School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

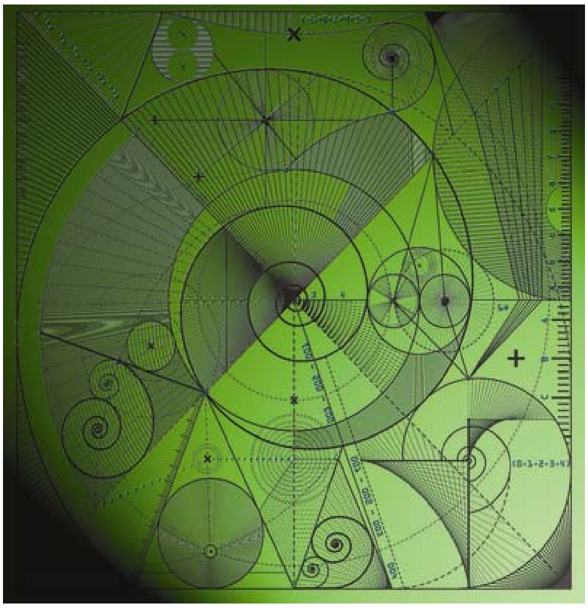
Preprint MPS-2013-05

14 February 2013

Clustering-based improvement of nonparametric functional time series forecasting. Application to intraday household-level load curves

by

Mohamed Chaouch



Clustering-based improvement of nonparametric functional time series forecasting. Application to intraday household-level load curves

Mohamed Chaouch
Centre for the Mathematics of Human behaviour
Department of Mathematics and Statistics
University of Reading, UK
email : *m.chaouch*@reading.ac.uk

February 14, 2013

Abstract

Energy suppliers are facing ever increasing competition, so that factors like quality and continuity of offered services must be properly taken into account. Furthermore, in the last few years, many countries are interested in Renewable Energy's (RE) such as solar and wind. RE resources are mainly used for environmental and economic reasons such as reducing the carbon emission. It also be used to reinforce the electric network especially during high peak periods. However, the injection of such energy resources in the Low-Voltage (LV) network can lead to a high voltage constrains. One possible solution is for electricity companies to motivate customers to use thermal or electric storage devices during high-production periods of PV to foster the integration of RE generation into the network. In this paper, we are interested in forecasting household-level electricity demand which represents a key factor to assure the balance supply/demand in the LV network. We propose a novel methodology able to improve short term functional time series forecasts. An application to the Irish smart meter data set showed the performance of the proposed method for forecasting intra-day household level load curves.

Keywords: Household-level forecasting, nonparametric statistics, unsupervised classification, curve discrimination, functional data, intra-day load curve, smart grids.

1 Introduction and Motivations

In recent years we have seen the arrival of new technologies such as Electrical Vehicle (EV) and electric heating as well as the increase of RE sources such as wind and solar. Therefore, the power grid is going through change. In fact, the stochastic nature of the RE sources will lead the power grid to a highly stochastic system. Within this new context, two main problems arise: (1) because of the electrification of appliances and mobility applications, the peak demand will increase and the load curve shape will change. In fact, at the moment a great deal of attention is attracted by EV, both hybrid and not, that will allow users to recharge their vehicles directly at home. It is therefore important to understand and expect what might be the impact on the power grid capacity of this recharging activity. This question has been studied recently by several authors, see for instance [1], [2], for more details. (2) It is well-known that one of the expected solutions to reduce the peak demand is to reinforce the power grid by RE generation. In fact, one can use energy storage system during surplus energy periods, e.g. PV generation during the day, and discharges during peak load moments. Nevertheless, the integration of a large quantity of RE might lead to a serious problems in the power network. Since most feeders in the LV network have a decreasing cross section, only uni-directional electric charges might be received. Moreover, with the injection of RE sources the electric charge becomes multi-directional depending on consumption and RE production. Obviously, LV networks are designed to support such scenarios but only for short time periods (few hours). It is worth noting that in both cases (1) and (2) explained above, the problem of load forecasting represents a crucial issue for operational planners. In fact, reference [3] shows that short-term load forecasting (STLF) is a key step for proper operation of a battery energy storage system. They used an artificial neural network forecaster for hourly based forecasting of the distributed power generation and load consumption. Recently, reference [4] used updated load forecast for peak shaving and battery lifetime prolonging.

Regularly, the network constraints are evaluated on a specific area in order to prevent over-voltage problems on the network. Very localised consumption and PV/wind production forecasts are needed to detect constraints of current intensity and voltage on each node and each line of the LV network. On the other hand, in order to improve the management of energy demand, the customer is always considered, by the Distribution Network Operators (DNOs), as an important actor that might be involved in the regulation of the electricity network. In fact, to reduce the peak demand, DNOs may ask a selected number of influential customers to reduce their demand during some specific days in the year in conjunction with an incentivised tariff. Another way to make the customer an actor in the management of the energy in the power network is to transform himself as a producer of PV energy for instance. It is well-known that the development of smart meter and its massive deployment in Europe (80% households will be equipped by 2020) and North America allows us to get individual electricity consumption measures on a very fine time scale.

One-day-ahead forecasting of aggregated electricity demand has been widely studied in statistical literature. Different approaches have been proposed to solve this issue. Time series analysis methods like (S)ARIMA models or exponential smoothing can be found in [5]-[9]. Those based on state-space models in [10]. Machine Learning approaches such as artificial neural networks and support vector machine have also been used in [11]-[13]. Among Nonparametric and semiparametric methods, [14] used kernel-based regression model and [15] applied a dimension reduction approach named "Moving Average Variance Estimation" (MAVE, see [16] for more details) to forecast French aggregated load curve. Generalized additive models for short term electricity load curve forecasting were studied in [17]-[18] for instance. For an extensive review on forecasting electric load we refer to, e.g. [19] and [20].

