
VOLUME 30, ISSUE 2
APRIL–JUNE 2014
ISSN 0169-2070
ELSEVIER
international journal of forecasting
Special Section: Electricity Forecasting
Guest Editors: Tao Hong, Shu Fan and Pierre Pinson
International Institute of Forecasters

International Journal of Forecasting 30 (2014) 375–381

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

GEFCom2012: Electric load forecasting and backcasting with semi-parametric models

Raphael Nedellec [a], Jairo Cugliari [b], Yannig Goude [a,*]

[a] *Electricité de France Research & Development Division, 1 av du Général de Gaulle, 92141 Clamart Cedex, France*

[b] *INRIA, Research Team SELECT, Université Paris Sud, Bât. 425, 91405 Orsay Cedex, France*

A R T I C L E I N F O

Keywords:
Demand forecasting
Forecasting competitions
Multivariate time series
Nonlinear time series
Regression

A B S T R A C T

We sum up the methodology of the team TOLOLO for the Global Energy Forecasting Competition 2012: Load Forecasting. Our strategy consisted of a temporal multi-scale model that combines three components. The first component was a long term trend estimated by means of non-parametric smoothing. The second was a medium term component describing the sensitivity of the electricity demand to the temperature at each time step. We use a generalized additive model to fit this component, using calendar information as well. Finally, a short term component models local behaviours. As the factors that drive this component are unknown, we use a random forest model to estimate it.

1. Introduction

This document briefly presents the methodology which our team, TOLOLO, developed for the GEFCom2012 competition. Based on our experience of load forecasting and successive experiments on this dataset, we introduce a model that can be viewed as a temporal multi-scale model. Such a multi-scale approach has already been studied by Cho, Goude, Brossat, and Yao (2013) in the context of functional regression. It has also been used with semi-parametric regression, which is a popular statistical method (see the seminal work of Hastie & Tibshirani, 1986, 1990) that has proven to be efficient for aggregated (national or regional) electric load forecasting (see Ba, Sinn, Goude, & Pompey, 2012; Fan & Hyndman, 2012; Pierrot & Goude, 2011). Moreover, Goude, Nedellec, and Kong (2013) used this strategy to model 2260 time series recorded at the distribution grid level, and produced forecasts at daily and yearly horizons in an automatic and non-human-supervised way. We therefore model each electric load curve by means of three additive components that describe the long, medium and short patterns. Each region will be modeled using this approach, except for regions 9 and 10, which behave differently and will require some manual intervention.

We denote the electrical load of the region j at time t by Z_t^j, and the total consumption of the area by $Z_t = \sum_{j=1}^{20} Z_t^j$. T_t^i is the temperature at station i at time t.

$$Z_t = Z_t^{\mathrm{lt}} + Z_t^{\mathrm{mt}} + Z_t^{\mathrm{st}}, \qquad (1)$$

where Z_t^{lt} is the long-term part of the load, corresponding to low-frequency variations such as trends, economic effects, and slow changes in electricity usages (due to the increase of electrical heating, heat pumps, etc.). Z_t^{mt} is the medium-term part, being daily to weekly effects. Typically, Z_t^{mt} incorporates all of the meteorological effects (here, the temperature effects) and the calendar effects. The short term part, Z_t^{st}, contains everything that could not be captured on a large temporal scale but could be captured locally in time (close to the date of the prediction). The short-term part largely consists of special events: extreme weather, network reconfigurations, holidays, and so on. As we have three additive components, our approach is

* Corresponding author.
E-mail addresses: raphael.nedellec@edf.fr (R. Nedellec), jairo.cugliari@inria.fr (J. Cugliari), yannig.goude@edf.fr (Y. Goude).

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.07.004

divided into three estimation steps plus the final forecasting step:

1. fit a simple model on monthly data to estimate the trend;
2. fit a detailed middle-term model on the de-trended data;
3. fit a short-term correction model on the residuals;
4. produce a final forecast which is the sum of the forecast trend (extrapolation), the de-trended forecast and the short term correction.

