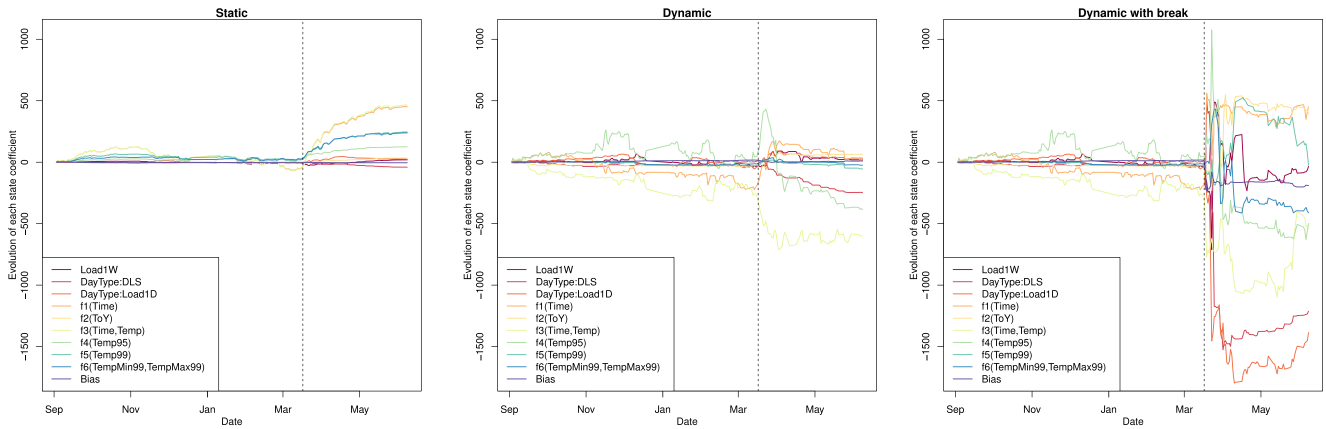
We present the application of our methods to the French dataset. While accuracy metrics are of paramount importance, we also focus on the interpretation of our results and on model behavior.

A. Model Dynamics

The moving average of the error of the different models are represented in Fig. 5. At the beginning of the lockdown all the models will tend to overpredict the load. However most of our adaptive methods quickly accommodate to the lower demand and progressively reduce their bias, notably Kalman with Dynamic Break and GAM fine-tuned. On the contrary regular GAM does not succeed in reducing the error (even with the help of an ARIMA) as it keeps overpredicting the demand. GAM-δ on the other hand is very good during the first couple of days, efficiently taking advantage of the change in patterns observed in Italy. However it quickly drifts away over time because the Italian consumption recovers faster than the French one during the second month of lockdown (see Fig. 1). However since the objective of GAM-δ is to provide an initial boost of performance during the first couple of weeks while the other models adjust, this is only a minor issue (see Section IV-B).

We test the Kalman filter in a static and a dynamic setting as described in Section II-A2. For both we assess the introduction of a break state noise covariance matrix at lockdown. The evolution of the state estimate $\hat{\boldsymbol{\theta}}_t$ is displayed in Fig. 6 for different settings.

In the static setting the Kalman filter optimizes a state which is assumed to be constant, hence explaining a slow evolution compared to the faster changes of the dynamic one. However both variants change faster during lockdown than they did before. As expected the introduction of a break covariance matrix at the beginning of the lockdown allows the model to adapt much faster.



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

Fig. 6. Evolution of the state coefficients for different Kalman variants at 8-8:30 PM (subtracting the coefficients on September 1st 2019).

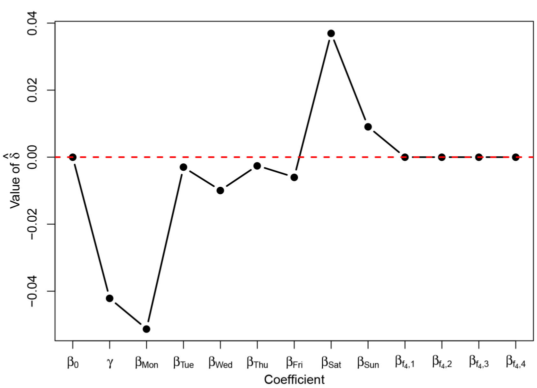


Fig. 7. Value of $\hat{\boldsymbol{\delta}}_t$ fine-tuned on the period 16/03-15/04 at 8-8:30 PM.

