Forecasting Intra Day Load Curves Using Sparse Functional Regression

Mathilde Mougeot, Dominique Picard, Vincent Lefieux, and Laurence Maillard-Teyssier

Abstract In this paper we provide a prediction method, *the prediction box*, based on a sparse learning process elaborated on very high dimensional information, which will be able to include new – potentially high dimensional – influential variables and adapt to different contexts of prediction. We elaborate and test this method in the setting of predicting the national French intra day load curve, over a period of time of 7 years on a large data basis including daily French electrical consumptions as well as many meteorological inputs, calendar statements and functional dictionaries. The prediction box incorporates a huge contextual information coming from the past, organizes it in a manageable way through the construction of a *smart* encyclopedia of *scenarios*, provides experts elaborating strategies of prediction by comparing the day at hand to referring scenarios extracted from the encyclopedia, and then harmonizes the different experts. More precisely, the prediction box is built using successive learning procedures: elaboration of a data base of historical scenarios organized on a high dimensional and functional learning of the intra day load curves, construction of expert forecasters using a retrieval information task among the scenarios, final aggregation of the experts. The results on the national French intra day load curves strongly show the benefits of using a sparse functional model to forecast the electricity consumption. They also appear to meet quite well with the business knowledge of consumption forecasters and even shed new lights on the domain.

M. Mougeot (✉) • D. Picard
Université Paris-Diderot, CNRS-LPMA, UFR de Mathématiques, 75013 Paris, France
e-mail: mougeot@math.univ-paris-diderot.fr; picard@math.univ-paris-diderot.fr

V. Lefieux
RTE-EPT & UPMC-ISUP, 92919 La Défense Cedex, France
e-mail: vincent.lefieux@rte-france.com

L. Maillard-Teyssier
RTE-R&D-I, 78000 Versailles, France
e-mail: laurence.maillard-teyssier@rte-france.com

A. Antoniadis et al. (eds.), *Modeling and Stochastic Learning for Forecasting in High Dimensions*, Lecture Notes in Statistics 217,
DOI 10.1007/978-3-319-18732-7_9

1 Prediction Box: Forecasting the Electrical Consumption

This paper is the result of a cooperation between industrial and academic research. RTE, the French electricity transmission system operator, is responsible for operating, maintaining and developing the high and extra high voltage network. RTE is required to guarantee the security of supply, so anticipating French electricity demand helps to ensure the balance between generation and consumption at all times, and directly influences the reliability of the power system.

Demand forecasts are carried out for several different timeframes: for the long-term, in the form of the Generation Adequacy Report or network development studies, for the medium-term (annual, monthly and weekly forecasts) and lastly on a day-ahead basis.

From a short term point of view, electricity demand fluctuates depending on cycles (annual, weekly and daily), on temperature and cloud cover (to take into account variations in the outside temperature affecting the use of heating equipment in winter and air-conditioning in summer), and on other factors such as economic activity (e.g. holiday periods), demand response offers or daylight saving time changes. Note that the French load curve is very sensitive to temperatures, it contributes to half of the European thermo-sensitivity.

Today RTE uses a complex nonlinear parametric regression model with around one thousand coefficients estimated twice a year, and also a SARIMA model. These decision-making tools cannot predict exceptional events which may disrupt the demand profile (heavy snow fall, sporting events, strikes), and final day-ahead forecasts are provided by the French national dispatchers (see: http://www.rte-france.com/en/sustainable-development/eco2mix/electricity-demand).

If this process currently provides good forecasts, the context of the smart grids and the energy transition will lead to more variability in the load curve. Moreover, the aim is not only to obtain a low mean error but also to avoid big forecast errors which have a direct influence on the reliability of the power system.

Taking into account new explanatory variables (e.g. wind, new tariffs, electricity prices), economic uncertainties (e.g. economic crisis), new innovative heating systems (e.g. heat pumps) requires to work with more adaptive and dynamic models. Many models and approaches have already been considered, from the robust SARIMA [5, 13] to the semi-parametric model MAVE [10] or functional regression using wavelets [2, 3].

Referring to this context, in this paper, we address the following program:

(i) Construct a prediction method, *the prediction box*, based on a sparse learning process including high dimensional information, which will be able to include new – potentially also high dimensional – influential variables or to adapt to different settings of prediction in terms of time ahead (one day ahead, 48-h ahead, medium-term) or geographical context (European, or at the opposite regional or even more local) and eventually to more general situations of forecasting.

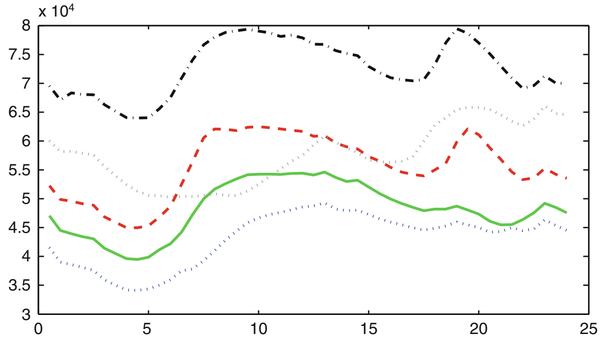

Fig. 1 Intraday load curves for various days. 2010-02-03 winter: *black dashed dot line*, 2010-05-21 spring: *red dashed line*, 2009-10-23 autumn: *green solid line*, 2010-08-19 summer: *blue dot line*, 2010-01-01 public day: *gray dot line*

(ii) Elaborate and test this method in the setting of predicting the national French intraday load curve, over a period of time of 7 years on a large data basis including daily electrical consumption as well as many meteorological inputs, calendar statements and functional dictionaries, described in the next subsection.

Figure 1 illustrates the observed dissimilarity between some daily consumption signals. For instance electrical consumption is mostly higher and more *active* in winter than in summer and is characterized by two large peaks of consumption. However some winter public days may show weak consumption with unusual pattern, and spring days can reach high consumption with also specific features characterized by a unique peak of consumption.

A crucial step in the forecasting process is the modeling. It is commonly admitted that many variables are influential for the prediction in this context. On the other hand, it is well known that a model relying on a small number of well chosen predictors is more robust and efficient than a model without variable pre-selection. The challenge here is then for each day to produce a small number of predictors, after considering all the variables which can be potentially significant.

