Local Short and Middle Term Electricity Load Forecasting with Semi-Parametric Additive Models

Yannig Goude, EDF R&D, Raphael Nedellec, EDF R&D, and Nicolas Kong, ERDF.

***Abstract*—Electricity load forecasting faces rising challenges due to the advent of innovating technologies such as smart grids, electric cars and renewable energy production. For distribution network managers, a good knowledge of the future electricity consumption stands as a central point for the reliability of the network and investment strategies.**
In this paper, we suggest a semi-parametric approach based on generalized additive models theory to model electrical load over more than 2200 substations of the French distribution network, and this at both short and middle term horizons. These generalized additive models estimate the relationship between load and the explanatory variables : temperatures, calendar variables, etc. This methodology has been applied with good results on the French grid. In addition, we highlight the fact that the estimated functions describing the relations between demand and the driving variables are easily interpretable, and that a good temperature prediction is important.

***Index Terms*—load forecasting, generalized additive model, semi-parametric model, time series, electricity networks**

I. Introduction

Electricity load forecasting from short term (hourly and daily), middle term (monthly to yearly) to long term horizon (5 to 30 years) has received a lot of attention from industrial and academics in the recent years. For electricity providers, forecasting electricity demand is a key activity as it is one of the most important entries for production planning and trading on the electricity markets. For electricity network managers a good knowledge of the future electricity consumption stands as a central point for the reliability of the network and investment strategies.

Recently, the advent of innovating technologies such as smart grids, electric cars or renewable energy production induced a lot of new perspectives for energy management and consequently for electricity demand forecasting. In France, ERDF (Électricité Réseau Distribution de France the French manager of the public electricity distribution network) has launched recently an AMM (Automated Metering Management) project that aims to install 300000 individual smart meters for experimenting real time individual data collection and processing. There are many challenges, two of them are local optimization of electricity production and real time management of individual demand -see e.g. [1]-; this leads to develop new methods to forecast a changing electricity demand at different scales -individual loads, a block, a region etc...-.

The literature on load forecasting is rich and many works have been done concerning forecasting electricity demand at an "aggregated" level -for an entire country, big towns or regions-. Classical statistical methods have been applied to electricity demand forecasting at a short term horizon such as SARIMA models in [2] or [3] and exponential smoothing in notably [4] and [5]. To deal with covariates -mostly meteorological- high dimensional linear and non-linear regression models have been successively proposed for short term and middle term horizon in [6] [7], [8] and [9]. State-space models presented in [10] and [11] show an interesting ability to adapt to smooth changes in electricity data. Other approaches based on machine learning methods also provide good results, as in [12], [13] and [14].

In this paper we focus on forecasting local electricity demand on the distribution network in France. More precisely, we study electricity load collected every 10 minutes by ERDF at 2260 substations located at the frontier between the high voltage grid and the distribution network in France. To manage the distribution grid, quantify the constraints on the network and optimize the configuration on the grid consequently, ERDF needs to produce short term -day ahead- and middle term -year ahead- forecasts for each substations. Each of them is in average linked to 40 big customers -industries, supermarkets etc- and 16000 small ones -residential, small businesses etc- but can have very different properties depending on their location and the type of costumers connected to it. To deal with the variability of the data, we propose to use semi-parametric additive models which are popular statistical models that show interesting capacity to adapt quite automatically to different data sets as explained in [15] and [16]. This methodology has already been applied for short term forecast on the french electricity data at a national level in [17] and on regional data in the National Electricity Market of Australia in [18]. We also applied it during the GEFcom competition on US data sets, see [19]. In both cases, semi-parametric models show an interesting trade off between an ability to capture complex relationships in the data and a quite automatic estimation processes that do not require intensive human intervention. In addition, the forecast performances obtained are really good in comparison with other methods and these models can be computed at a relatively low computational cost.

The purpose of this study is to suggest a methodology, based on semi-parametric models, to forecast the 2260 time series recorded on the distribution grid at a daily and yearly horizon. We apply it to a data set provided by ERDF and validate our

Y. Goude and R. Nedellec are with the Electricité de France Research & Development Division, 1 av du Général de Gaulle, 92141 Clamart Cedex, FRANCE. Contact: yannig.goude@edf.fr or raphael.nedellec@edf.fr.

N. Kong is with Électricité Réseau Distribution France, Direction Technique, Tour Winterthur, 102 terrasse Boieldieu 92085 Paris La Défense. Contact: nicolas.kong@erdfdistribution.fr

approach out-of-sample over the last year of the data set.