The model dynamics can be analysed for the fine-tuning too. For GAM-δ the only coefficients of $\hat{\boldsymbol{\delta}}_t$ with a significant evolution after fine-tuning are the ones pertaining to the lagged load (γ for Load1W and $\beta_i, i = 1..7$ for Load1D) and have been represented in Fig. 7. The other ones are zero and have been omitted for clarity. The coefficients of the working days drop, especially the Monday, whereas the ones of the weekend increase, notably Saturday. It can be interpreted as follows: the historical model learned a certain transition between the different days of the week. With the lockdown now all the days are similar and close to a Saturday, which has a lower demand than Monday and thus the associated coefficient plummets. The coefficient of Saturday soars because the demand on Fridays is now much lower than it used to be and that daily profiles are similar. Finally since during the first weeks the electricity demand progressively decreases (see Fig. 1) the coefficient of γ drops as well.

B. Aggregation

We proposed 2 load forecasting models (ARIMA, GAM) and different variants to adapt them to the lockdown period (exp-LS, Kalman adaptation, transfer learning) leading to 11 candidates which we call experts in the following. A natural approach is then to aggregate them in a single forecast which will take benefit of the best one in function of time. This is the main idea behind online aggregation methods which have already demonstrated their benefits in the field of electricity load forecasting (see [35], [36]). Since Fig. 2 shows the convergence of the daily profiles towards the Saturday shape, this as well as [16] motivates adding another expert named GAM Saturday, whose prediction is made by the regular GAM as if every day was a Saturday.

We recall briefly the main principles of the online aggregation approach and refer the interested reader to [37] for a complete presentation. A bounded sequence of observations (here half-hourly French electricity consumption) $y_1, \ldots, y_n \in [0, B]$ is observed (B being an unknown constant). We have access to a set of N experts producing forecasts of the sequence at each instant t based on past values of y. After that, aggregation is computed step by step: $\hat{y}_t = \sum_{j=1}^{N} \hat{p}_{j,t} \hat{y}_t^j$ where the weights are updated according to past performances of each experts which are measured with a convex loss function. In accordance to the RMSE criterion used in our case study we consider the square loss $\ell_t(x) = (y_t - x)^2$. At time t expert e suffers loss $\ell_t(\hat{y}_t^e) = (y_t - \hat{y}_t^e)^2$ and the aggregation $\ell_t(\hat{y}_t) = (y_t - \hat{y}_t)^2$. We call Oracle an optimal forecast which is unknown in advance and usually hard to beat in terms of forecasting accuracy (see [37]). We denote it by $\hat{y}_t^*$. For example, it could be the best fixed convex aggregation or the best expert (best w.r.t the entire time interval performance, of course unknown a priori). The goal of aggregation algorithms is to minimise the total loss $\sum_{t=1}^{T} (y_t - \hat{y}_t)^2$ that can be expressed:

$$\frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \quad \triangleq \quad \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t^*)^2 + R_T \,,$$

where R_T is the so-called regret term, it is the error suffered by our algorithm relatively to the error of the oracle (see again [37]). The aim is thus to propose algorithms that, regarding competitive oracles, achieve low regrets. In our study we use the ML-Poly algorithm of [38], implemented in the R package `opera` [39]

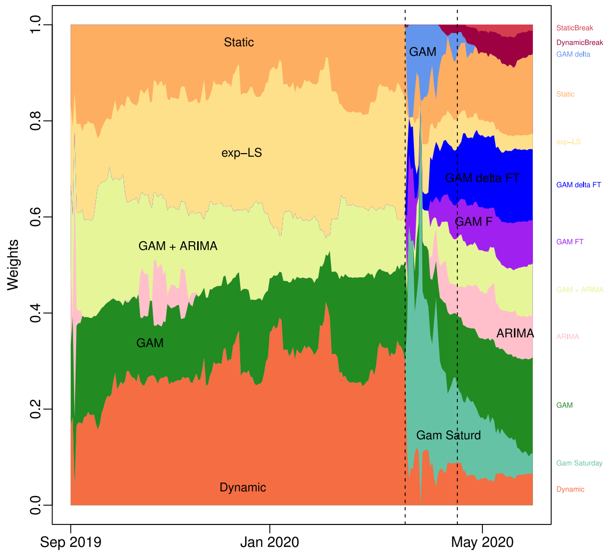


Fig. 8. Weights attributed to each expert by the aggregation method at 8-8:30 PM. Dashed lines split the test sets.