The arrival of smart meters allows us to receive energy demand measurements at a finite number of equidistant time points, e.g. every half hour or every ten minutes. Thus, in order to forecast the load demand of the next day, one has to predict the load demand at forty-eight or one hundred forty four, respectively, time points. From a statistical point of view, it is convenient to think of the daily load demand recorded at these forty-eight or one hundred forty four points as a *segment* and to perform load prediction for the whole segment of time points rather than forecasting the load demand at each one of these time points separately. This implies that we adopt the *functional* time series framework. Functional approach can be also seen as a solution to overcome the problem of incorporating a high number of past values into the statistical model such as in SARIMA model. The idea of forming a functional time series has been considered by several authors, including [21]-[23]. Within this framework of functional time series, several approaches has been proposed e.g. [24] used a semi-functional partial linear model for one-day-ahead forecasting of electricity demand and price, [25] forecasted peak load demand by using functional linear model and [26] developed a functional linear regression model when the response variable and the covariate are both functional. The authors in [27] proposed a nonparametric functional approach based on functional kernel regression estimator. Their developed methodology supposes that all the available information for predicting a segment is essentially contained in the last observed segment. Moreover, an application to sub-aggregated stationary load curve has shown the efficiency of this method with respect to SARIMA model. Recently, [28] performed the approach proposed by [27] by means of a weighted average of past daily load segments. In that case, the past load segments are identified by mean of their closeness to some reference load segment which captures some expected qualitative and quantitative characteristics of the segment to be predicted.

In this paper, we are interested in short term forecasting of household-level intra-day electricity load curve. In contrast to aggregated load curves, which are characterised by their seasonality, regularity and sensibility to meteorological conditions, the household load curves are very volatile, their shape depends mainly on the *customer behaviour* and are less dependent to weather conditions. It is easy to see that the presence of customer behaviour, which is difficult to quantify, as a determinant factor of the shape of the individual load curve makes the issue of household-level forecasting difficult to solve. In this paper, we propose an improved version of the approach proposed by [27] adapted to household-level forecasting. The improvement procedure here is based on the use of an unsupervised clustering step of the historical segments which allows us to find segments describing a common consumption behaviour. Then, we use a nonparametric curve discrimination approach to assign a cluster to the last segment. This step allows us to identify segments which will be used to forecast the target segment.

The paper is organised as follows. In Section 2, we introduce the concept of *functional time series* methodology. Then, we summarize the functional wavelet-kernel approach proposed by [27] and describe the methodology proposed in this paper. Section 3 is devoted to an application of our method to intra-day household level load curve forecasting. A comparison study and an extension to 2000 Irish customers load forecasting is given in the same section. Some concluding remarks are given in Section 4.

2 Functional time series forecasting

Let us consider the household electricity demand as a (real-valued) continuous-time stochastic process $X = (X(t);\ t \in \mathbb{R})$. We are interested in the evolution of this process in the future. We suppose that we observe the process X over an interval $[0, T]$ and one would like to predict the behaviour of X on the entire interval $[T, T+\delta]$, where $\delta > 0$, rather than at specific time points. To this end we can divide the interval $[0, T]$ into subintervals $[\ell\delta, (\ell+1)\delta]$, $\ell = 0, 1, \ldots, k-1$ with $k = T/\delta$, and to consider the (functional-valued) discrete-time stochastic process $\mathcal{S} = (S_n;\ n \in \mathbb{N})$, where $\mathbb{N} = \{1, 2, \ldots\}$, defined by

$$\mathcal{S}_n(t) = X(t + (n-1)\delta); \quad n \in \mathbb{N}, \forall t \in [0, \delta). \tag{1}$$

In this paper we are interested in one-day ahead intra-day load curve forecasting, the segmentation parameter δ corresponds to the daily electricity demand. In practice, the electricity demand is recorded at a finite number of equidistance time points within each day, say $t_1, t_2, \ldots, t_P$, for instance, every half hour (in that case $P = 48$) or every 10 minutes (then $P = 144$). Let us denote by $S_n(t_i)$ the observation at time point t_i, $i = 1, 2, \ldots, P$, within curve $\mathcal{S}_n$, $n \in \mathbb{N}$. We denote by

$$S_n = [S_n(t_1), S_n(t_2), \ldots, S_n(t_P)], \quad n \in \mathbb{N},$$

the segment of the total number of observations of the n-th curve $\mathcal{S}_n$, $n \in \mathbb{N}$. Therefore, given a "sample" $S_1, S_2, \ldots, S_L$ of segments, our purpose is then to predict the *whole next segment* S_{L+1}. In other words we want to predict

$$S_{L+1} = [S_{L+1}(t_1), S_{L+1}(t_2), \ldots, S_{L+1}(t_P)].$$

This forecasting issue has been a subject of several publication in statistical literature. The Functional Autoregressive (FAR) process has been introduced and studied theoretically by [29] and extensively used in both practical and theoretical studies since then, see [30]-[31] among numerous other contributions. Under the FAR model, the best predictor, $\widehat{\mathcal{S}}_{L+1}$, of the curve $\mathcal{S}_{L+1}$, given the historical curves $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_L$ is the conditional mean of $\mathcal{S}_{L+1}$ given the last curve $\mathcal{S}_L$.