II. Methodology

A. Statistical framework

Consider that we want to fit the following statistical model:

$$y_i = f_1(x_{1,i}) + f_2(x_{2,i}) + ... + f_p(x_{p,i}) + \varepsilon_i$$

where y_i is a univariate response variable, $x_{q,i}$ are the covariates that drive y_i. In the following application, y_i will be the electricity demand, $x_{q,i}$ will be the meteorological predictors, the calendar effects and etc. ε_i denotes the model error at time i. The non-linear functions f_q are supposed to be smooth, which means here that it can be relatively well estimated by penalized regression in a spline basis. Thus, each function is expressed like this:

$$f_q(x) = \sum_{j=1}^{k_q} \beta_{q,j} b_j^q(x)$$

where k_q is the dimension of the spline basis to model the effect f_q and $b_j^q(x)$ the corresponding spline functions, for example B-splines or cubic regression splines. A classical way to estimate this smooth effects is penalized regression, more precisely ridge regression, where we minimize the following criteria:

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

where the penalty parameter $\Lambda = (\lambda_1, ..., \lambda_p)$ which controls the degree of smoothness of each effect -the higher λ_q the smoother f_q is- has to be optimized. Denoting B the matrix formed by concatenation of the b_j^q, we have to solve the following problem:

$$\min_{\beta,\lambda} \|\|Y - B\beta\|\|^2 + \sum_{q=1}^{p} \lambda_q \ \beta^T S_q \beta$$

where β is the vector of the unknown regression parameters, S_q is a smoothing matrix depending on the spline basis. This problem is solved using the methodology presented in [20] and [21] which consists in minimizing the GCV -Generalized Cross Validation- criteria proposed in [22]. We will use for that the R package *mgcv* (see [23] and [16]) that implements this method.

B. A semi-parametric model for electricity data

Our approach consists in designing a general semi-parametric model that can then be applied to each of the 2260 substations to provide short and middle term forecast. In other words, we suppose that a single equation model can be declined locally and that the semi-parametric model is sufficiently flexible to adapt to each time series. This hypothesis is quite restrictive and should be relaxed, for example using automatic model selection methods like grouped lasso or any shrinkage method. One reason for our choice is that our approach already gives good forecasting results at a lower computational cost but we keep these ideas in mind for future works.

It is well known -see [11], [4] and [17]- that the French electricity demand exhibits a trend corresponding to economic and demographic growth, different seasonalities as an intra-day cycle, a weekly cycle and a yearly cycle and is driven by meteorological data, mostly by temperature. As in [17] and [18] we fit one model per instant of the day. The dates are recorded each 10 minutes so that we fit 144 models corresponding to the 144 instants per day. We also tried to fit a single model instead of 144 models, to capture the time structure of the data and the correlation between the instants of the day but obtained better results in terms of goodness of fit and computation time with one model per instant. For simplicity, we don't make the dependency on the instant of the day apparent in the equations bellow, as each time series composed by the data measured at one instant of the day are treated independently. For example, y_t^{100} the t^{th} day of electrical consumption measured at the 100^{th} instant of the day will be denoted y_t. Note also that we tried using a log-transformation of load curves. We did not notice any improvement and chose to use raw demand.

Based on that and successive experiments, we suggest the following model for Middle term (MT) forecasting.

MT model

$$\begin{aligned} y_t &= \sum_{j=1}^{7} m_j \, \mathrm{I}_{\text{DayType}_t=j} + k \, \mathrm{I}_{\text{SpecialTariff}_t=1} \\ &+ g_1(\theta_t) + g_2(T_t) + g_3(T_{t-1}) + g_4(T_{t-2}) \\ &+ \sum_{j=1}^{11} o_j \, \mathrm{I}_{\text{OffSet}_t=j} + h(toy_t) \\ &+ \varepsilon_t \end{aligned} \tag{1}$$

where:

- y_t is the electric demand recorded at time t -for one instant of the day-
- DayType_t is the type of day for the observation t: 1 for Sunday, 2 for Monday, 3 for Tuesday-Wednesday-Thursday, 4 for Friday, 5 for Saturday, and 6 and 7 for bank holidays
- SpecialTariff_t is a factorial predictor, taking values 0 when there is no special tariff at time t and 1 otherwise
- θ_t is a smooth temperature which is an exponential smoothing of the real temperature T_t: $\theta_t = (1-0.99)T_t + 0.99\theta_{t-1}$, T_{t-1} is the lag 1 temperature -the temperature of the day before- and T_{t-2} is the lag 2 temperature
- OffSet_t is a categorical variable indicating the holidays and daylight saving time
- toy_t is the time of year which is the position of the observation t within the year -from 0 January the 1^{st} to 1 December the 31^{st}-, $h(toy_t)$ corresponds to the smooth variation of the yearly cycle of the load per instant of the day

An important point is that this model realizes a good trade-off between the fit and the complexity of the model -which is one property of the models obtained by minimizing the GCV criteria- and that is crucial to avoid computational issues.