Algorithm 3: ML-Poly.

Initialization: $\hat{\boldsymbol{p}}_1 = (1/N, \ldots, 1/N)$ and $\boldsymbol{R}_0 = (0, \ldots, 0)$;
Recursion: at each time step $t = 1, 2, \ldots$

- Pick the learning rates:
  $\eta_{e,t-1} = 1/(1 + \sum_{s=1}^{t-1} (\ell_s(\hat{y}_s) - \ell_s(y_s^e))^2)$.
- Compute the weights $\hat{\boldsymbol{p}}_t$:
  $\hat{p}_{e,t} = \eta_{e,t-1}(R_{e,t-1})_+ \,/\, \boldsymbol{\eta}_{t-1} \cdot (\boldsymbol{R}_{t-1})_+$
  where $\boldsymbol{R}_+$ is the non-negative parts of $\boldsymbol{R}$.
- Output prediction $\hat{y}_t = \sum_{e=1}^{N} \hat{p}_{e,t} \hat{y}_t^e$.
- For each expert e update the regret:
  $R_{e,t} = R_{e,t-1} + \ell_t(\hat{y}_t) - \ell_t(y_t^e)$,
  $\boldsymbol{R}_t = (R_{1,t}, \ldots, R_{N,t})$.

---

and successfully used for load and price forecasting in [35], [40]. It is described in Algorithm 3. An expert who has a high regret, which means that he suffers a higher loss than the aggregation, will receive less weight for the next round. The time varying learning rate $\eta_{e,t}$ could be seen as a vector step size parameter of gradient descent varying with time so that no parameter tuning is needed.

Finally a few experts are introduced in the aggregation only at lockdown. Indeed before lockdown the transfer learning experts don't make sense (there is no target data), the Kalman experts modelling the break coincide with the other ones, and the expert GAM Saturday was only introduced for the lockdown period. These specialized experts are added to the aggregation at the lockdown period with a uniform weight $(1/12)$, and the experts present before share the rest of the weight proportionally to their previous weight [41].

The evolution of the weights of the experts over time is given in Fig. 8. It gives insight on which predictions are the most useful in the aggregation at a given time. The lockdown acts as a break and causes a significant shift in the weights distribution. As such, GAM Saturday immediately takes a large weight: this is due to the aforementioned resemblance between the daily profiles during the lockdown with Saturdays. Moreover, this expert predicts a lower consumption than reality, compensating for the overestimation of the other experts at the beginning of the lockdown. GAM-δ also has high importance, as it has knowledge of what happened in Italy and thus suits the new patterns of load demand in France. For instance on the two first days of lockdown (16 and 17th of March) GAM-δ yields 1984 MW of RMSE, compared to 2674 and 3005 for Kalman Dynamic Break and regular GAM respectively. However their importance dwindle with time as the adaptive Kalman and fine-tuning methods have seen enough data and have become more competitive.

C. *Numerical Results*

As usual in electricity load forecasting, the performance metrics are the root mean squared error (in MW) and the mean absolute percentage error (in %):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{t=1}^{n} (y_t - \hat{y}_t)^2}\,,$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{t=1}^{n} \Big|\frac{y_t - \hat{y}_t}{y_t}\Big|\,,$$

where n is the number of instances in the test set.

We display the numerical performance of our methods in Table I. We observe that any of our methods have lower RMSE or MAPE than GAM + ARIMA on both post-COVID test sets. As expected, the Kalman Dynamic with break yields the best results for the two error metrics during the first part of the lockdown period but the fine-tuned methods are very close to it. Similarly, the two break approaches are the best ones after the lockdown. The additional benefits brought by expert aggregation is emphasized by the two last rows. The algorithm manages to take advantage of the individual specificities of the different predictions, leading to further error reduction. While individually poor, the inclusion of GAM Saturday in the mixture brings significant improvement for the first testing period (see end of Section IV-B) because it compensates for the overestimation of the demand at the beginning of the lockdown.