2.1 Functional wavelet-kernel approach (FWK)

A nonparametric approach based on kernel method has been developed by [27] to solve the same forecasting issue. In contrast to the FAR model, authors in [27] supposed that the regression operator is unknown and they estimated it *non-parametrically*. More precisely, the prediction of segment S_{L+1} was obtained by kernel smoothing, conditioning on the last observed segment S_L, while the resulting predictor was expressed as a weighted average of the past segments, placing more weight on those segments the preceding of which is "*similar*" to the present one. The notion of *similarity* between two segments (or curves) plays an important role in the calculus of weights and therefore in the prediction of segment S_{L+1}. The authors in [23] defined some semi-metrics which allows to measure the similarity between curves. Another approach based on a distance metric on the discrete wavelet coefficients of suitable wavelet decomposition of the available segments has been proposed by [27]. This approach consists in applying the discrete wavelet transform to the historical segments in order to decompose the temporal information of those segments into discrete wavelet coefficients that are associated both with time and scale. Let us consider two segments S_n and S_m, $n \neq m$, and let $\theta_{j,k}^{(n)}$ and $\theta_{j,k}^{(m)}$ be the discrete wavelet coefficients of S_n and S_m respectively at scale j and location k. Then, the measure the closeness of the two segments S_n and S_m can be summarized in the following two steps:

(a) *at each scale* j, the closeness of the two segments, S_n and S_m, might be defined by measuring the euclidean distance between their discrete wavelet coefficients

$$d_j(\theta^{(n)}, \theta^{(m)}) = \left\{ \sum_{k=0}^{2^j-1} \left(\theta_{j,k}^{(n)} - \theta_{j,k}^{(m)} \right)^2 \right\}^{1/2},$$

(b) to quantify the *similarity* between any two segments S_n and S_m it suffices to combine all scales, then the distance is defined as fellow

$$\mathcal{D}(S_n, S_m) = \sum_{j=j_0}^{J-1} 2^{-j/2} d_j(\theta^{(n)}, \theta^{(m)}).$$

Recall that the predictor $\widehat{S}_{L+1}$ (of the segment S_{L+1}) is a *weighted average* of all segments, then we have

$$\widehat{S}_{L+1}(t_i) = \sum_{\ell=1}^{L-1} w_{L,\ell} \, S_{\ell+1}(t_i), \qquad i = 1, 2, \ldots, P, \tag{2}$$

where the weights $w_{L,\ell} := w(S_L, S_\ell)$, $\ell = 1, 2, \ldots, L-1$ satisfy $w_{L,\ell} \geq 0$, $\ell = 1, 2, \ldots, L-1$ and $\sum_{\ell=1}^{L-1} w_{L,\ell} = 1$. In the nonparametric literature the weights $w_{L,\ell}$, $\ell = 1, 2, \ldots, L-1$, are known as Nadaraya-Watson weights and are defined as follow

$$w_{L,\ell} = \frac{K_h(\mathcal{D}(S_L, S_\ell))}{\sum_{\ell=1}^{L-1} K_h(\mathcal{D}(S_L, S_\ell))}, \quad \ell = 1, 2, \ldots, L-1, \tag{3}$$

where $K_h(\cdot) = h^{-1} K(\cdot/h)$ for some symmetric function $K : \mathbb{R} \to \mathbb{R}$ centered at zero (called Kernel) such that $K(x) \geq 0$, $\int K(x)dx = 1$ and $\int x^2 K(x)dx < \infty$. The tuning parameter h (the so-called bandwidth) controls the effective number of segments for which $w_{L,\ell}$ is positive and therefore the smoothness of the predictor.

Remark: *an implicit assumption was assumed in the approach proposed by [27] which supposes that all the available information for predicting segment* S_{L+1} *is mainly contained in the last observed segment* S_L.

2.2 Clustering-based improvement of the FWK approach (CFWK)

The proposed prediction procedure consists in the following three main steps (a) classification of the sample of historical segments into M (could be fixed or not) clusters containing typical daily load curves. In contrast to aggregated load curves, for which it is easy to observe a common pattern for the working days, week-ends and holidays (see [32] for the use of a type-of-day classification to a national-level load forecasting), household load curves do not contain such kind of similarity between days. For that reason, we use in this step, an unsupervised classification method to identify days describing a common consumption behaviour pattern. (b) Assign to the last observed segment S_L to the most "appropriate" cluster. The main purpose of this step is to find days that contain the same information as the last observed day. In other words, we look, in the historical segments, for those that describe a similar behaviour as what I observe today. (c) Apply the FWK method to forecast the segment S_{L+1} by using segments that belong to the cluster obtained in step (b). The following algorithm describes in more detail how we improve the FWK forecasts using clustering and curve discrimination approaches.

- Step1: *Unsupervised curve classification*

  Suppose that we have $L-1$ historical segments $S_1, S_2, \dots, S_{L-1}$. In this step we are interested in splitting *automatically* these $L-1$ curves into M clusters, say $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_m, \dots, \mathcal{G}_M$. Because we do not have at hand any categorial response variable and the data set are clearly of functional nature then this problem can be seen as an *unsupervised* curves classification. Since the number M of clusters is unknown in our case, then the unsupervised curves classification problem becomes harder to solve. In the statistical literature few authors gave a solution to that problem. These contributions are mainly restricted to the works by [33]-[34] and [35] in which k-means techniques for classification analysis are extended to curves data. In this paper we used the hierarchical algorithm proposed by [36]. The reader is referred to [36] to get more about this algorithm and the methodology behind.