The weekly cycle of the electricity consumption is introduced in the linear part of the model. For each instant of the

day, each model as one coefficient per type of day, globally resulting to one coefficient per instant and per type of day at the end -considering the 144 models associated to each instant-. The special tariff effect correspond to a tariff option for big customers that can be activated by energy suppliers about 20 days a year to moderate peaks in electricity consumption -in winter in France-. These tariffs are very attractive all the year but customers have to pay big penalties for consuming when this option is activated, resulting to a significant decrease of the load during special tariff days. Other tariffs exist which have an impact on the electricity load in France, but as we'll see bellow, these tariff depend on the hour of the day and are embedded in the weekly and daily cycles.

The temperature T_t is associated to the substation considered, we will explain how in section III. To model the inertia of the demand to the temperature -mainly due to the isolation of the buildings- we add two lag temperatures and a smooth temperature effect. We optimized the smoothing parameter to 0.99 using the GCV criteria on a sub-sample of 120 substations chosen randomly among the 2260 substations but we noticed that around this optimal value the results are quite insensitive to this parameter.

At a short term horizon, the forecaster has access to recent electricity load observations to produce a forecast and we derive a short term model from (1) adding a lag load effect $l(y_{t-1})$ as presented in (2).

ST model

$$\begin{aligned} y_t &= \sum_{j=1}^{7} m_j \, \mathrm{I}_{\mathrm{DayType}_t=j} + k \, \mathrm{I}_{\mathrm{SpecialTariff}_t=1} \\ &+ g_1(\theta_t) + g_2(T_t) + g_3(T_{t-1}) + g_4(T_{t-2}) \\ &+ \sum_{j=1}^{11} o_j \, \mathrm{I}_{\mathrm{OffSet}_t=j} + h(toy_t) \\ &+ l(y_{t-1}) \\ &+ \varepsilon_t \end{aligned} \tag{2}$$

C. Modeling the trend

At a substation level, modeling the trend for middle term horizon is a tricky issue since it is often driven by a lot of unknown covariates. As an example, the building of a commercial mall in a substation area will induce a big jump in the electricity consumption and there is no way to predict it without any information about it. For this study, no commercial or sociological data useful to forecast the trend were provided at this local scale but this concern will be addressed in future work. In our experiments we tried to add into our MT model a trend based on past observations -linear trend, non-linear trends-. However, we did not observe any improvements, because this trend is often hard to catch and hidden by the other covariates. To analyze the impact of forecasting trends we propose another approach based on two successive steps. First, we detrend the data at a monthly scale -we explain it below-, then we fit a model on the detrended data. Then, the forecast are obtained summing the detrended forecasts and the estimated monthly trend. In the following we will produce some forecasting results of this approach supposing that the trend is known in advance which is obviously not true. The aim is to compare the performances of the model (1) with an optimistic benchmark and to evaluate what could be the gain induced by an improvement of the trend modeling. In the following we will call this approach the detrending approach.

Detrending the data is a complex process as for the electricity data either the calendar effects and the meteorological effects could have a low frequency behavior hard to distinguish from a trend. For example, a very cold winter in a substation area could induce an increase of the electrical heating in this area and thus a change in the global level on the electricity consumption. So, the impact of temperature and other covariates on the load is so important on most substations that classical univariate detrending methods for time series can't be applied here. We suggest a two step semi-parametric model as a natural and efficient way to solve that problem.

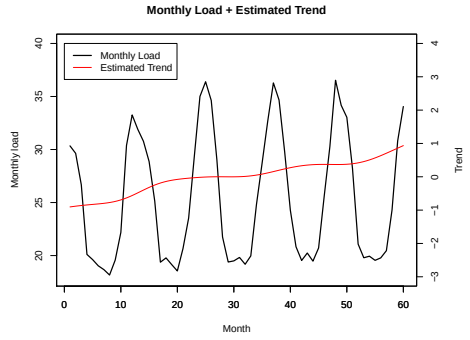

Fig. 1. Monthly Load and Estimated Trend

As the objective is to estimate a trend which is a low frequency effect independent of the meteorological effects, we first estimate a very simple GAM model at a monthly scale -monthly data-, retrieve some residuals, then estimate the trend with a kernel smoothing method, then use this estimate to detrend the data and finally estimate the final semi-parametric model on the detrended data. We can see on Figure 1 the trend estimated in red and the monthly data in black. Due to temperature effects, we could not have seen the same trend on original data.