The first step consists of modeling a smooth transformation of the electric load as a function of time and smooth meteorological data, to estimate Z_t^{lt}. In the second step, we model the de-trended electric load $Z_t - \widehat{Z}_t^{\mathrm{lt}}$ as a function of all of the covariates which are available in our dataset that typically drive the electric load at a daily level. We then estimate a local model on the residual signal $Z_t - \widehat{Z}_t^{\mathrm{lt}} - \widehat{Z}_t^{\mathrm{mt}}$ using only the observations in the neighborhood of the prediction time window.

In Section 2, we briefly recall the different statistical methods which we use to estimate the various components. A full description of our procedure is provided in Section 3. Finally, Section 4 is devoted to the presentation of all of the components that contribute to reducing the values of the error criterion which is used in the competition, using numerical experiments over a validation test.

2. Materials and methods

In this section, we briefly discuss the three statistical methods we use: the long-term forecast uses generalised additive models and kernel regression, the medium-term forecast uses generalised additive models, and the short-term forecast uses random forests.

2.1. Kernel regression

Kernel regression (see for example Hastie, Tibshirani, & Friedman, 2009, Chapter 6) is a non-parametric technique for estimating the function M for the non-linear regression model

$$y = M(x) + \varepsilon. \tag{2}$$

The estimated function M_η is warranted to be smooth and depends on a kernel function K (e.g., a probability density) and the smoothing parameter η, known as the bandwidth. One popular choice is to use Nadaraya–Watson estimators. Then, the estimation over the observations $\{y_i, x_i;\ i = 1, \ldots, n\}$ of M is obtained locally on each point x of the definition domain using only the observations y_i, with a corresponding x_i in the neighbourhood of x. It can be written as

$$M_\eta(y)(x) = \sum_{i=1}^{n} y_i K_\eta(x_i, x) \Big/ \sum_{i=1}^{n} K_\eta(x_i, x). \tag{3}$$

We use the convention that the estimator is zero if the denominator is not positive. Then, the estimator can be seen as a weighted mean of the response variable, where the weight associated with y_i is bigger, the closer x_i is to x.

2.2. Generalized additive models

Consider that we want to fit the following statistical model:

$$y_i = f_1(x_{1,i}) + f_2(x_{2,i}) + \cdots + f_p(x_{p,i}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{4}$$

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

$$f_q(x) = \sum_{j=1}^{k_q} \beta_{q,j} b_j^q(x), \tag{5}$$

where k_q is the dimension of the spline basis for modelling the effect f_q, and $b_j^q(x)$ is the corresponding spline functions, such as thin plate regression splines, B-splines, or cubic regression splines. A classical way to estimate these smooth effects is by penalized regression, and more precisely ridge regression, where we minimize the following criterion:

$$\sum_{i=1}^{n} \left(y_i - \sum_{q=1}^{p} f_q(x_i) \right)^2 + \sum_{q=1}^{p} \lambda_q \int \| f_q''(x) \|^2 dx, \tag{6}$$

where the penalty parameter $\Lambda = (\lambda_1, \ldots, \lambda_p)$ which controls the degree of smoothness of each effect (the higher λ_q is, the smoother f_q is) has to be optimized. We denote by B the matrix formed by the concatenation of the discretized versions of the spline functions b_j^q, evaluated on the observed data points (see Wood, 2006, p. 163). Then, we have to solve the following problem:

$$\min_{\beta,\lambda} \|\|Y - B\beta\|\|^2 + \sum_{q=1}^{p} \lambda_q \beta^T S_q \beta, \tag{7}$$

where β is a vector of the unknown regression parameters and S_q is a smoothing matrix of known coefficients, determined by the spline basis (see for example Wood, 2006, p. 156). This problem is solved using the methodology presented by Wood (2004, 2011), which consists of minimizing the GCV (Generalized Cross Validation) criterion proposed by Craven and Wahba (1979). For this purpose, we will use the R package `mgcv` (see Wood, 2001) that implements this method. The functions b_j^q are penalized thin plate regression splines unless it is explicitly stated otherwise.