- Step2: *Curve discrimination*

  The curve-discrimination step can be stated as follows. Given the historical segments $S_1, S_2, \dots, S_{L-1}$, then from step1 we know in which cluster each segment belongs to. Let us denote by $\mathcal{G}_\ell$ the cluster of the segment S_ℓ. Assume that each pair of variables $(S_\ell, \mathcal{G}_\ell)$ has the same distribution as a pair of random variables $(S, \mathcal{G})$. Given a new segment S_L (the last observed daily load curve) the purpose now is to identify its class membership. For that we estimate, for each $m \in 1, 2, \dots, M$, the following conditional probability:

  $$p_m(S_L) = \mathbb{P}(\mathcal{G} = \mathcal{G}_m \mid S = S_L). \tag{4}$$

  This means that, whenever the last observed segment is S_L, what is the probability that it belongs to cluster $\mathcal{G}_m$. Observe that this conditional probability, given by

(4) can be seen as a regression function. Therefore, a nonparametric estimator of these probabilities has been proposed in [37]. For all $m \in 1, 2, \dots, M$,

$$\widehat{p}_m(S_L) = \frac{\sum_{\{\ell:\mathcal{G}_\ell = \mathcal{G}_m\}} K_h(\mathcal{D}(S_L, S_\ell))}{\sum_{\ell=1}^{L-1} K_h(\mathcal{D}(S_L, S_\ell))}.$$

Therefore, say $\mathcal{G}_m$, the cluster corresponding to the highest probability. We suppose that $\mathcal{G}_m = \left\{S_1^{(m)}, S_2^{(m)}, \dots, S_{K(m)}^{(m)}\right\}$, where $S_d^{(m)}$, $\forall d = 1, 2, \dots, K(m)$ are the segments that belong to the cluster $\mathcal{G}_m$ and $K(m)$ is the total number of segments in $\mathcal{G}_m$.

- Step3: *Forecasting*

  Using results obtained in step2, we can now build the following sample of segments $\left\{\left(S_d^{(m)}, S_{d+1}^{(m)}\right)\right\}_{d=1,\dots,K(m)}$, where $S_d^{(m)}$ is the segment (corresponding to day d) that belongs to the cluster $\mathcal{G}_m$ and $S_{d+1}^{(m)}$ is the segment observed at the day $d+1$. Observe that $S_{d+1}^{(m)}$ doesn't necessarily belongs to the cluster $\mathcal{G}_m$. Recall that our target is to forecast the segment S_{L+1}. Therefore we propose the following estimator

$$\widehat{S}_{L+1}(t_i) = \sum_{d=1}^{K(m)} w_{L,d}^{(m)}\, S_{d+1}^{(m)}(t_i) \qquad i = 1, 2, \dots, P, \tag{5}$$

  where $w_{L,d}^{(m)} := w(S_L, S_d^{(m)})$, for all $d = 1, 2, \dots, K(m)$, and $w(\cdot,\cdot)$ are as defined in (3).

Remark: *the hierarchical classification and the curve discrimination algorithms have been implemented in* R *language. The program is available on-line through the npfda package*[1].

3 Application to intra-day load forecasting

3.1 Description of the data set

To evaluate the proposed approach to the household-level load curve, we used the smart meter data from the Irish smart meter trial[2]. The data set we used consists of $N = 2000$ *residential customers* with a half-hour electricity demand between 14/07/2009 to 31/12/2010. Figures 1 gives some examples of residential load curves. We can easily observe the high volatility of those curves. In this section, our target is to forecast, one-day ahead, the daily half-hour electricity demand (here $P = 48$) for the 2000 residential customers. We also compare results obtained the proposed method `CFWK` to those obtained by `FWK` approach.

[1] http://www.math.univ-toulouse.fr/staph/npfda/

[2] http://www.ucd.ie/issda/data/commissionforenergyregulation/

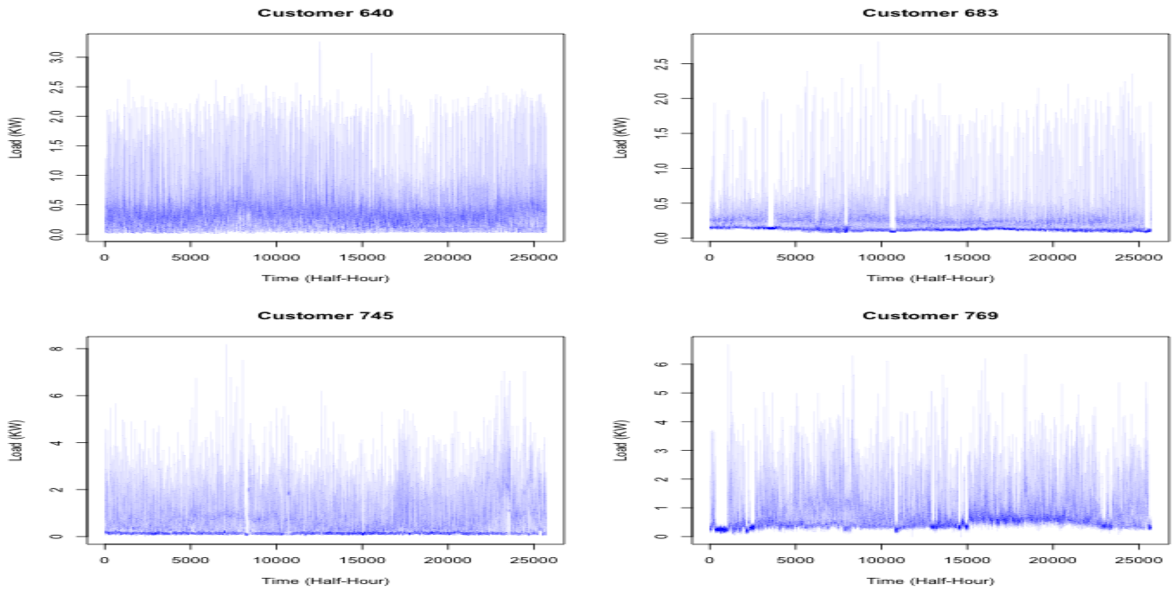


Figure 1: First sample of residential customer's load curve.