So, we first aggregate the data by month which result in monthly electricity loads and temperatures time series for every substations. Denoting y_t^m and T_t^m these times series, we estimate the following semi-parametric additive model (3):

$$y_t^m = \sum_{j=1}^{12} c_j \, \mathrm{I}_{\mathrm{Month}_t=j} + f(T_t^m) + \varepsilon_t \tag{3}$$

where:

- $\mathrm{I}_{\mathrm{Month}_t=j}$ is an indicator variable equal to 1 the month of the observation t is j from 0 to 12 and 0 otherwise
- f is the effect of the monthly temperature of the monthly electricity load

Then we consider the monthly estimated residuals $\widehat{\varepsilon}_t^m = y_t^m - \widehat{y}_t^m$ where $\widehat{y}_t^m$ is the estimated load from the model 3, and estimate the smooth residuals $M_\eta(\widehat{\varepsilon}^m)(t)$, where M_η is

$$M_\eta(x)(t) = \sum_{i=1}^{n} x_i K_\eta(i,t) / \sum_{i=1}^{n} K_\eta(i,t) \tag{4}$$

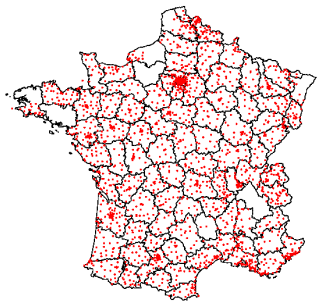

Fig. 2. Location of the 1976 among 2260 substations.

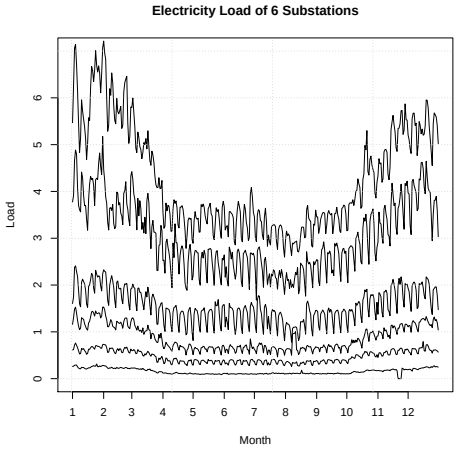

Fig. 3. One year of weekly electricity consumption for 6 substations.

with K the classical Gaussian kernel $K_\eta(x,y) = \exp(-\eta(x-y)^2)$ and $\varepsilon^m = (\varepsilon_1^m, \varepsilon_2^m, ..., \varepsilon_n^m)$ the vector of the residuals over the n month of the estimation-forecasting period.

The bandwidth parameter is set to $\eta = 0.024$ -corresponding to a Gaussian law with standard deviation of 3 months- to insure the regularity of the trend estimates based on ERDF expert advises. As $M_{0.024}(\widehat{\varepsilon}^m)(i)$, $i = 1, .., n$, is a monthly time series, we interpolate it linearly to obtain $\widehat{Tr}_t$ the trend estimated at a 10 minutes frequency. We finally apply the model (1) on the signal $y_t - \widehat{Tr}_t$. The final forecasts are the sum of the forecast of this model and the estimated trend $\widehat{Tr}_t$. We will call this model the MTD (Middle Term Detrending) model in the following.

III. Application

In this section we apply the model of the section II-B to real electricity data. We first describe the dataset and the estimation procedure and then propose two forecasting cases. In the first forecasting run, we test the capacity of our model to adapt automatically -without human intervention- to each substations. We proceed to the forecast of all the substations assuming that the realization of the meteorological covariates is known in advance which is of course not the case in practice but allows us to quantify the performances of our model without embedding the meteorological forecasting errors. The second forecasting run simulates a real forecast in real operational condition and show the feasibility of our method in practice. The last part of this section is devoted to the analysis of these models. We focus on a few particular substations to illustrate the ability of our semi-parametric models to capture and represent the different features of the electricity consumption.