The significativity of our results was assessed with two statistical tests: a Diebold-Mariano (DM) test [42] and a Wilcoxon test as proposed in [43], both on the absolute error. For two methods A and B it allows to test the null hypothesis that method B outperforms or is equivalent to method A, against the alternative hypothesis that method A outperforms method B. In Table II we display the results of the tests for the most relevant forecasting models at the significance level 0.01. At each row i and column j we display the p-values of Wilcoxon test in blue and of the Diebold-Mariano test in purple, the alternative hypothesis is "method i outperforms j". We use the symbol ε when the p-value is below 0.01 and otherwise we give a 0.01 approximation. For clarity we consider only the best non-adaptive method and selected adaptive ones, and we order them according to the performance on the last test set. These tests confirm that on both post-COVID test sets, the improvement brought by our adaptive procedures on an ARIMA correction of

TABLE I
NUMERICAL PERFORMANCE IN MAPE (%) AND RMSE (MW)

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**
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	-	**2.54 %, 1636 MW**	**1.49 %, 766 MW**

TABLE II
WILCOXON TEST AND DIEBOLD-MARIANO TEST ON THE ABSOLUTE ERROR ON THE LAST TWO TEST SETS. ε FOR P-VALUE BELOW 0.01

2020/03/16-2020/04/15	1	2	3	4	5	6
1. Aggregation	**1 1**	**0.44** ε	ε ε	ε ε	ε ε	ε ε
2. Kalman Dynamic Break	**0.56 1**	**1 1**	**0.12 0.06**	ε ε	ε ε	ε ε
3. GAM δ Fine-Tuned	**1 1**	**0.88 0.94**	**1 1**	ε ε	ε ε	ε ε
4. exp-LS	**1 1**	**1 1**	**1 1**	**1 1**	ε ε	**0.08 0.28**
5. GAM + ARIMA	**1 1**	**1 1**	**1 1**	**1 1**	**1 1**	**0.99 1**
6. GAM δ	**1 1**	**1 1**	**1 1**	**0.92 0.72**	**0.01** ε	**1 1**
2020/04/16-2020/06/07	1	2	3	4	5	6
1. Aggregation	**1 1**	ε ε	ε ε	ε ε	ε ε	ε ε
2. Kalman Dynamic Break	**1 1**	**1 1**	ε ε	ε ε	ε ε	ε ε
3. Gam δ Fine-Tuned	**1 1**	**1 1**	**1 1**	ε ε	ε ε	ε ε
4. exp-LS	**1 1**	**1 1**	**1 1**	**1 1**	ε ε	ε ε
5. GAM + ARIMA	**1 1**	**1 1**	**1 1**	**1 1**	**1 1**	ε ε
6. GAM δ	**1 1**	**1 1**	**1 1**	**1 1**	**1 1**	**1 1**

the GAM is statistically significant, and so is the improvement of the aggregation compared to any of our method. Results coincide for the two tests, ergo consolidating our results even further.

V. CONCLUSION

In this paper, we proposed two novel approaches of adaptive generalized additive models to improve load forecast during the COVID-19 pandemic, one relying on Kalman filtering and the other on transfer learning with GAM fine-tuning. We showed that Kalman filtering approaches can be significantly improved by re-initalizing the online update at the beginning of the lockdown period (Break approach). This helps the Kalman filter to adapt quickly to a change in the data and update the forecasts taking advantage of recent observations. Transfer learning was successfully adapted to this problem in two ways: we fine-tuned a GAM learned before the COVID-19 crisis on the lockdown period, and we transferred information from Italian data to French data. We illustrated the benefits of the transfer from Italy at the beginning of the lockdown, as well as the efficiency of adaptive methods to significantly improve predictions, all without relying on the inclusion of new exogenous features. As all these new approaches have time varying performances (the best forecasts vary with time), we proposed to use online expert aggregation to enhance results ever further.

While in this paper we focused on adapting GAM, the proposed framework can be applied to other approaches. The use of neural networks for instance will soon be investigated. We also plan to include exogenous information as mobility data proposed in [17], macro-economic indicators or data from social media such as Twitter. Regarding load data, exploiting regional data could be relevant as the propagation of the pandemic and its impact on consumption was different depending on the region in France and Italy. The inclusion of more countries could be useful as well. For these next steps, transfer approaches will be of fundamental importance but also adaptive ones, as the effects of exogenous variables are likely to vary with time or even be added at some point.