The *prediction box* will provide three drawers of unequal sizes using different learning procedures at each different scale:

(a) The first drawer contains a *smart* encyclopedia of *scenarios* coming from the observed past. A smart encyclopedia is a very large but very well organized structure. For each day of the past sample, the encyclopedia provides the *background* of the day measured by a large (but manageable) number of significant explanatory variables. It also contains, associated to the *background* of the day, a sparse approximation of the consumption of the day (using much fewer explanatory variables than the number of initial variables).
(b) The second drawer contains a bunch of experts. Each of them provides a strategy of prediction for the load curve after consulting the referred encyclopedia. Each expert essentially bases its strategy on comparing the day at hand to referring scenarios extracted from the backgrounds of the past. This step allows to find a

day in the past which is closest (according to the expert) to the day at hand. The prediction then uses the sparse approximation of this closest day.

(c) The final action of the box will be to harmonize the experts using an aggregation process.

In the following section we describe the data basis. Section 3 is devoted to the construction of the encyclopedia. This section is crucial and will especially describe the choice of variables for the backgrounds as well as the sparse approximation. Section 4 details the experts, their respective performances and the aggregation process. The last section is devoted to analyze the results of the prediction box as well as the perspectives of the method.

2 The Data Basis

2.1 Electrical Consumption

The French national electrical consumption has been recorded every half hour from January 1st, 2003 to August 31st, 2010 and stored in a database. We focus our study on daily recording signals. For this period of time, the global consumption signal is split into $N_{obs} = 2{,}800$ sub signals $(Y_1, \ldots, Y_t, \ldots, Y_{N_{obs}})$. $Y_t \in R^n$ defines the intra day load curve for the tth day of size $n = 48$. These intra day signals will constitute our data basis to approximate and then to forecast the daily consumption.

Figure 2 shows a week of electrical consumption defined by seven successive intra day curves.

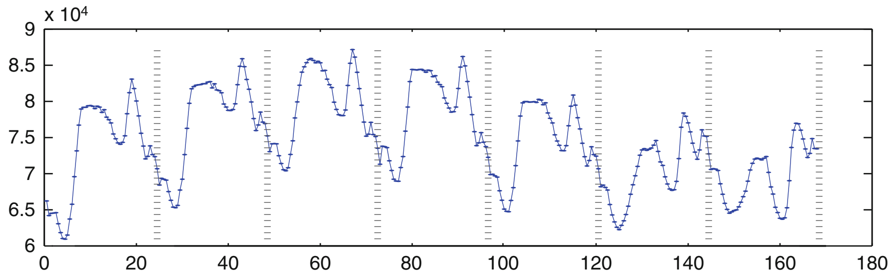

Fig. 2 Electrical consumption week from Monday January 25th to Sunday January 31st 2010 regrouping seven successive intra day load curves of size $n = 48$

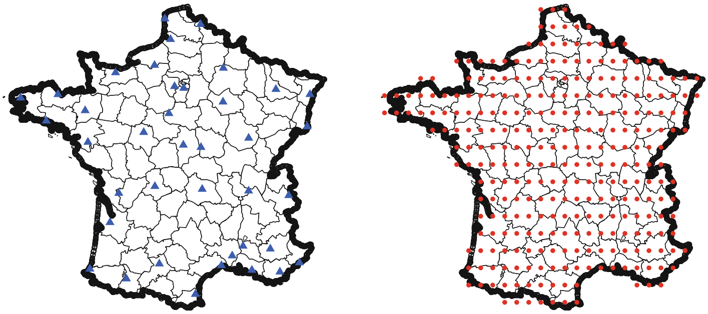

Fig. 3 Temperature and cloud cover (*left*) measurement stations. Wind strength network available points (*right*)

2.2 Meteorological Inputs

For this study, available meteorological inputs are recorded each half hour on the same period of time, from January 1st, 2003 to August 31st, 2010:

Temperature: T^k for $k = 1, \ldots, 39$ denotes the temperatures measured in 39 weather stations scattered all over the French territory as indicated in Fig. 3 (left).
Cloud Cover: N^k for $k = 1, \ldots, 39$ is an indicator of the cloud cover which is also measured in the same 39 weather stations. The cloud cover is a fraction between 0 (free of clouds sky) and 80 tenths of octas (completely clouded sky). Cloud cover are nowadays built on satellite based observations on the same meteorological stations than temperature.
Wind: $W^{k'}$ for $k' = 1, \ldots, 293$ denotes the 100 m wind speed analyses available at the 293 network points scattered all over the French territory (see Fig. 3, right).

As the electrical consumption, the meteorological inputs (temperature, cloud Cover and wind) are sampled each half hour and available for the same time period. Every day, for each meteorological input and weather spot, a $n = 48$ signal can be extracted as shown in Fig. 4.

T_t^k (resp. N_t^k, $W_t^{k'}$) denotes the daily temperature (reps. cloud cover, wind) for day t, $1 \leq t \leq 2{,}800$, and station k with $1 \leq k \leq 39$ or network point k' with $1 \leq k' \leq 293$.

Figure 4 illustrates the variability of weather conditions in France. Temperature, cloud cover and wind signals are chosen in three different meteorological spots localized in West (Brest), North (Lille) and South (Marseille) of France. In Marseille, large variations of temperature can be observed during this day, with a stationary high cloud cover and an increasing wind. On the opposite during the same day, stationary temperature and wind can be observed in Brest with a decreasing cloud cover. All these meteorological factors are known to have an impact on the electrical regional consumption which has been established for the French PACA and Bretagne regions for instance.

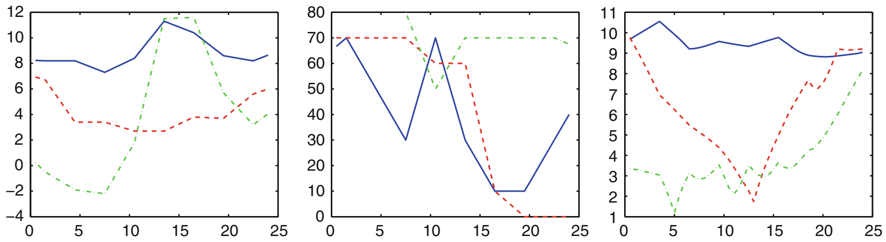

Fig. 4 Temperature (*left*), cloud cover (*middle*) and wind (*right*) intra day signal for the 3rd February 2010 in Brest (*blue line*), Lille (*red line*) and Marseille (*green*) cities

In this study, a total of 371 (= 2×39 + 293) raw meteorological variables are available. Both types of data (surface weather points and grid points) are used which constitutes a new way of integrating meteorologic data into a load curve modeling.