A. Model estimation

The data set used in this section is composed with electricity demand provided by ERDF from 2260 substations in France recorded every 10 minutes from January the 1^{st} 2006 to December the 31^{st} 2011. We also have access to temperature data recorded every 3 hours at 63 weather stations in France provided by MeteoFrance. The locations of 1976 substations among the 2260 are represented on the Fig. 2 -some departments have missing GPS data-. A weather station is affected to each substation by a meteorologist, corresponding to the closest weather station in terms of climate properties -it can occur that the closest weather station is not relevant to explain the climate in the area of a substation, particularly in mountain regions-. The temperature data are interpolated linearly to fit with the frequency of the electrical recording. We separate the data between an estimation set, going from January the 1^{st} 2006 to December the 31^{st} 2010 and a forecasting set, the year 2011. We estimate on each substation the model summarized in (1) and the corresponding short term forecasting model. As presented on Fig. 3, the substations can have very different properties and modeling it automatically is a challenging task.

We notice that a lot of outliers are present in the data. These outliers can come from database errors but also can be the consequence of network reconfigurations or physical injuries on the grid. We exclude 360 time series among 2260 as they contain too much outliers to estimate our model. Among the 360 excluded series, a proper cleaning of the data would make some of them predictable with ours models, but this pre-processing step would be off-topic.

We implement these models in R with the mgcv package develop and maintain by Simon Wood, see [16] for a complete and friendly description of it. As each substation is treated independently, we parallelize the calculation using the multicore package doMC from Revolution Analytics. These package can be found on CRAN http://cran.r-project.org/. For this experiment we use a personal computer hp z600 -12Go RAM, 8 proc. intel xeon E5620@2.4GHz- which is a powerful machine but still a pc. The estimation of the 3 models -MT, MTD and ST- for the 1900 substations takes 52 hours -including the access to the data base and the saving of the results on the hard drive-. All the data, results and models correspond to 1.2 To of information.

B. Forecasting

We measure the forecasting performances of our model on the forecasting set (year 2011) with the Mean Absolute Percentage Error (MAPE). This measure is a very classical tool in time series and is particularly relevant here as we deal with time series of different scales. We compare the MAPE of our MT and ST models with three naive benchmarks:

- **D1**: the load of the day before.
- **D7**: the load of the week before.
- **Y1**: the electricity load of the year before translated so that the days of the week corresponds. For instance, as January the 3^{rd} 2011 is a Monday, so we forecast it with January the 4^{rd} 2010 which is also a Monday.

The forecasting results we obtained are presented in detailed on the Fig. 4 and Fig. 5 and summarize in Table I. The median MAPE of the middle term forecast is 8% and the very bad forecasts (more than 20% in MAPE, corresponding to 71 substations) are due to network reconfigurations or local trends -construction or destruction of new buildings, installation of new companies etc- that are impossible to predict without side informations. Detrending the data has a large impact since it reduces the median MAPE of the MTD model to 6% but also the number of substations with a MAPE larger than 20% to 31. At a short term horizon, adding the lag load effect is a good way to capture those variations and the median is considerably reduced to 5%.

quantile	CT	MT	MTD	D1	D7	Y1
10%	0.04	0.05	0.04	0.06	0.08	0.10
25%	0.04	0.06	0.05	0.07	0.09	0.12
50%	0.05	0.08	0.06	0.09	0.10	0.14
75%	0.06	0.11	0.08	0.11	0.12	0.18
90%	0.09	0.15	0.12	0.14	0.15	0.22

TABLE I
PERFORMANCES OF THE DIFFERENT MODELS.

On the Fig. 4, the solid thick line represents the ordered MAPE for the MT model, the dashed thick line represents the ordered MAPE for the MT model with the detrending step and the dashed line -noisy signal- the MAPE of the Y1 benchmark. We clearly see on this graph that for most of the substations our models are far more better than the Y1 benchmark. A more in depth analysis shows that the MT (resp. MTD) model forecasts are more than 2 times better than the Y1 for 42% (resp. 66%) of the substations and better for 93% (resp. 99%) of them. The substations for which our models obtain the worst results are mostly those where the trend is difficult to predict as the MTD model is largely better than the MT model in those cases.

The results for the short term horizon are presented on Fig. 5. The solid thick line represents the MAPE for the ST model and the dashed line the the MAPE of the D1 benchmark -the best benchmark in terms of median MAPE-.

As for the middle term horizon, the short term forecasts of the ST model are largely better than the benchmark performances for most of the substations. More precisely the ST model obtained better MAPE than the benchmark on 96% of the substations and for 66% of them it is 1.5 times better.