2.3. Random forests

The random forest method (see Breiman, 2001) is a machine learning method that allows one to estimate the link function f on the following non-parametric model:

$$y_i = f(x_{1,i}, \ldots, x_{p,i}) + \varepsilon_i, \tag{8}$$

using the data $\{y_i, x_{1,i}, \ldots, x_{p,i};\ i = 1, \ldots, n\}$. The response variables and the covariates can be either continuous or discrete. In this short presentation, we focus on the case where the response is continuous and the covariates are both continuous and discrete. A detailed introduction to the topic is provided by Hastie et al. (2009, Chapter 15).

The estimation is done by averaging many simple tree models. Each of the tree models is a recursive partitioning of the space of covariates, in order to obtain classes of observations that maximize some purity criterion for the response (e.g., reduce the intra-class variance). If the tree models are built to be de-correlated, the averaging step will reduce the variance of the random forest estimator significantly. In order to build a de-correlated tree model, the random forest adds two layers of controlled randomness to the data. The first layer is generated by a bootstrap sampling of the observations, while the second one is produced from a random draw from a subset of covariates on each partitioning step.

For a new vector of covariates $(x_{1,n+1}, \ldots, x_{p,n+1})$, we can predict the value of the non-observed response y_{n+1} using the estimation $\hat{f}_n$ of f, obtained using n data points. To do this, each simple tree gives its prediction, and these are then aggregated using the mean of the individual predictions. Each tree model assigns a class to the new vector of covariates, based on the recursive partitioning. Then, the prediction is the mean value of the responses that correspond to this class.

From a computational point of view, the fit of each independent simple tree model can be performed using parallel computing, which reduces the computation time. We use the `randomForest` package (Liaw & Wiener, 2002) to fit our random forest models.

3. Calculations

In this section, we give the details of our multi-scale model, presenting the three components which correspond to the long-, medium- and short-term models.

3.1. Long-term model

We begin by aggregating the data by month (monthly electricity loads and temperature time series for every region and weather station). Denoting these time series by $Z_{t,j}^{\text{monthly}}$ and $T_{t,k}^{\text{monthly}}$, we estimate the following semi-parametric additive model for each region j, using the method presented in Section 2.2:

$$Z_{t,j}^{\text{monthly}} = \sum_{q=1}^{12} c_q \mathbb{I}_{\text{Month}_t=q} + f(T_{t,k_j}^{\text{monthly}}) + \varepsilon_t, \tag{9}$$

where:

- $\mathbb{I}_{\text{Month}_t=q}$ is an indicator variable which is equal to 1 when the month of observation t is q (from 1 to 12), and 0 otherwise.
- f is the effect of the monthly temperature of the station k_j associated with zone j (our procedure for choosing that station will be explained in Section 3.4).

The monthly estimated residuals for zone j can then be obtained as follows:

$$\widehat{\varepsilon}_{t,j}^{\text{monthly}} = z_{t,j}^{\text{monthly}} - \widehat{z}_{t,j}^{\text{monthly}}, \tag{10}$$

where $\widehat{z}_{t,j}^{\text{monthly}}$ is the estimated load from Eq. (9). This represents the long term trend.

We then estimate the smoothed residuals $M_\eta(\widehat{\varepsilon}_j^{\text{monthly}})(t)$, where M_η is defined in Eq. (3), with K being the classical Gaussian kernel, $K_\eta(x, y) = \exp(-\eta(x-y)^2)$, and $\varepsilon_j^{\text{monthly}} = (\varepsilon_{1,j}^{\text{monthly}}, \varepsilon_{2,j}^{\text{monthly}}, \ldots, \varepsilon_{n,j}^{\text{monthly}})$ being the vector of monthly residuals. We test various different smoothing methods: a Gaussian kernel, local polynomial regression, and splines regression. We chose a Gaussian kernel and set the bandwith to 12, such that the quartiles (viewed as probability densities) are at $\pm 0.25 * \eta$. As $M_{12}(\widehat{\varepsilon}_j^{\text{monthly}})$ is a monthly time series, we interpolate it linearly to obtain $\widehat{\text{Tr}}_t$, the trend estimated at a half-hourly frequency.