2.3 Calendar Statements

As illustrated in Figs. 1 and 2, the intra day load curves are quite different, depending on the day and on the season. Qualitative variables are introduced in order to characterize days and seasons.

Variable D takes seven modalities characterizing the type of day {1:Monday, …, 7:Sunday }.

As in [22], C is a qualitative variable, taking one of the 20 modalities depending on the 5 groups of days (Monday, Friday, Saturday, Sunday and the others) subdivided by the four seasons Winter, Spring, Summer and Autumn.

D_t and C_t describe the modalities of both variables for day t.

3 Building the *Smart* Encyclopedia

Electrical consumption, meteorological inputs and calendar statements contributes to the elaboration of the *smart encyclopedia*, recording the historical data base.

The encyclopedia proposed in this work has been considered as a *smart* encyclopedia since part of its elements are built using a learning process. In particular, the encyclopedia contains sparse approximations of the intra day load curves which can be considered as a ‘clever’ and significant representation of the consumption signals; the choice of the dictionary being a key point.

Parts of the construction of the *encyclopedia* use different steps interesting by themselves which are described in [17]. We just recall here the principal ones and refer to [17] for more details.

3.1 Patterns of Consumption

It is well known that intra day load curves can be explained using two types of variables: on one hand specific patterns of consumption (also called endogenous variables), on the other hand meteorological variables (also called exogenous variables) [6, 18, 19]. Patterns of consumption are usually built using calendar information and it is important to address the problem of determining which calendar information will be used to best represent the curves, with the serious issue that some choices can lead to representation which are highly correlated with meteorological variables and very often will disappear in sparse representations.

In most applications, in order to integrate calendar information, the set of days is split on deterministic statements. Taylor [22] uses a partition of size 20 already mentioned in Sect. 2.3. To study the Spanish consumption, [14] uses Kohonen maps to build adaptive groups of consumption.

Our point of view is slightly different and consists in providing typical *patterns of consumption* using a three step pre-processing: we first provide a sparse modeling of the profiles of daily electrical consumption as functions of the time. Then, using the sparse representation of each load curve, clusters of consumption are defined. A final interpretation of the clusters yields *typical* profiles, the *group centroids* or it patterns of consumption that will be time variables entering into the final model, in addition to meteorological variables. More precisely,

1. The first task is defined by the compression of the intraday load curves Y_t using a nonparametric regression on a dictionary of functions of the time variable, with the help of the sparse algorithm LOLA described in Table 1 [12, 15, 17].

   In other words, the intraday signals Y_t are treated as functions of the time [20] and sparsely represented in a dictionary, which has to be well adapted to produce a full reduction of the problem. A combination of Fourier basis and Haar basis has been chosen as dictionary for this task.

Table 1 Description of LOLA Algorithm

Step 1: Selection by thresholding
A first thresholding procedure allows to reduce the dimensionality of the problem in a rather crude way by a simple inspection of the empirical correlations between the signal and each element of the dictionary.
This first threshold is data driven and is chosen adaptively [15].
Step 2: OLS
Ordinary Least Square method is then used on the linear sub-model obtained by considering the variables retained after the first step.
Step 3: Denoising by thresholding
A second thresholding is performed on the estimators of the parameters of the sub-model.
This second step is more refined and corresponds to a denoising phase of the algorithm.
As the first one, the second threshold is also data driven and is chosen adaptively [15].

Table 2 Groups, 1 . . . 8, are defined using a calendar interpretation of clusters from Monday (day 1) to Sunday (day 7) and from January (month 1) to December (month 12) computed form January 1st to August 31st [17]

	Months											
Days	1	2	3	4	5	6	7	8	9	10	11	12
1	7	8	5	3	3	3	3	1	3	3	5	7
2	7	8	5	3	3	3	3	1	3	3	5	7
3	7	8	5	3	3	3	3	1	3	3	5	7
4	7	8	5	3	3	3	3	1	3	3	5	7
5	7	8	5	3	3	3	3	1	3	3	5	7
6	6	8	4	4	2	2	2	2	2	2	4	6
7	6	6	4	4	2	2	2	2	2	2	4	6

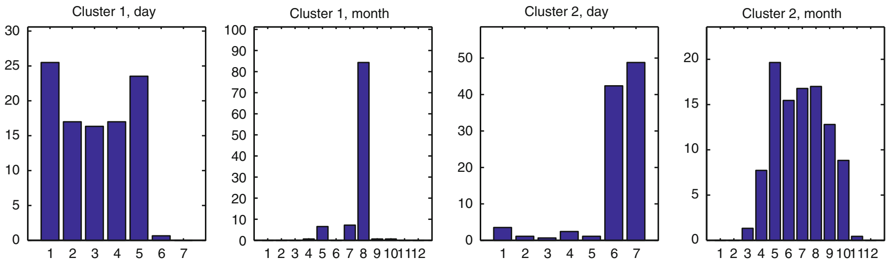

Fig. 5 Illustration of the calendar repartition of days and months for the two clusters 1, 2

2. The second task consists in clustering the previous sparse representations of the signals. The K-means algorithm is here chosen for its simplicity. An additional investigation [17] shows that the clustering end up with 8 different groups, as described in the following Table 2.
3. The last step defines the *group centroid* variables or *patterns of consumption* as the mean signal inside each group. This yields eight signals $S_1, \ldots, S_8$ summarizing the typical behavior of each cluster.

An analysis is then performed in order to retrieve a correspondence between clusters and calendar statements. For an illustration of the different calendar statements extracted between the clusters, Fig. 5 provides, as an example, the occurrences observed between days and months, for two clusters. For cluster 1, a large majority of load curves are week-days (1–5) for the month of August (8). For cluster 2, a large majority of load curves are week-ends (6–7) for months from May to October (5–10). It should be noted that, at this stage, no specific treatments have been done for bank holidays or other special days. Interpretation for all clusters is available in [17].

The *pattern variable* G is a variable taking eight (functional) modalities and assigning each day t to the center of the group where it belongs: $S_{g(t)}$, where $g(t)$ is simply the labeling function of the calendar situation of t, according to Table 2.