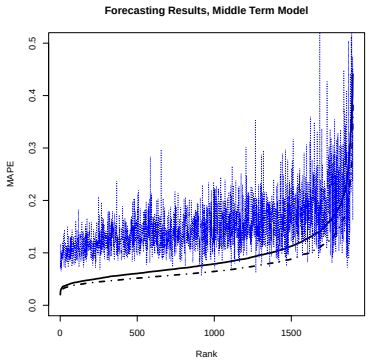

Fig. 4. MAPE obtained on the forecasting set on the 1900 substations. Black line: middle term model ordered from the best -left- to the worst -right- MAPE. Black dashed line : corresponding middle term model with detrending. Blue line : Naive benchmark Y1

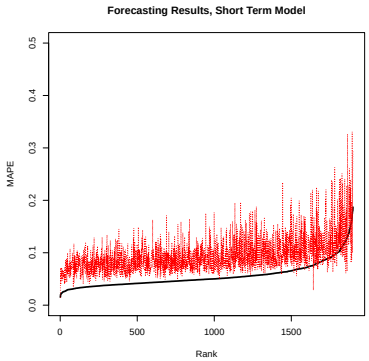

Fig. 5. MAPE obtained on the forecasting set on the 1900 substations with the short term model. Red : Naive benchmark D1. Black : ST model

C. Forecasting in operational condition

We proceed here to the forecast of one day, January the 30^{th} 2013 for a substation in the region of Lyon in France. We use for that the ST model. The estimation is carried out as described in section III-A. The one-day ahead forecast temperature is provided by MeteoFrance and we generate two forecasts: one with the real observed temperature and one with the forecast temperature. We therefore compare two forecasts with predicted temperatures -3 days ahead with approximately 30% error in MAPE- and real ones on the Fig. 6 -in red prediction with forecast temperatures, in blue with real ones and in black is the real curve-. The MAPE of this two forecasts are in Table III-C.

The forecast we obtain are quite good -about 1.5% in MAPE- and we can see that our model fit pretty well the real load curve along the day. Clearly, the weather forecasts have a significant impact on the electricity consumption forecast, specially here for a winter day with a mean temperature of about $7.8\,^{\circ}\mathrm{C}$ where the electrical heating enters as a major component of the load. More precisely, the temperature forecast has a 30% MAPE on this day and it entails a 0.5 % in MAPE of our forecast which is arround 30% of the MAPE of the electricity forecast done with the real temperature. This result is just for one day but give a good intuition of what can

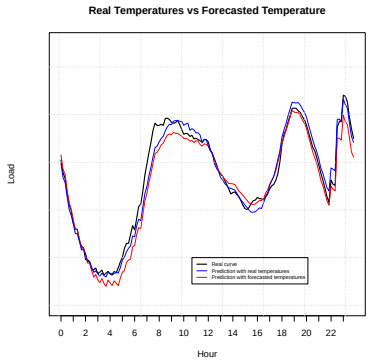

Fig. 6. Forecast with both real and predicted temperatures.

happen in practice using weather forecast. It also highlights the importance of a good weather forecast.

	Real Temperature	Forecast Temperature
MAPE	0.014	0.019

TABLE II
PERFORMANCES WITH REAL AND PREDICTED TEMPERATURES.

D. Focus on one substation

Semi-parametric models produce an interesting interpretation of electricity consumption. Their additive structure allows to separate and represent different features of the signal which is a very important point for industrial applications. Thus the operators can both have a good understanding of the models and develop a practical expertise. As an example we represent on Fig. 7 and on Fig. 8 the temperature effects of 5 instants associated to one substation. Fig. 7 represents the effect of the real temperature for different instants of the day, while Fig. 8 represents the effect of the smoothed temperature. One can notice that temperature has different effects depending on the instant. When temperature decreases, both effects tend globally to increase electricity consumption (and so do the lagged effects of the temperature). Nevertheless, during daytime, the effect of the current temperature seems to be more important than during the night when temperatures are cold, which makes sense because people could less reactive to temperature's changes during the night. Note that we projected these effects over a regular grid from -10 °Celsius to 30 °Celsius. On Fig. 7, one can observe the linear approximation of the splines at the extremities to avoid border effects.

Other interesting features are the estimated daily shapes provided by our models. As an example we represent those shapes for 2 substations on the Fig.9. We can clearly see the differences between the different days of the week. Obviously, the level of week-ends is lower than week days but also the shapes are different. As an example, the load increase in the morning begins later for Saturdays and Sundays. Tariffs effects are also visible. In France, residential customers can subscribe to a special tariff with two regimes: peak and off-peak, to encourage them to reduce their consumption during peaks.