These smooth residuals are a good estimate of the low frequency effects included in each zone. They contain neither seasonality (annual seasonality) nor weather effects. These residuals are smooth by construction, and thus they are easy to forecast at a one-week horizon with simple constant extrapolation.

3.2. Medium-term model

We apply Eq. (11) to the signal $Z_{t,j}^{\text{det}} = Z_{t,j} - Z_{t,j}^{\text{lt}}$ (note that $Z_{t,j}^{\text{det}}$ is not exactly $Z_{t,j}^{\text{mt}}$, as it also includes $Z_{t,j}^{\text{st}}$, which is in the residual part of the model presented in Eq. (3)). Like Fan and Hyndman (2012) and Pierrot and Goude (2011), we fit one model per instant of the day, so that we have twenty-four models for each zone. We also considered a single model approach like that of Wood (2011), but it led to worse forecasting performances with our present dataset. For the sake of simplicity, we do not make the dependency on the instant of the day explicit, as each time series is composed of the data measured at one instant of the day, and the series are treated independently here. For example, $Z_{t,j}^{10}$, the tth day of electrical consumption for zone j measured at the 10th hour of the day, will be denoted Z_t. The proposed medium-term model is:

$$\begin{aligned} Z_{t,j}^{\text{det}} &= \sum_{q=1}^{11} m_q \mathbb{I}_{\text{DayType}_t=q} + g_1(\theta_{t,k_j}) + g_2(T_{t,k_j}) \\ &\quad + g_3(T_{t-1,k_j}) + g_4(T_{t-2,k_j}) + h(\text{toy}_t) + \varepsilon_t, \end{aligned} \tag{11}$$

where:

- $Z_{t,j}^{\text{det}}$ is the de-trended electrical demand for zone j recorded at time t (for one instant of the day).
- DayType_t is the type of day for observation t: 1 for Sunday, 2 for Monday, 3 for Tuesday, 4 for Wednesday, 5 for Thursday, 6 for Friday, 7 for Saturday, 8 for Christmas and New Year's Day, 9 for Christmas Eve, 10 for Independence Day, and 11 for Thanksgiving.
- θ_t is the smoothed temperature, obtained via exponential smoothing of the real temperature T_{t,k_j} : $\theta_{t,k_j} = (1-0.85)T_{t,k_j} + 0.85\theta_{t-1,k_j}$; T_{t-1,k_j} is the lag 1 temperature (the temperature of the day before at the same hour); and T_{t-2,k_j} is the lag 2 temperature (the temperature at the same hour two days before).
- toy_t is the time of year, which is the position of the observation t within the year (from midnight of January

the 1st to midnight December the 31st), while $h(\text{toy}_t)$ corresponds to the smooth yearly cycle of the load. This function is represented using cyclic cubic regression splines.

Note that we introduce both smooth temperature and lag temperature effects in order to model the inertia of the demand to temperatures (which is due mainly to the building inertia). We optimized the smoothing parameter to 0.85 using the GCV criterion (see Section 2.2), but we noticed that the prediction results are quite insensitive to this parameter around this optimal value.

3.3. Short-term model

A simple visual inspection of the residuals from the medium-term model confirms that there are still some local patterns remaining. As we have already extracted the long-term and medium-term behaviours, what remains can be explained by very localized short-term factors which were not provided for the competition (grid configuration, tariff option, other weather covariate effects, production, etc.). We used a random forests approach to capture these hidden effects, which were due to local (in time) observations. This component can be viewed as a local bias correction of the medium-term model. A short-term model was fitted to each series, and, unlike the medium-term model, we did not build a different model for each hour.