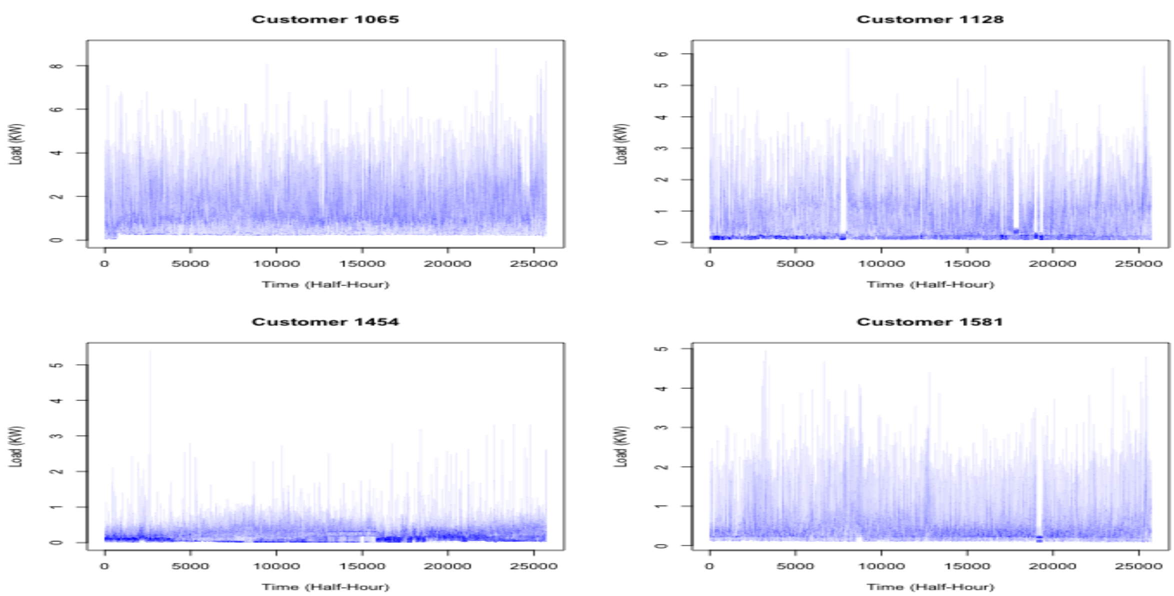


Figure 2: Second sample of residential customer's load curve.

3.2 An illustration of CFWK approach to customer 1016

In this section, we focus on the application of the CFWK method to one randomly chosen customer. We take as example the customer number 1016 in the Irish data. Later, we suggest to extend the results to the entire sample of 2000 customers. Figure 3 (a) shows the original time series which represents the half-hourly electricity demand of this cus-

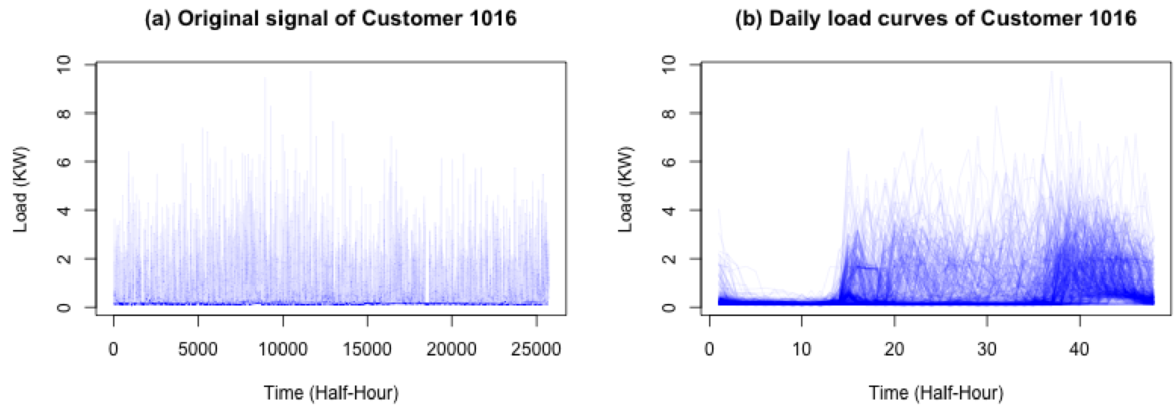


Figure 3: (a) Half-hour electricity demand of customer 1016 between 14/07/2009 and 31/12/2010. (b) A sample of 535 daily load curve (segments) of the same customer.

tomer between 14/07/2009 and 31/12/2010. In Figure 3 (b), we split the original signal into daily load curves ($P = 48$) in order to be able to apply the proposed functional approach. Thus we obtain a sample, say $S_1, S_2, \ldots, S_{535}$, of 535 daily load curves (segments). One can easily observe, from Figure 3 (b), that the electricity demand for that customer is very low between 00:00 and 07:00. Then, the demand increase around 07:30 which corresponds to the morning activity in the household. During the day, consumption decreases in the most of days. Finally, we can observe the classical evening peak demand between 19:00 and 20:00.