REFERENCES

[1] D. Bunn and E. D. Farmer, *Comparative Models for Electrical Load Forecasting*. New York, NY,USA: Wiley, 1985.

[2] S.-J. Huang and K.-R. Shih, "Short-term load forecasting via arma model identification including non-Gaussian process considerations," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 673–679, May 2003.

[3] A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi, "A prediction interval for a function-valued forecast model: Application to load forecasting," *Int. J. Forecasting*, vol. 32, no. 3, pp. 939–947, 2016.

[4] H. Cho, Y. Goude, X. Brossat, and Q. Yao, "Modeling and forecasting daily electricity load curves: A hybrid approach," *J. Amer. Stat. Assoc.*, vol. 108, no. 501, pp. 7–21, 2013.

[5] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *Int. J. Forecasting*, vol. 30, no. 2, pp. 357–363, 2014.

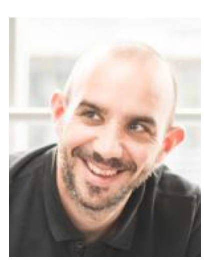
[6] J. R. Lloyd, "Gefcom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes," *Int. J. Forecasting*, vol. 30, no. 2, pp. 369–374, 2014.

[7] D. C. Park, M. El-Sharkawi, R. Marks, L. Atlas, and M. Damborg, "Electric load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 442–449, 1991.