For this purpose, we constructed a data matrix where the responses are the fitted residuals $\hat{\varepsilon}_t$ from Eq. (11) and the covariates are temperature values from all weather stations (both real temperatures and smoothed temperatures), with the following additional covariates: $\texttt{Year}_t$, $\texttt{Month}_t$, $\texttt{Day}_t$, $\texttt{Hour}_t$, $\texttt{Toy}_t$, and $\texttt{DayType}_t$. We did not select the optimal subset of covariates for each series, but we did use a reasonable identical set of them for every station. We chose to include exhaustive temperature information in case the random forest could highlight hidden local correlations.

Two distinct estimation approaches were used for backcasting and forecasting. For backcasting, we used the four weeks surrounding the missing week (two weeks before and two weeks after) as a training set. For the forecasting set, we used the four weeks before the unknown week.

3.4. Weather station selection and temperature forecasting

The organizers of the competition chose not to provide either the real temperatures or temperature forecasts for the last week to be predicted; real temperatures were only available for backcasting. In our applications at *Electricité de France*, weather covariates have been identified as capital information for predicting the future consumption. As a consequence, weather forecasts are bought from the French meteorological company MeteoFrance, and are always provided for electric load forecasting. Thus, our models are designed in such way that weather information is considered to be accurate. In reality, this means that we fit the models on observed meteorological data rather than on forecasts.

For this competition, we proceed in the same way. We first fit our models using the observed temperature data, then plug in forecast temperatures to produce our final forecasts. We tried various different methods for producing temperature forecasts:

- the normal temperature: the mean of the temperature at the same period for the last 4 years.
- SARIMA models.
- semi-parametric models with calendar information and lag temperature as explanatory variables.
- semi-parametric models with SARIMA errors.
- a kernel wavelet functional forecast, which involves finding similar patterns in the past temperature data, based on a wavelet distance (see Antoniadis, Brossat, Cugliari, & Poggi, 2010).

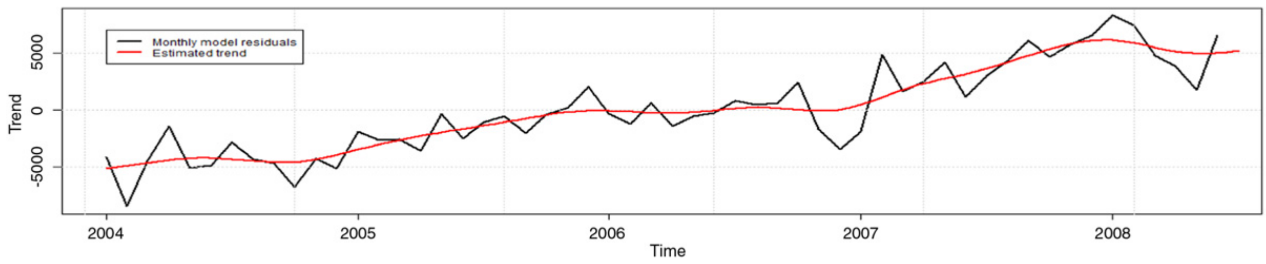
We calibrate these different approaches to the simulation of forecasts on the last week of available data and on the previous years at the same period. Then, each approach in turn is used to generate a new set of forecasts, which are submitted with the same set of backcasts. Finally, we choose the temperature forecasts which minimize the score on the public leaderboard (semi-parametric models with SARIMA errors). Note that we do not claim that this method is the best method for temperature forecasting, but it is the one that gives the best performance for our load forecasting method on the public leaderboard set. In fact, we even think that we overfitted during this process, considering the public and private leaderboards.

In designing the models in Eqs. (9) and (11), we attribute to each zone j only one meteorological station, k_j. The final choice of working with one station per zone was driven by the fact that alternative strategies (e.g., combining the temperature zones by taking the mean, or doing a principal components analysis on the temperature trajectories, with the selection of an optimal number of components) led to worse prediction performances. The optimal temperature zone for each region was found by using a step-wise procedure for Eq. (11) and selecting the weather station by minimizing a V-fold cross validation criterion.