During our preliminary study an additional pattern variable emerged and appeared to be as well a key endogenous variable, defined by the intra day load curve, Y_{t-7}, recorded 1 week before. This signal can be considered as a general trend for Y_t and provides also indirectly some calendar information related to the type of day (Sunday, Monday, ...) and seasons.

The endogenous variables are then defined by these two patterns and will represent the consumption signal as a function of the time, denoted in the sequel by $P_t = [G_t\, Y_{t-7}]$.

3.2 Meteorological Variables

In this study, the target variable is the French national electrical consumption, which is impacted by all the meteorological conditions of the French territory, but, of course, with different contributions regarding each region. Inside each group of temperatures $\{T^k, 1 \leq k \leq 39\}$, cloud cover $\{N^k, 1 \leq k \leq 39\}$ or winds $\{W^{k'}, 1 \leq k' \leq 293\}$, variables appear to be highly correlated and show strong scale effect. A first non linear pre-processing is applied to build meteorological *indicators*, both to sum up the information as well to reduce the redundancy between the meteorological variables as already observed in [9] for instance. The new following standard indicator variables are then computed.

For each label $U \in \{T, N, W\}$, we introduce four non linear transformations of the meteorological inputs computed each half hour for the set of $n_U = 39$ stations for (T, N) and for the set of $n_U = 293$ network points for W:

- $U^{min} = \min(U^1, \ldots, U^{n_U})$
- $U^{max} = \max(U^1, \ldots, U^{n_U})$
- $U^{med} = \text{median}(U^1, \ldots, U^{n_U})$
- $U^{std} = \sqrt{\text{Var}(U^1, \ldots, U^{n_U})}$

These standard indicators provide a non linear sum-up of the variations and sizes of temperature, cloud cover or wind all over the French territory. Hence, $12 = 4 \times 3$ indicators are computed, half-hourly sampled, and stored from January 1st 2003 to August 31st 2010.

With a slight abuse of notation, for each label $U \in \{T, N, W\}$, we denote now by $U = [U^{min}, U^{max}, U^{med}, U^{std}]$ the reduced meteorological variable, where U^{min}, U^{max}, U^{med}, U^{std} are the standard indicators defined previously.

Finally, piling up all the days, a total of 12 indicators defines the meteorological process $M = [T\, N\, W]$ over the time. The meteorological conditions for day t are defined then by $M_t = [T_t N_t W_t]$, which is a 48×12 matrix.

Even if at this step of reading, the forecasting methodology has not been yet presented, we should say a brief remark in order to justify, at this stage, the choice of these four indicators. During this project, different methods have been investigated to forecast intra day load curves. One of them investigated to model the intra day load curves with a high dimensional model using all 371 meteorological available variables (371 = 273 + 2*39). As the fit of the intraday load curves was extremely accurate, the performance of the forecast was quite poor. This is explained by the fact that the global consumption, as we have already said, is actually composed of various regional consumptions. In this study, we did not have access to the regional consumptions which are known additionally to show quite high variability in space and in time. For a forecasting point of view, the method which sums up the meteorological data using four indicators performs the best at this stage.

3.3 Sparse Approximation of Intraday Load Curves

For each day t, we model the daily electrical consumption signal Y_t in a linear way using the following equation

$$Y_t = Z_t \beta_t + u_t \tag{1}$$

where the unknown parameter β_t (so depending on the day t) belongs to R^p with $p = 14$ and where the variable $Z_t = [P_t \ M_t] = [G_t \ Y_{t-7} \ T_t \ N_t \ W_t]$ is the concatenation of the pattern variables and meteorological variables previously described. G_t is the pattern variable previously defined. G_t takes eight modalities (Table 2) depending on days (variable D_t) and months of the year. The size of Z_t is $(n \times p) = (48 \times 14)$.

β_t is estimated using the LOLA algorithm especially chosen to produce a sparse representation: $\hat{\beta}_t = LOLA(Z_t, Y_t)$.

Note that LOLA is an algorithm providing good sparse approximation in very high dimension (see [17] in the case of the intra day load curve) and very accurate selection properties in medium high dimension (see [16]), which is the case here ($p = 14$). The adjusted electrical consumption is then:

$$\hat{Y}_t = Z_t \hat{\beta}_t \tag{2}$$

To evaluate the quality of this preliminary fit, we report here the *MAPE* and the *RMSE* errors, computed.

We observe that the selected covariables offer a quite high sparsity representation (Table 3): in average, $S = 2.5$ non zero coefficients are used to approximate the 365 intra-day load signals, with an average MAPE error of 1.2 % (median 1 %).

Figure 6 shows the sparse approximations $\hat{Y}_t = Z_t \hat{\beta}_t$ of Y_t for 4 days belonging to different seasons. For each graph, the number of selected coefficients with LOLA

Table 3 Statistical indicators for the sparsity, the MAPE and the RSME between the daily signal Y_t and its fit signal $\hat{Y}_t$, computed from September 1st 2009 to August 31st 2010. Groups are computed using previous years of available data

Statistical indicator	Sparsity	MAPE (%)	RMSE
Mean	2.49	1.24	833
Median	2.00	1.05	695
Std	0.81	0.79	531

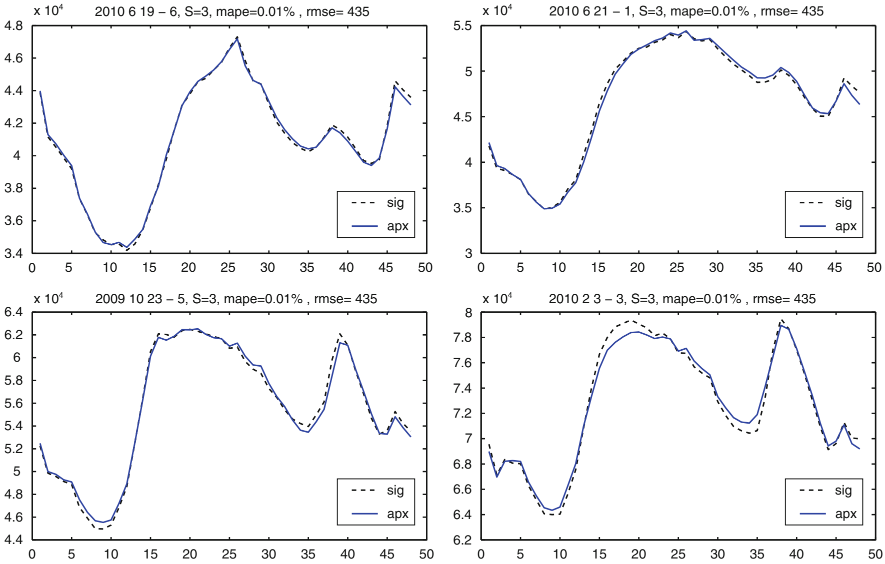

Fig. 6 Spring, Summer, Autumn and Winter intra day signals (*dashed line*) are here approximated using at most 3 coefficients selected by LOLA algorithm (*solid line*)

algorithm equals at most 3. Hence, we achieved here one step in our box: well approximating intra day consumption signals for various shapes and sizes using only few coefficients.