To validate our method, we split this sample in two parts. Firstly, denoted by $\mathcal{L} = \{S_1, S_2, \ldots, S_{170}\}$, a learning sample containing daily load curves from 14/07/2009 to 31/12/2009. This sample will be used to build clusters and find the "optimal" bandwidth h. The second part, denoted by $\mathcal{T} = \{S_{171}, S_{172}, \ldots, S_{535}\}$, is the test sample which will be used to compare our forecasts to the observed daily load curves for the period between 01/01/2010 to 31/12/2010 (365 days). Each segment in the test sample $\mathcal{T}$ is forecasted independently. In fact, to forecast the segment S_{171} we use as historical segments $S_1, S_2, \ldots, S_{169}$ and the last observed segment is S_{170}. Then to forecast the segment S_{172}, we consider the historical data $S_1, S_2, \ldots, S_{170}$ and the last observed segment now is S_{171} (the true one and not its forecast). This procedure will be repeated until we forecast all segments that belong to the test sample $\mathcal{T}$. Based on the sample $S_1, S_2, \ldots, S_{170}$ of segments, the goal now is to forecast the segment S_{171} (which corresponds to the 1st Jannuary 2010) using `CFWK` approach. To this end, the following steps are taken:

1. *How many clusters do we have?*

Based on segments $S_1, S_2, \ldots, S_{169}$ and using the hierarchical algorithm proposed by [36],

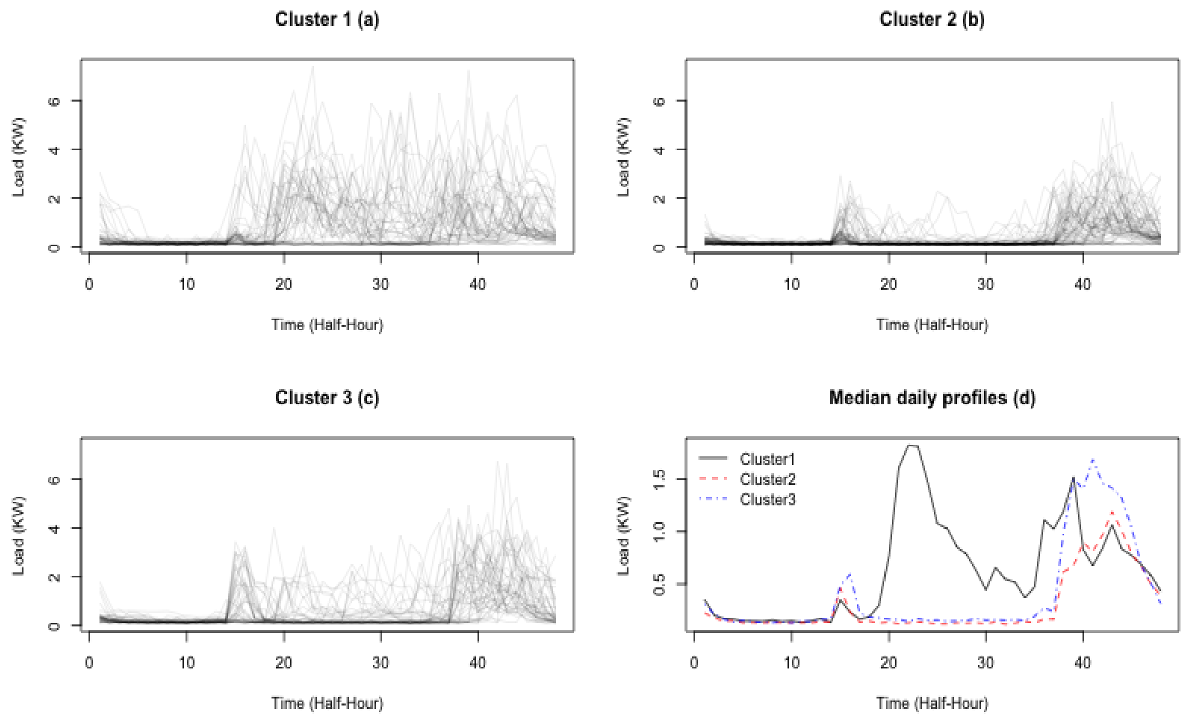


Figure 4: Clusters obtained for cluster 1016.

we find three clusters which are represented in Figure 4 (a)-(c). The median daily profile has been plotted for each cluster in Figure 4 (d). One can easily observe the peak around 07:30 for all clusters. In contrast to the other clusters, cluster 1 contains three other important peaks at 10:30, at 19:30 and at 21:00. We can also observe that in the first cluster 50% of segments have electricity demand between 0.5 KW and 2 KW during the day period from 10:00 to 15:00. On the other hand, in addition to the small peak observed at 10:30, clusters 2 and 3 are characterized by an important peak at the evening (more important for cluster 3) and very small electricity demand during the day.

Table 1 summarizes the type of days within each cluster. We can observe that the first cluster contains mainly week-ends (Saturdays and Sundays) which explain why this customer consumes more electricity during the morning and the afternoon. Working days are in majority within Cluster 2 and for that reason we observe a small peak in the morning (around 07:30), roughly no consumption during the day and then another peak in the evening when people come back to home. Cluster 3 contains mainly Fridays and Sundays (about 50% of the total number of days in that cluster) and some holidays like 25/12/2009. In comparison to cluster 2, this may explain the high values, during the day, of the electricity demand.