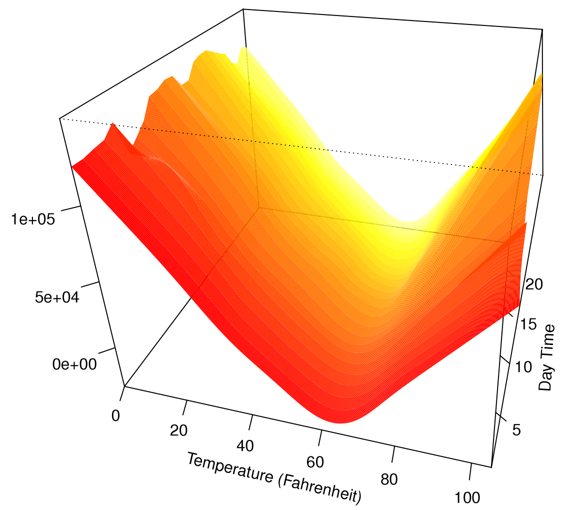
3.5. Validation

In order to validate our different options, we mostly used a kind of V-fold cross validation on the year 2007 (the last entire year), where V corresponds to one week of observations. More precisely, we randomly chose one week per month (so that most of the different properties of the data during the year are captured), excluded it from the estimation set, forecast it with the selected/estimated model, and measured the prediction score using the Root Mean Square Error (RMSE). This method has many advantages. It allows us to conserve much of the data in the estimation set relative to a basic validation set, and it can mimic both backcasting and forecasting problems.

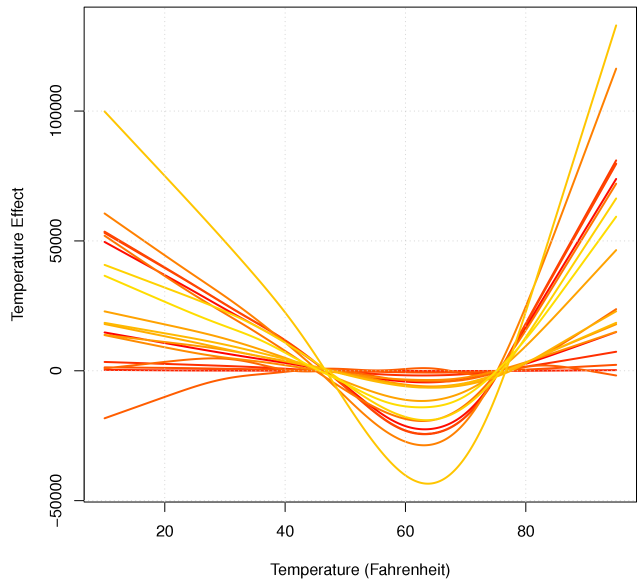
Note that the RMSE is relevant here because it is a good indicator of a global error, and the competition scores are based on (weighted) quadratic errors. We complete it with a precise analysis of our forecasts. We pay particular attention to diagnostic and validation tools of our cross validation errors, such as graphical analysis, in order to detect problems such as forecast biases.



(a) Estimation of the long term trend on the monthly residuals.



(b) Sum of all temperature effects for Zone 3.



(c) Temperature effects for all of the zones at 10 pm.

Fig. 1. Some of the effects of the LT & MT model.

3.6. *Zones requiring special treatment*

As has been mentioned, two zones were dealt with following slightly different procedures: Zones 9 and 10. This is because Zone 9 was found to be very insensitive to temperature effects, and a huge break point occurs in Zone 10 between 2007 and 2008.

The proposed solution is to not include temperature effects in the different estimation steps for Zone 9, and to build two different models for Zone 10: one for the years 2004–2007, and another for the year 2008. Otherwise, we use the approach described in Section 3 for the remaining stations. This was the only special treatment of the data. In particular, no data cleansing method was used.

The prediction for the whole system is obtained using the bottom-up approach: we sum the predictions for the twenty zones.