3.4 Smart Encyclopedia Contents

To summarize, the raw temporal data of electrical consumption as well as the meteorological inputs, available on the given historical period of 7 years and an half of data, have been daily processed as described in the previous subsections to produce the *Encyclopedia*, $\mathscr{E}$, providing for each day t, $1 \leq t \leq 2{,}800$:

- The daily electrical consumption Y_t,
- A qualitative description of t, given by calendar statements: D_t, C_t,
- A qualitative description of t, defined after an adaptive clustering on the data: G_t,
- The meteorological indicators over the French territory $M_t = [T_t\, N_t\, W_t]$,
- The estimated coefficient $\hat{\beta}_t$,
- The approximation of the daily consumption $\hat{Y}_t = Z_t \hat{\beta}_t$.

4 Intra Day Forecasting

We are now interested in the one day ahead forecast of the intraday load curve Y_t. Consumption and meteorological variables are then supposed to be known until day $t-1$, and we want to propose a forecast of the intraday load curve for the next day t. In this case, it is obviously not possible to use for the forecast, the approximated coefficient $\hat{\beta}_t$, since its computation involves the knowledge of the electrical consumption Y_t we precisely want to forecast.

To forecast the intra day load curve of the next day, called $\tilde{Y}_t$, we refer again to the previous linear model and write:

$$\tilde{Y}_t = Z_t\, \tilde{\beta}_t$$

- The matrix $Z_t = [P_t\ M_t]$ is known at t.

  $P_t = [G_t\ Y_{t-7}]$ defines the pattern of day t. Using the calendar interpretation of the clusters described in Table 2, the centroid of G_t is known as Y_{t-7} which is the intra day load curve, one week ahead and can be easily computed.

- The meteorological variables M_t are here supposed to be known. In real applications, these variables will be provided by Meteo France, the French company for weather prediction.
- The main issue here is to provide $\tilde{\beta}_t$.

Our approach will be to choose a "good candidate" for $\tilde{\beta}_t$, among the set of already estimated coefficients $\hat{\beta}_u$ with $u < t$. This strategy is motivated by the fact that the linear model introduced in Eq. (1) appears to be quite a good model to approximate the intra day load curve. Moreover, this model is sparse and thus relies only on a small number of coefficients.

In the forecast problem there is a typical balance to find between the need for increasing the number of coefficients to better approximate (bias correction), and the fact that each added coefficient increases the variability of the forecast, which strongly justifies the use of a sparse methodology.

4.1 *The Experts*

The forecasting begins with an information retrieval task. As explained before, at this stage, we will produce a variety of experts. Each expert essentially bases its approach on comparing the day t at hand to referring scenarios extracted from the backgrounds of the past i.e. finding a day t^* in the past which is closest (according to this expert) to the day t.

The motivation is that, for one day, similar causes of weather or calendar conditions or identical groups of consumption should provide similar effects and then a similar electrical consumption.

In order to retrieve t^*, different *strategies* s are introduced. Each strategy, s, is a function defined from $\mathscr{T}$ to $\mathscr{T}$ such that for any $t \in \mathscr{T}$, $s(t) = t^* < t$, where $\mathscr{T}$ denotes the set of indices of the different days. A forecasting *Expert* is then simply associated to a strategy s and provides a forecast of the intra day load signal of the next day t by plug-in the approximated coefficients $\hat{\beta}_{s(t)}$ calculated at day $s(t)$ chosen by strategy s:

$$\tilde{Y}_t^s = Z_t \hat{\beta}_{s(t)}$$

How to choose the experts? Many factors are known to have a potential impact on the electrical consumption and the next paragraph provides the 17 strategies introduced here to potentially forecast the intra-day load curves. Of course, much more strategies can be included but, for a sake a clarity we detail here some of the simplest as well as most efficient ones (Table 4).

Time-lags: Studies of historical intra day load curves typically show that the day before as well as the day one week before are significantly influential,[6, 22]. Consequently, strategy *Week* recalls the approximated coefficients of the same

Table 4 The forecasting experts

Strategies	Time lags impact
Yday	$t-1$
Week	$t-7$
Strategies	Meteorological scenarios, $s(t) = t^* = ArgMin_u sup\|.\|$
T	$[T_u - T_t]$
T^{med}	$[T_u^{med} - T_t^{med}]$
T^{med}/N	$[T_u^{med} - T_t^{med}]$ with $\|\|N_u^{med} - N_t^{med}\|\|_1/\|\|N_t^{med}\|\|_1 < 2\,\%$
T^{med}/W	$[T_u^{med} - T_t^{med}]$ with $\|\|W_u^{med} - W_t^{med}\|\|_1/\|\|W_t^{med}\|\|_1 < 2\,\%$
T	$[T_u - T_t]$
T/G	$[T_u - T_t]$ with $G_u = G_t$
T/D	$[T_u - T_t]$ with $D_u = D_t$
T/C	$[T_u - T_t]$ with $C_u = C_t$
N	$[N_u - N_t]$
N/G	$[N_u - N_t]$ with $G_u = G_t$
N/D	$[N_u^{med} - N_t^{med}]$ with $D_u = D_t$
N/C	$[N_u^{med} - N_t^{med}]$ with $C_u = C_t$
W	$[W_u - W_t]$
W/G	$[W_u - W_t]$ with $G_u = G_t$
W/D	$[W_u^{med} - W_t^{med}]$ with $D_u = D_t$
W/C	$[W_u^{med} - W_t^{med}]$ with $C_u = C_t$

day, one week before, $s(t) = t - 7$ and strategy *Yday* involves the "yesterday" approximated coefficients, $s(t) = t - 1$ (which is known to provide useful information from Tuesday to Wednesday) [21].