[8] S. Ryu, J. Noh, and H. Kim, "Deep neural network based demand side short term load forecasting," *Energies*, vol. 10, no. 1, 2017, Art. no. 3.
[9] A. Pierrot and Y. Goude, "Short-term electricity load forecasting with generalized additive models," in *Proc. IEEE Intell. Syst. Appl. Power Syst. Conf.*, 2011, pp. 410–415.
[10] S. Fan and R. J. Hyndman, "Forecasting electricity demand in australian national electricity market," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2012, pp. 1–4.
[11] Y. Goude, R. Nedellec, and N. Kong, "Local short and middle term electricity load forecasting with semi-parametric additive models," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 440–446, Jan. 2014.
[12] M. Fasiolo, S. N. Wood, M. Zaffran, R. Nedellec, and Y. Goude, "Fast calibrated additive quantile regression," *J. Amer. Stat. Assoc.*, pp. 1–11, 2020.
[13] S. N. Wood, Y. Goude, and S. Shaw, "Generalized additive models for large data sets," *J. Roy. Stat. Soc.: Ser. C: Appl. Statist.*, vol. 64, pp. 139–155, 2015.
[14] M. Narajewski and F. Ziel, "Changes in electricity demand pattern in europe due to covid-19 shutdowns," 2020, *arXiv:2004.14864*.
[15] IEA, "Year-on-year change in weekly electricity demand, weather corrected, in selected countries," 2020. [Online]. Available: https://www.iea.org/data-and-statistics/charts/year-on-year-change-in-weekly-electricity-demand-weather-corrected-in-selected-countries-january-december-2020
[16] K. Nagbe, "France short-term load demand forecasting using a functional state space adaptative model: Case of covid-19 lockdown period," 2020.
[17] Y. Chen, W. Yang, and B. Zhang, "Using mobility for electrical load forecasting during the covid-19 pandemic," 2020, *arXiv:2006.08826*.
[18] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME - J. Basic Eng.*, vol. 82, pp. 35–45, 1960.
[19] A. Harvey and S. J. Koopman, "Forecasting hourly electricity demand using time-varying splines," *J. Amer. Stat. Assoc.*, vol. 88, no. 424, pp. 1228–1236, 1993.
[20] V. Dordonnat, S. J. Koopman, and M. Ooms, "Dynamic factors in state-space models for hourly electricity load signal decomposition and forecasting," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2009, pp. 1–8.
[21] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[22] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big data*, vol. 3, no. 1, 2016, Art. no. 9.
[23] L. Cai, H. Wen, J. Gu, J. Ma, and Z. Jin, "Forecasting customers' response to incentives during peak periods: A transfer learning approach," *Int. Trans. Elect. Energy Syst.*, vol. 30, no. 7, 2020, Art. no. e 12251.
[24] S. N. Wood, *Generalized Additive Models: An Introduction With R*. Boca Raton, FL, USA: CRC Press, 2017.
[25] A. Ba, M. Sinn, Y. Goude, and P. Pompey, "Adaptive learning of smoothing functions: Application to electricity load forecasting," in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 2510–2518.
[26] P. J. Brockwell, R. A. Davis, and M. V. Calder, *Introduction to Time Series and Forecasting*. Berlin, Germany: Springer, vol. 2, 2002.
[27] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*. London, U.K.: Oxford Univ. Press, 2012.
[28] L. Fahrmeir and G. Tutz, *Multivariate Statistical Modelling Based on Generalized Linear Models*. NewYork, NY, USA: Springer Science & Business Media, 2013.
[29] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
[30] N. Laptev, J. Yu, and R. Rajagopal, "Reconstruction and regression loss for time-series transfer learning," in *Proc. Special Interest Group Knowl. Discov. Data Mining 4th Workshop Mining Learn. Time Ser.*, London, U.K., vol. 20, 2018.
[31] D. Obst, B. Ghattas, J. Cugliari, G. Oppenheim, S. Claudel, and Y. Goude, "Transfer learning for linear regression: A statistical test of gain," 2021, *arXiv:2102.09504*.
[32] S. Wood and M. S. Wood, "Package 'MGCV'," *R package version*, vol. 1, p. 29, 2015.
[33] H. Akaike, "Time series analysis and control through parametric models," in *Appl. Time Ser. Anal. I*, 1978, pp. 1–23.
[34] T. Hale *et al.*, "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)," *Nature Human Behaviour*, pp. 1–10, 2021.
[35] P. Gaillard and Y. Goude, "Forecasting electricity consumption by aggregating experts; How to design a good set of experts," in *Proc. Model. Stochastic Learn. Forecasting High Dimens.*, 2015, pp. 95–115.
[36] B. Goehry, Y. Goude, P. Massart, and J.-M. Poggi, "Aggregation of multi-scale experts for bottom-up load forecasting," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 1895–1904, May 2019.
[37] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. New York, NY, USA: Cambridge Univ. Press, 2006.
[38] P. Gaillard, G. Stoltz, and T. Van Erven, "A second-order bound with excess losses," in *Proc. Conf. Learn. Theory*, 2014, pp. 176–196.
[39] P. Gaillard and Y. Goude, "opera: Online prediction by expert aggregation," vol. 1, 2016. [Online]. Available: https://CRAN. R-project. org/package= opera. r package version
[40] P. Gaillard, Y. Goude, and R. Nedellec, "Additive models and robust aggregation for gefcom2014 probabilistic electric load and electricity price forecasting," *Int. J. Forecasting*, vol. 32, no. 3, pp. 1038–1050, 2016.
[41] M. Devaine, P. Gaillard, Y. Goude, and G. Stoltz, "Forecasting electricity consumption by aggregating specialized experts," *Mach. Learn.*, vol. 90, no. 2, pp. 231–260, 2013.
[42] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 20, no. 1, pp. 134–144, 2002.
[43] Z. Zhang, S. Ding, and Y. Sun, "A support vector regression model hybridized with chaotic Krill herd algorithm and empirical mode decomposition for regression task," *Neurocomputing*, vol. 410, pp. 185–201, 2020.

David Obst is currently working toward the Ph.D. degree with EDF R&D in conjunction, Aix-Marseille Université, Marseille, France. His research interests focuses on the use of transfer learning and textual data for time series forecasting.

Joseph de Vilmarest is currently working toward the Ph.D. degree with LPSM, Sorbonne Université, Paris, France and with EDF R&D. His research interests include state space models, Bayesian methods, and time series forecasting.



Yannig Goude received the Ph.D. degree in statistics and probability from Université Paris-Sud Orsay, Orsay, France, in 2008. He is currently a Senior Researcher in the field of Data Science with EDF R&D and an Associate Professor with Université Paris-Sud Orsay, France. His research interests include time series forecasting for electricity markets, time series analysis, nonparametric models, and aggregation of experts.