4. Results

We apply the approach described in Section 3 over the twenty zones and report some of the results here.

4.1. *Estimation of the long- and medium-term models*

We show in Fig. 1(a) the residuals of Eq. (9) and the fit for the long term trend. The curve corresponds to an upward trend that is certainly non-linear.

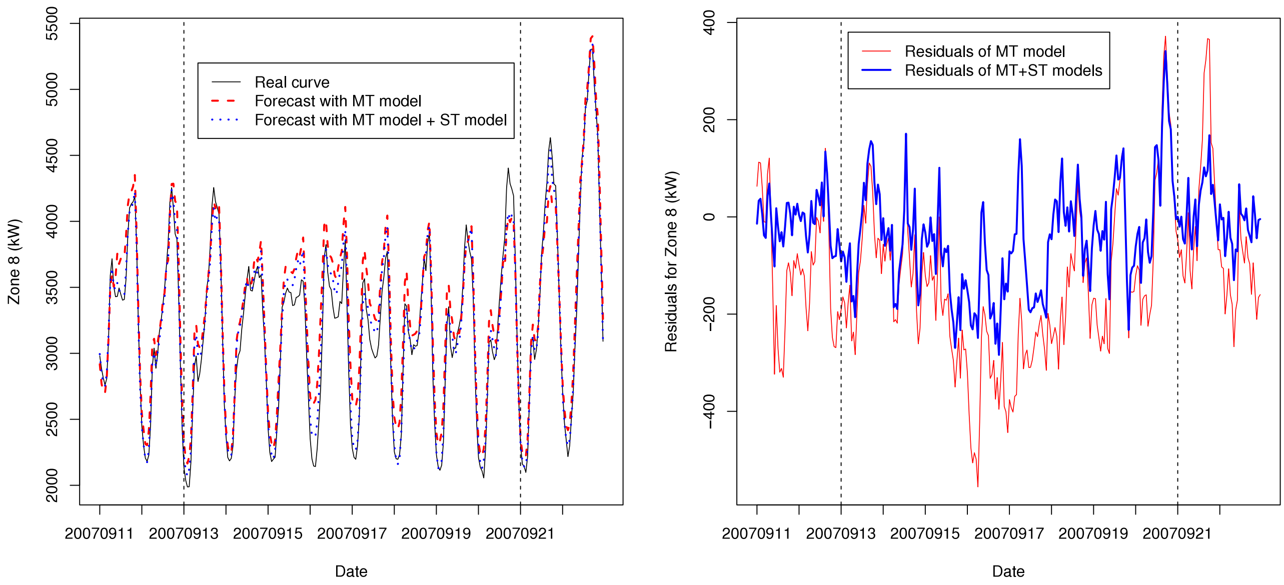
Next, we estimate the 24 models described in Eq. (11) for each zone. One interesting property of semi-parametric models is that the estimated effects of the different covariates are easy to represent and interpret. This is a major point for industrials, who generally want to understand the factors that drive the electrical consumption.

As an example, we show the temperature effects for the different hours of the day for zone 3 in Fig. 1(b). We clearly see both heating and cooling effects, with their evolution through the day, which may be explained by tariff options or consumer habits. We also show the estimated temperature effects for every single zone at 10 pm in Fig. 1(c). It is interesting to observe the similar shapes for the different zones: we can see both cooling and heating effects. We also note that some zones are more sensitive to the temperature than others (which can be explained by the zones being more or less industrial regions), and that our model captures that.

4.2. *Estimation of the short-term part of the electric load*

We now consider the residuals obtained after the two previous steps, and estimate the short term effects presented in Section 3.3.

As an example of short term corrections, Fig. 2 shows a simulated short-term backcast for a week in September 2007. On the left are the real electricity consumption with



(a) Fitted and original values.

(b) Residuals.

Fig. 2. Short term correction effects for zone 8 over a week.

Table 1
Reduction in RMSE from the ST model as a proportion of that of the LT & MT model.