Meteorological scenarios (MS): Temperature is commonly admitted to be an important factor in France as, in winter, 80 % of the French heating comes from electrical devices. So called *windchill temperature*, which is a more complex phenomenon depending both on temperature, wind and cloud cover has also an impact on electrical consumption.

The strategies we provide in this domain will be contextual and retrieve in the past a day corresponding to the nearest neighbor, regarding meteorological intra day signals (temperature, wind and/or cloud cover). Different metrics have been investigated but the sup distance seems to be especially suitable. The distance is measured taking into account the median signal or all the indicators (min, max, med, std) of meteorological variables introduced in Sect. 3.2.

Strategy T (resp. T^{med}) refers to the day having the closest temperature indicators (resp. the closest median temperature) among all the days in the past. Strategy N (resp. W) refers to the day having the closest cloud cover (resp. wind) indicator.

Cloud cover and wind may have special effects, so we consider strategies that specifically single out their effect.

Temperature constrained by cloud cover and wind: Strategy T^{med}/N refers to the day having the closest median temperature given cloud cover. More precisely, this means that we begin by selecting the days in the past which have a cloud cover signal in a small vicinity of 'today', and among these ones we choose the day with the closest T^{med}.

In the same way, strategy T^{med}/W refers to the day having the closest median temperature given the wind.

Impact of meteorological factors on electrical consumption also depends on day type (week days, week end or public day).

MS constrained by groups: Clustering methods applied on historical consumption data have exhibit specific groups of consumption [17]. Strategy T/G refer to the day having the closest temperature indicator given the group of t. Strategy N/G (resp. W/G) refers to the day having the closest cloud cover (resp. wind) indicator given the group of day t.

MS of the day constrained by the type of day: [1] shows the importance of the type of day. Strategy T/D refers to the day having the closest indicators (min, max, med, sd), given the kind of day of t Strategy N/D (resp. W/D) refers to the day having the closest cloud cover (resp. wind) indicator (min, max, med, sd) given the kind of day of t.

MS of the day constrained by a calendar group: [21] introduced five calendar groups to forecast (see Sect. 2.3). Strategy T/C refers to the day having the closest temperature indicators(min, max, med, std) for days belonging to the same calendar group as t. Strategy N/C (resp. W/C) refers to the day having the closest cloud cover (resp. wind) indicators (min, max, med, std) for days belonging to the same calendar group as Y_t.

Here a meaningful discussion could be engaged on the choice of the experts. For instance, experts could integrate information about "special events" such as announced strike, big soccer games... We did not include such experts in the present study, since we observed in our data, that these events are at the same time rare and with high variable responses, in such a way that the learning process on them did not seem to show clear effects. Consequently we preferred to postpone this delicate point to a further study.

4.2 Smoothing Parameter

In this approach, the approximated intra day load is computed day by day and up to now, no continuity assumptions have been introduced between two consecutive days. In this perspective, several approaches can be used. For sake of simplicity, we chose to present here the simplest one.

In order to address this problem of introducing a regularity constraint between days, a parameter called δ_t^s reflecting the possible lag between day $t-1$ and prediction at day t is introduced and maintain in a *security zone*: $\tilde{Y}_t^s \leftarrow \tilde{Y}_t^s + \delta_t^s$

In this application, where the forecast is computed for 24 h, we simply choose $\delta_t^s = Y_t(1) - \tilde{Y}_t^s(1)$, and maintained it to be zero. $Y_t(1)$ (resp. $\tilde{Y}_t^s(1)$) is the value of the first point $(00 : 30')$ of the load curve (predicted load curve using strategie s) for day t. It means that the forecast is actually beginning each day at $00 : 30'$ for the next 23 h $30'$. When this method is used in other contexts (for a 36 h forecast for instance), a more refined δ_t^s parameter can be computed.

4.3 Performances of the Various Experts

Table 5 presents the forecast performances, for the $K = 17$ experts considered previously, for 1 year of data, from September 1st 2009 to August 31st 2010. The historical set of data to retrieve the $\hat{\beta}_t^*$, contains 3 years of data, from September 1st 2006 to August 31st 2009.

For sake of comparison, an additional *naive* expert is introduced which forecasts the daily electrical consumption of day t by simply using the intra day raw consumption signal of the previous day: $\tilde{Y}_t = Y_{t-1}$.

An approximation expert (called "Apx") is also introduced. The oracle expert cannot be used in practice, but gives a benchmark of the method: in this case $\tilde{\beta}_t = \hat{\beta}_t$.

The difference between these two performances measures the gain that can be expected with respect to a crude prediction.

Table 5 MAPE performances for each expert in forecasting 1 year of data from September 1st 2009 to August 31st 2010

Names	Average	Median	Std
Naive	0.0634	0.0415	0.0514
Apx	0.0170	0.0145	0.0145
The experts:			
Yday	0.0300	0.0231	0.0231
Week	0.0293	0.0236	0.0236
T^{med}	0.0315	0.0261	0.0261
T^{med}/W	0.0351	0.0252	0.0252
T^{med}/N	0.0320	0.0257	0.0257
T	0.0311	0.0238	0.0238
T/G	0.0310	0.0232	0.0232
T/D	0.0321	0.0262	0.0262
T/C	0.0295	0.0249	0.0249
N	0.0406	0.0293	0.0293
N/G	0.0282	0.0210	0.0210
N/D	0.0284	0.0220	0.0220
N/C	0.0287	0.0220	0.0220
W	0.0384	0.0294	0.0294
W/G	0.0309	0.0241	0.0241
W/D	0.0381	0.0305	0.0305
W/C	0.0317	0.0256	0.0256

4.3.1 Detailed Performances of the Experts

The naive approach shows a 6.3 % MAPE error. Time lag strategies (Yday, Week) behave well compared to the overall strategies (MAPE average of 3.0 % or 2.9 %). Forecast results are significantly improved by plug-in the sparse estimated coefficients computed the day before instead of taking the raw intra day load curve of the previous day. In order to stress the importance of variable selection, we have computed ordinary Least Square regression (OLS), for theses two time lag strategies. Compared to the LOLA algorithm actually used, no sparsity constraint is introduced in the least square method. For the same period of analyze, the OLS approach shows a MAPE average of 3.98 % for the Yday strategy and of 4.15 % for the Week strategy. With no variables selection, the MAPE average error increases of approximatively 30 %. Similar deteriorations of performances are observed using OLS instead of LOLA for the other strategies.