Table 1: Type of days within each cluster

	Cluster 1	**Cluster 2**	**Cluster 3**
Mon.	3	**17**	4
Tue.	8	**9**	8
Wed.	1	**15**	8
Thu.	4	**15**	5
Fri.	4	8	**12**
Sat.	**22**	1	1
Sun.	**9**	5	**10**
Total	51	70	48

2. *Which cluster to be assigned to the last observed segment $S_L = S_{170}$?*

A nonparametric curve discrimination method introduced by [37] has been used to assign a cluster for each last observed segment S_L in the training sample. In this example the last segment S_{170} corresponds to the load curve observed on 31/12/2009. Our main task is to predict the corresponding class (which will be in our case cluster 1, 2 or 3) for this segment. To apply the discrimination method explained in sub-section 2.2 several tuning parameters should be fixed. The kernel is chosen to be quadratic and the optimal bandwidth is chosen by the cross-validation method on the k-nearest neighbors (see [23], p. 115 for more details). Another important parameter needs to be fixed is the semi-metric $\mathcal{D}(\cdot,\cdot)$. In this example, because of the roughness of the load curves, we used a semi-metric computed with the functional principal components analysis (see [38]) with an optimal dimension equal to 2. The optimality here was measured with respect to the rate of misclassified curves obtained within the learning sample (17% in this case). Finally, the discrimination method assigned the cluster 1 to the segment S_{170}. This result looks to be compatible with the shape of the load curve of the segment S_{170} presented in Figure 5. In fact, since 31/12/ 2009 is a Christmas Holiday, the customer behaviour in that period is expected to be the same as on the week-end. We can easily see, from Figure 5, the absence of the small peak demand usually observed at 07:30 on working days. We also observe the presence of two important peaks during the day: the first one around 12:30 which corresponds to lunch time and another more important one around 15:30.

3. *Day-head forecasting and validation criteria*

Figure 5: Last observed segment in the training sample for customer 1016: $S_L = S_{170}$ which corresponds to 31/12/2009.

Recall that our main purpose in this example is to forecast the half-hour load curve of the 1st of January 2010 which corresponds to the segment S_{171}. Using results obtained in step 1 and 2 we can then consider the following sample $\left\{\left(S_d^{(1)}, S_{d+1}^{(1)}\right)\right\}_{d=1,2,\dots,51}$, where 51 is the number of segments in cluster 1. Therefore, the forecast of S_{171} is obtained as follow

$$\widehat{S}_{171}(t_i) = \sum_{d=1}^{51} w_{170,d}^{(1)}\, S_{d+1}^{(1)}(t_i) \qquad i = 1, 2, \dots, 48,$$

where $\left\{w_{170,d}^{(1)}\right\}_{d=1,\dots,51}$ are the Nadaraya-Watson weights obtained by measuring the similarity between the load curve observed day-ahead (segment S_{170}) and load curves within the cluster 1. Those weights are determined by equation (3). In this step several tuning parameters should be fixed: the kernel K is chosen to be the gaussian density function and the bandwidth h being selected by the empirical risk of prediction methodology suggested by [39]. In order to extend our study for one-year day-ahead forecasting, we need just to repeat steps one, two and three, 365 times.

The accuracy of each model (CFWK and FWK) will be measured using half-hourly and daily errors. For each fixed day d, with $d = 1, \dots, 365$, in the test sample, the Half-Hour Absolute Errors (HHAE) are defined by

$$\mathsf{HHAE}_d(t_i) = \left|\widehat{S}_d(t_i) - S_d(t_i)\right|, \qquad i = 1, 2, \dots, 48,$$

where $S_d(t_i)$ and $\widehat{S}_d(t_i)$ are the observed and the forecasted value at the i-th half hour in the d-th day of the year to be predicted. The Daily Median Absolute Errors (DMAE) are defined, for all d=1, ..., 365, by

$$\mathsf{DMAE}_d = \mathtt{Median}\{\mathsf{HHAE}_d(t_1), \dots, \mathsf{HHAE}_d(t_{48})\}.$$

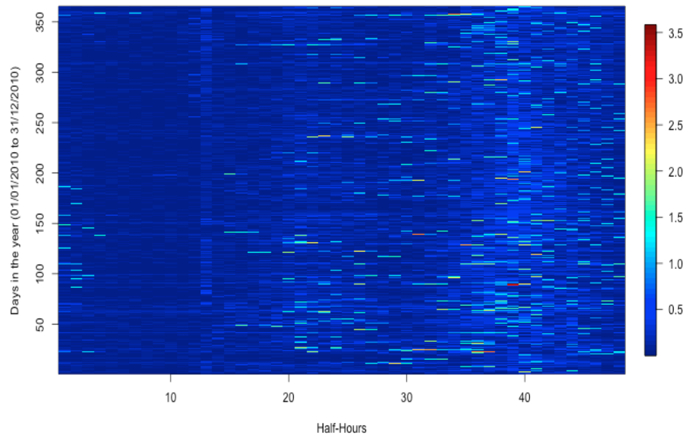


Figure 6: Half-hour Absolute Errors (HHAE) obtained by CFWK method (case of customer 1016).