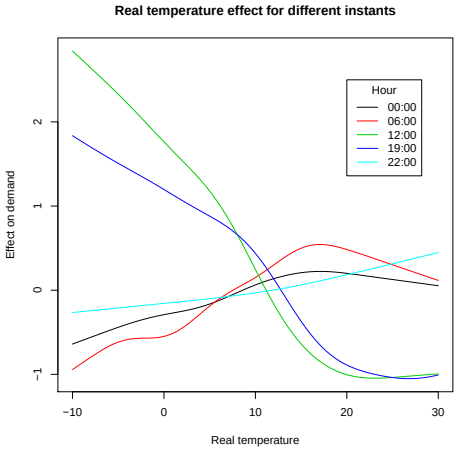

Fig. 7. Effect of Real Temperature (°C) on demand.

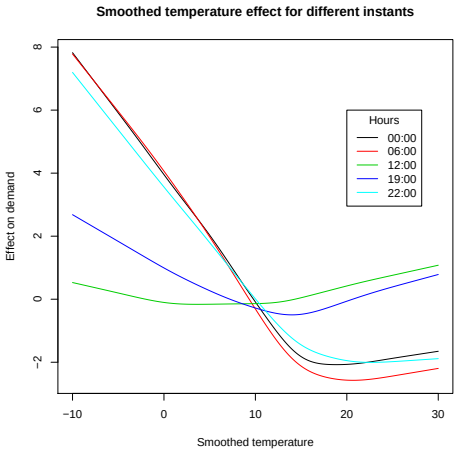

Fig. 8. Effect of Smoothed Temperature (°C) on demand.

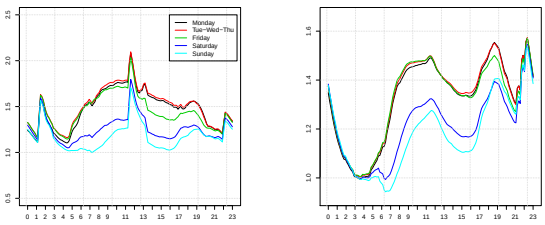

Fig. 9. Estimated day shapes for different days for 2 substations

This peak and off-peak periods can be different for each area but the off-peak tariffs often occur around midday and at night -when the cost of electricity production is low-. We can see for these two substations an increase of the load at midday and after 10.P.M in the evening due to the automatic tripping of some domestic devices -water heating, washing machine etc-. It is also interesting to notice that our models capture two very different tariff effects for the two stations which is a probably due to a the different behaviours of the customers connected to each substation.

IV. Conclusion

In this paper we apply an existing and well known method, semi-parametric additive models, to a new industrial problem: the forecast of a huge number of electricity consumption series on the distribution grid in France. We show the ability of this approach to capture automatically, without any human intervention, the variety of about 2000 consumption series measured on the French grid. The performances for middle term and short term horizons are good and sufficient for the industrial perspectives at this stage. Furthermore, our models are easily interpretable and it is easy to distinguish and estimate the different features of electricity consumption: the effect of special tariff, the electrical heating or cooling, the different seasonality of the signal etc. Another interesting point for applications is that the estimation of these models over big datasets is feasible on a personal computer.

Nevertheless, these models should be improved in many ways, and we identify a few perspectives listed below:

- *automatic selection of covariates*. Some work has to be done to select automatically, for each substations the covariate and the way to include it in the model. This is a natural generalization of our work and it could be solved using recent penalization methods like group lasso.
- *introducing other covariates at a different resolutions*. For this study, we restricted the covariates to temperature over 63 stations and simple calendar variable. This is a good starting point but there is some work to identify new covariates that drive electricity consumption at a local scale. For example, other weather variables like solar radiations, wind direction and speed, humidity could have an impact on local consumption. Demographical and tariff information could probably be pertinent also. All this information could be provided at a fine granularity to improve the models.
- *multivariate analysis*. We don't exploit here the potential dependency between two or more substations. Fitting vectorial GAM models could be a solution to capture and exploit such properties. In that case, we'll probably have to deal with a big data issue and need to develop new estimation algorithms.

References

[1] S. Widergren, K. Subbarao, D. Chassin, J. Fuller, and R. Pratt, "Residential real-time price response simulation," *Proceedings of the 2011 IEEE Power & Energy Society General Meeting*, July 24-28 2011.