These results strongly show the benefits of using a sparse functional model to forecast the electrical consumption.

Using LOLA algorithm, strategies associated to the closest cloud cover conditions with constraints (N/G, N/D, N/C), behave especially well, compared for instance, to strategies only based on temperature (T), cloud cover (N) or wind (W). The strategy retrieving the day relying on cloud cover given group information

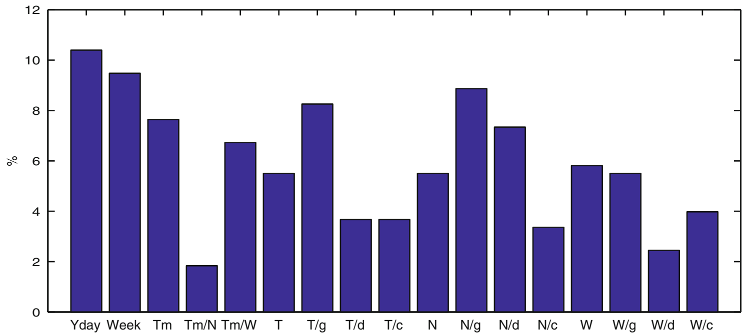

Fig. 7 Frequencies for the expert which perform at best computed for 1 year of data from September 1st 2009 to August 31st 2010

provides, in average, the best MAPE error (mean: 2.82 %; median: 2.10 %). Hence, these results strongly suggest that cloud cover seem to have a important impact in the electricity consumption and then the forecast.

Figure 7 provides for each strategy, the frequency of hits, i.e. the frequency of days over 1 year when the strategy performs at best.

The different strategies seem quite competitive and we observe that all of them performs at best at least 2 % of the days. Strategies based on finding the closest day regarding temperature, cloud cover given the group of day performs in general well (N/G: 9 % of hits), as it was already observed in Table 5. Time dependent strategies, Yday or Week seem also quite competitive with hit frequencies equal to 10 % or 9 %.

These quantitative forecasting results seem to reflect quite well the opinion of the human experts of the discipline.

4.4 Aggregation of Experts

If we were able to find each day the best strategy to apply regarding the MAPE error, the oracle MAPE average error over 1 year would be equal to 1.44 % (standard deviation 0.74 %), as presented in Table 6. This is similar to the approximation MAPE error (Table 4, Apx: 1.70 %) and quite a good performance for these prediction experts which purpose are, as explained before, more adaptation to high dimensional information.

As seen in the previous part, the experts perform successively well depending on days, or meteorological issues. But no one among them achieves the best performance most of the time. There is an obvious need to combining them.

Table 6 MAPE performances in % for aggregation method

Names	Average	Median	Std
Oracle	1.44	1.29	0.74
Exponential weights	2.25	1.92	1.24

In the recent years, many interesting theoretical results as well as practical simulations have been obtained using aggregation and especially exponential penalization: see [4, 7, 8, 11, 23]. These techniques will be for us a good source of inspiration. However, a crucial problem then is to find a weighting, learning the performances of each expert and optimizing them. In this context of prediction, this is quite a challenging issue which can give rise to very sophisticated procedures.

For sake of simplicity we present here a very understandable and manageable one, which only records the approximation properties of each expert and penalizes those with poor approximation results. More precisely, let us recall that $\mathscr{M}$ is the set of strategies introduced above, and $\hat{Y}^s$ the expert forecast computed with the strategy s.

The aggregated expert is a weighted sum of all the forecast consumptions provided by the different experts:

$$\hat{Y}_t = \frac{\sum_{s \in \mathscr{M}} w_t^s \tilde{Y}_t^s}{\sum_{s \in \mathscr{M}} w_t^s}$$

where w_t^s are positive weights depending on the day t and the strategy s.

As explained above, our procedure penalizes by putting small weights, on the strategies which were not able to well approximate the signal at $s(t)$: e.g. the weights w_t^s depend in an exponential way on the l_2 error of $\|Y_{s(t)} - \hat{Y}_{s(t)}||_2^2$:

$$w_t^s = exp(-\|Y_{s(t)} - \hat{Y}_{s(t)}\|_2^2/\theta)$$

$\theta > 0$ is a standard tuning parameter (also called temperature parameter with reference to statistical physics). In the performances presented in the Table 6, this parameter was chosen using cross validation on the past. Using aggregation with exponential weights, we observe that the MAPE decreases to 2.25 % in average and to 1.92 % in median, with a standard deviation of 1.20 %. This error is much smaller than the different errors computed for each individual experts, presented in Table 4 showing the benefits of the different contributions.

Figures 8 and 9 give a graphical illustration of forecast for two different weeks chosen in winter and spring. We observe that forecasts are more accurate during spring periods than winter periods. In Fig. 9, local maxima seem to be overestimated, while local minima are underestimated.

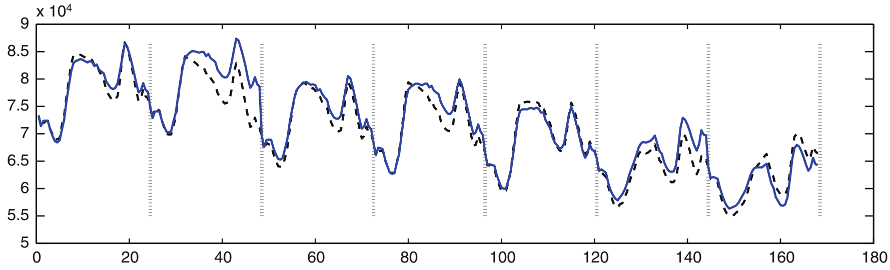

Fig. 8 Forecast (*solid blue line*) and observed (*dashed dark line*) electrical consumption for a winter week from Monday February 1st to Sunday January 7th 2010

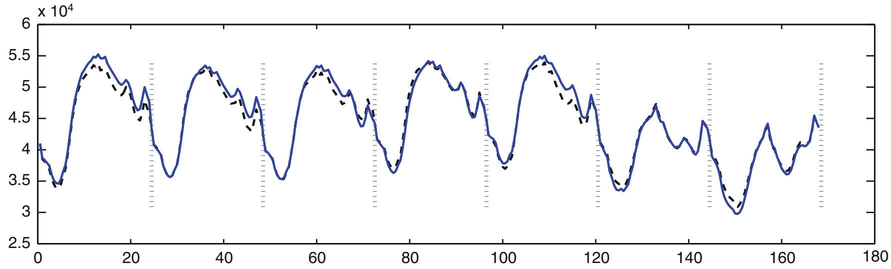

Fig. 9 Forecast (*solid blue line*) and observed (*dashed dark line*) electrical consumption for a spring week from Monday June 14th to Sunday June 21st 2010

5 Conclusion and Perspectives

The method described above will be implemented in a short-term consumption forecasting platform, aggregating various models, which is currently tested at RTE.