Zone	ST backcast	ST forecast
1	0.95	1.02
2	0.96	0.97
3	0.97	0.95
4	0.95	1.02
5	0.96	0.98
6	0.95	0.93
7	0.96	0.97
8	0.93	0.95
9	1.17	1.07
10	0.88	0.90
11	0.90	0.88
12	0.94	0.90
13	0.94	1.00
14	0.91	0.95
15	0.93	0.99
16	0.93	0.96
17	0.98	0.96
18	0.89	0.92
19	0.93	0.97
20	0.93	0.95

the medium- and short-term forecasts. On the right are the medium- and short-term forecast residuals. There is a clear negative bias (overestimation) in both the medium-term model on the backcast week and the short-term correction. The V-fold cross-validation gains on the prediction RMSE for the short-term (ST) model relative to the medium-term (MT) model are presented in Table 1. We see that we obtain a consistent RMSE gain of about 5% from the inclusion of the ST model component.

5. Conclusion

We have presented the multi-scale model we used in the GEFCom competition. This model, which is based mostly on semi-parametric modeling, is flexible and easy to use, and produces an accurate description of the electricity consumption. Thus, it has a good interpretability, despite the fact that it includes a "black box" component. Actually, this component represents a 5% gain on the prediction performances, which is rather marginal for the full model, but is necessary in order to obtain a better score in the competition.

Table 2
Prediction quality and computation time for one week prediction for the 20 time series (whole system), once each zone has been connected to a weather station.

	LT & MT	ST
RMSE (in kW for whole system)	58 164	53 537
Estimation time (in seconds)	400	80

Globally, our model performs well on the test dataset, a result which is confirmed in the public competition results. We note that if the weather covariates are provided (accurate forecasts, real observations or backcasts), our method provides low error predictions (the best in the competition). The sensitivity of the method to accurate weather forecasts is the price that must be paid for this.

In relation to the computational aspects, only open source software was used to generate our results. The global estimation and prediction times for one series are presented in Table 2. They are obtained using a common laptop (2.4 GHz per core and 4 Gb of RAM running on a 32-bit operating system and using only one computation core). Table 2 presents the prediction quality which the operator can get for the whole system and the time needed to obtain all of the predictions for each week once a weather station has been assigned to each zone. The estimation time of the LT & MT components includes the fit of the long trend and the 24 (one per hour) GAM models. The ST component is the fit of a random forest locally on the 2 weeks (where available) around the prediction. Note also that these times can be reduced considerably by means of parallel computing, since the implementations of the random forest and GAM which we use allow it. In addition, only the ST part of the model has to be re-estimated for each forecast (most of time the LT–MT parameters can remain the same for a few months).

References

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* (pp. 697–704).

Ba, A., Sinn, M., Goude, Y., & Pompey, P. (2012). Adaptive learning of smoothing functions: application to electricity load forecasting. In P. Bartlett, F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in neural information processing systems 25* (pp. 2519–2527).

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

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.

Craven, & Wahba (1979). Smoothing noisy data with spline functions: estimated the correct degree of smoothing by the method of general cross validation. *Numerische Mathematik*, *31*, 377–403.

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

Goude, Y., Nedellec, R., & Kong, N. (2013). Local and middle term electricity load forecasting with semi-parametric additive models. *IEEE Transactions on Smart Grids* (in press).

Hastie, T., & Tibshirani, R. (1986). Generalized additive models (with discussion). *Statistical Science*, *1*, 297–318.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall/CRC.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning. Data mining, inference, and prediction* (2nd ed.). Springer.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Pierrot, A., & Goude, Y. (2011). Short-term electricity load forecasting with generalized additive models. In *Proceedings of ISAP power* (pp. 593–600).

Wood, S. (2001). mgcv:GAMs and generalized ridge regression for R. *R News*, *1*(2), 20–25.

Wood, S. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*, 673–686.

Wood, S. (2006). *Generalized additive models, An introduction with R*. Chapman and Hall.

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B*, *73*(1), 3–36.