[2] J. Nowicka-Zagrajeka and R. Weron, "Modeling electricity loads in california: Arma models with hyperbolic noise," *Signal Processing*, vol. 82, pp. 1903 –1915, 2002.

[3] S. J. Huang and K. R. Shih, "Short-term load forecasting via arma model identification including nongaussian process considerations," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 673–679, 2003.

[4] J. W. Taylor, "Triple seasonal methods for short-term load forecasting," *European Journal of Operational Research*, vol. 204, pp. 139–152, 2010.

[5] J. Taylor, "Short-term load forecasting with exponentially weighted methods," *IEEE Transactions on Power Systems*, 2012.

[6] D. W. Bunn and E. D. Farmer, *Comparative Models for Electrical Load Forecasting*, N. Y. john Wiley, Ed., 1985.

[7] R. Campo and v. P. Ruiz., "Adaptive weather-sensitive short-term load forecasting," *IEEE Transactions on Power Systems*, vol. 3, pp. 592–600, 1987.

[8] R. Ramanathan, R. Engle, C. Granger, F. Vahid-Araghi, and C. Brace., "Short-run forecasts of electricity loads and peaks," *International Journal of Forecasting*, vol. 13, pp. 161–174, 1997.

[9] A. Bruhns, G. Deurveilher, and J. S. Roy, "A non-linear regression model for mid-term load forecasting and improvements in seasonnality," in *The 15th Power Systems Computation Conference, Liege, Belgium*, 2005.

[10] S. J. Harvey, A. C. & Koopman, "Forecasting hourly electricity demand using time-varying splines," *Journal of the American Statistical Association*, vol. 88, pp. 1228–1237, 1993.

[11] a. . K. S. J. Dordonnat, V. and M. Ooms, "Dynamic factors in state-space models for hourly electricity load signal decomposition and forecasting," *IEEE Power & Energy Society*, 2009.

[12] B. J. Chen, M. W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: a study on eunite competition 2001," *IEEE Transaction on Power Systems*, vol. 19, pp. 1821–1830, 2004.

[13] S. Fan and L. Chen, "Short-term load forecasting based on an adaptive hybrid method," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 395–401.

[14] A. Hinojosa and V. Hoese, "Short-term load forecasting using fuzzy inductive reasoning and evolutionary algorithms," *IEEE Transactions on Power Systems*, vol. 25, pp. 565–574, 2010.

[15] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*. Chapman & Hall/CRC, 1990.

[16] S. Wood, *Generalized Additive Models, An Introduction with R*, Chapman and Hall, Eds., 2006.

[17] A. Pierrot and Y. Goude, "Short-term electricity load forecasting with generalized additive models," in *Proceedings of ISAP power, pp 593-600*, 2011.

[18] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Transactions on Power Systems*, vol. 27(1), pp. 134–141, 2012.

[19] R. Nedellec, J. Cugliari, and Y. Goude, "Gefcom2012: Electricity load forecasting and backcasting with semi-parametric models," *submitted to International Journal of Forecasting*, 2013.

[20] S. Wood, "Stable and efficient multiple smoothing parameter estimation for generalized additive models," *Journal of the American Statistical Association*, vol. 99, pp. 673–686, 2004.

[21] ——, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *Journal of the Royal Statistical Society Series (B)*, vol. 73(1), pp. 3–36, 2011.

[22] Craven and Wahba, "Smoothing noisy data with spline functions: estimated the correct degree of smoothing by the method of general cross validation," *Numerische Mathematik*, vol. 31, pp. 377–403, 1979.

[23] S. Wood, "mgcv:gams and generalized ridge regression for r. r news 1(2)," *R News*, vol. 1(2), pp. 20–25, 2001.

Yannig Goude is a research-engineer at EDF R&D since 2008. He obtained his PhD in statistics and probability in 2007 at the university Paris-Sud 11 Orsay. His research interests are electricity load forecasting, more generally time series analysis and forecasting, semi-parametric models and individual sequences.

Raphael Nedellec is a research-engineer at EDF R&D since 2011. He graduated engineer from ENSAI (Ecole National de la statistique et de l'analyse de l'information) in 2011. He's mostly interested in time series, semi-parametric models and their applications to local electricity load modeling.

Nicolas Kong Nicolas Kong is senior Consultant in the Technical Management of ERDF, Network Development Department. He graduated engineer from ENSMA (Ecole Nationale Supérieure de Mécanique et d'Aérotechnique) in 1981 and work for EDF since 1982. He is currently working on the knowledge, the load forecast of power electric system and the impact of new uses and decentralized production on the french network.