This collaboration between academics and industrials provides results which, although coming from automatic statistical methods, happen to agree surprisingly well with the business knowledge of RTE. In fact, they go further and shed new lights:

– The differences between the performances of OLS predictions and the sparse method approach if they were expected from a theoretical point of view, happen to be surprisingly striking (30 % !), strongly motivating for going in this direction in the future.
– The approach for the construction of the experts (searching for *similar days* in the past) which has been established using a mathematical perspective is finally quite close to the strategies of the RTE forecasters. If we refer for instance to the comparison between the expert performances, some results are *common sense* (the lag-expert performances for instance). However, the impressive performances of cloud cover experts were a little more difficult to predict, but already observed by RTE forecasters.

– The competitiveness of these experts, also expected from the RTE forecasters, have been highlighted and quantified. In the near future, more experts will be introduced. As well, the method of aggregation will be diversified according to the feedback of the short term consumption forecasting platform.

Due to forecasting operational needs, different adaptations of the forecasting box will be provided. Particularly, the horizon forecast will be extended to 48 h, or more and the method will be adapted to choose the delivery time of prediction, according to business constraints. In this perspective, the smoothing parameter will be refined to integrate nonparametric regularization methods as well as designed strategies of RTE forecasters.

Acknowledgements The authors thank RTE for the financial support through two industrial contracts, LPMA for hosting our researches, and Karine Tribouley for taking part of an earlier elaboration of this project.

References

1. Antoniadis, A., Brossat, X., Cugliari, J., & Poggi, J. M. (2010). Clustering functional data using wavelets. In *Proceedings of the 19th international conference on computational statistics, COMPSTAT*, Paris, 697–704.
2. Antoniadis, A., Brossat, X., Cugliari, J., & Poggi, J. M. (2012). Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. Application à la consommation d'électricité. *Journal de la Société; Française de Statistique, 153*(2), 52–78.
3. Antoniadis, A., Paparoditis, E., & Sapatinas, T. (2006). A functional wavelet–kernel approach for time series prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68*(5), 837–857.
4. Catoni, O. (2004). *Statistical learning theory and stochastic optimization, volume 1851 of Lecture notes in mathematics*. Berlin: Springer.
5. Chakhchoukh, Y., Panciatici, P., & Bondon, P. (2009). Robust estimation of SARIMA models: Application to short-term load forecasting. In *IEEE workshop on statistical signal processing*, Cardiff
6. Cho, H., Goude, Y., Brossat, X., & Yao, Q. (2013). Modelling and forecasting daily electricity load curves: A hybrid approach. *Journal of the American Statistical Association, 108*(501), 7–21.
7. Dalalyan, A. S., & Tsybakov, A. B. (2008). Aggregation by exponentiel weighting, sharp oracle inequalities and sparsity. In *COLT*, Helsinki (pp. 97–111).
8. Devaine, M., Gaillard, P., Goude, Y., & Stoltz, G. (2012). Forecasting electricity consumption by aggregating specialized experts. *Machine Learning, 90*, 1–30.
9. Fan, S., & Hyndman, R. J. (2010). Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems, 25*(2), 1142–1152.
10. Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive approach. *IEEE Transactions on Power Systems, 27*(1), 134–140.
11. Juditsky, A. B., & Nemirovski, A. S. (2000). Functional aggregation for nonparametric regression. *The Annals of Statistics, 28*(3), 681–712.
12. Kerkyacharian, G., Mougeot, M., Picard, D., & Tribouley, K. (2009). Learning out of leaders. In *Multiscale, nonlinear and adaptive approximation* (Lecture notes in computer science). Berlin: Springer.

13. Lefieux, V. (2007). Modèles semi-paramétriques appliqués à la prévision des séries temporelles: cas de la consommation d' électricité.
14. Marin, F. J., Garcia-Lagos, F., & Sandoval, F. (2002). Global model for short term load forecasting using artificial neural networks. *IEE Proceedings – Generation, Transmission, and Distribution, 149*, 121–125.
15. Mougeot, M., Picard, D., & Tribouley, K. (2012). Learning out of leaders: Regression for high dimension. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74*, 475–513.
16. Mougeot, M., Picard, D., & Tribouley, K. (2014). LOL selection in high dimension. *Computational Statistics and Data Analysis, 71*, 743–757.
17. Mougeot, M., Picard, D., Tribouley, K., Lefieux, V., & Maillard-Teyssier, L. (2013). Sparse approximation and fit of intraday load curves in a high dimentional framework. *Advanced in Adaptive Data Analysis, 5*. http://www.worldscientific.com/doi/pdf/10.1142/S1793536913500167.
18. Muñoz, A., Sánchez-Úbeda, E. F., Cruz, A., & Marin, J. (2010). Short-term forecasting in power systems: A guided tour. In *Handbook of power systems II* (pp. 129–160). Berlin/Heidelberg: Springer.
19. Poggi, J. M. (1994). Prévision non paramétrique de la consommation d'électricité. *Revue de Statistique Appliquée, 42*, 83–98.
20. Ramsay, J. O., & Silverman B. W. (2005). *Functional data analysis*. New York: Springer.
21. Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research, 204*(1), 139–152.
22. Taylor, J. W. (2012). Short-term load forecasting with exponentially weighted methods. *IEEE Transactions on Power Systems, 27*, 458–464.
23. Tsybakov, A. B. (2003). Optimal rates of aggregation. In *COLT*, Washington, DC (pp. 303–313).