The choice of the median instead of the mean here is because it is less sensitive to outliers (highest values of HHAE). Figure 6 displays the HHAE errors obtained for one year (from 01/01/2010 to 31/12/2010) day-ahead forecasting using the proposed method CFWK. One can observe that the HHAE are between 0 KW and 3.5 KW and much important errors are obtained during the evening, a period of the day corresponding to a high volatility of demand. A comparison between the proposed method and the FWK approach has been made. Figure 7 displays, for each month in the year, the distribution of the DMAE errors obtained with each method. We can observe clearly that CFWK method provides smaller errors than the FWK one. One can observe that the median of DMAE obtained by CFWK are always less than those obtained by FWK. Other similar results are given in Figure 8 and 9 for customers number 39 and 708.

3.3 Extension of the study to 2000 Irish customers

To measure the efficiency of the proposed approach, we extend the analysis to a sample of 2000 customers randomly selected from the Irish data. We applied the forecasting algorithm given by the CFWK approach to this panel of customers. The tuning parameters has been fixed to be the same for all customers. For each customer, the number of clusters in step 1 of the algorithm has been found automatically. To measure the performance of the proposed method over the panel of 2000 customers, we define the following validation procedure: for each customer, $k = 1, 2, \ldots, 2000$, we calculate, as in the previous subsection, the 365 Daily Median Absolute Errors, say $\mathtt{DMAE}_1^{(k)}, \mathtt{DMAE}_2^{(k)}, \ldots, \mathtt{DMAE}_{365}^{(k)}$. Then, for each day, $d = 1, \ldots, 365$, we determine the Sample

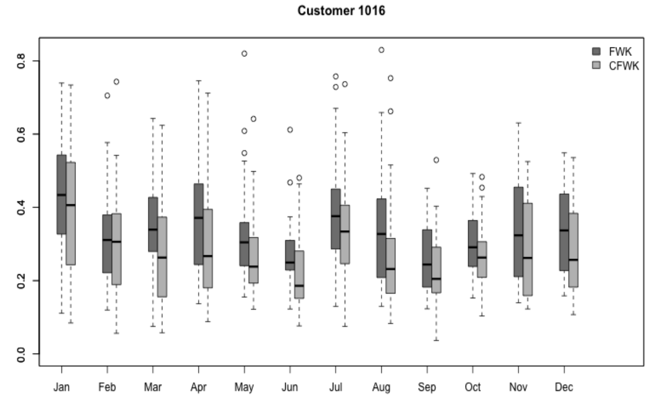


Figure 7: Distribution (by month) of the daily median absolute errors (DMAE) obtained by CFWK and FWK (case of customer 1016).

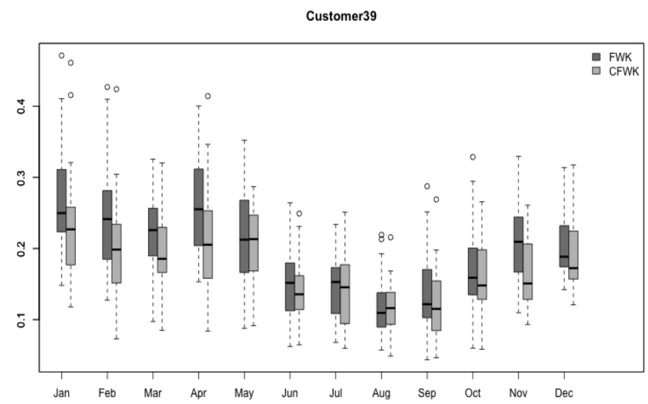


Figure 8: Distribution (by month) of the daily median absolute errors (DMAE) obtained by CFWK and FWK (case of customer 39).

Daily Median Absolute Error (SDMAE) defined as follow:

$$\mathtt{SDMAE}_d = \mathtt{Median}\left\{\mathtt{DMAE}_d^{(1)}, \mathtt{DMAE}_d^{(2)}, \ldots, \mathtt{DMAE}_d^{(2000)}\right\}.$$

Figure 10 displays the distribution, for each month, of the SDMAE errors provided by CFKW and FKW approaches. Table 2 gives numerical summary of results obtained in Figure 10. For instance, if we take the January 2010 as an example, one can observe that, with the CFWK (resp. FWK) approach, 50% (of the 2000 customers in the panel) have a daily median absolute error (DMAE) less than 0.206 KW (resp. 0.222 KW) and 75% of them have a DMAE errors between 0.195 KW and 0.215 KW (resp. 0.211 and 0.235). The same analysis might be made for the other months. Table 2 shows that CFWK approach is more efficient than the FWK one.

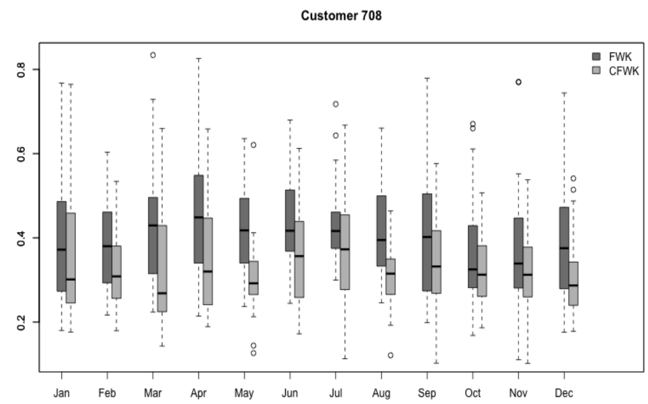


Figure 9: Distribution (by month) of the daily median absolute errors (DMAE) obtained by CFWK and FWK (case of customer 708).

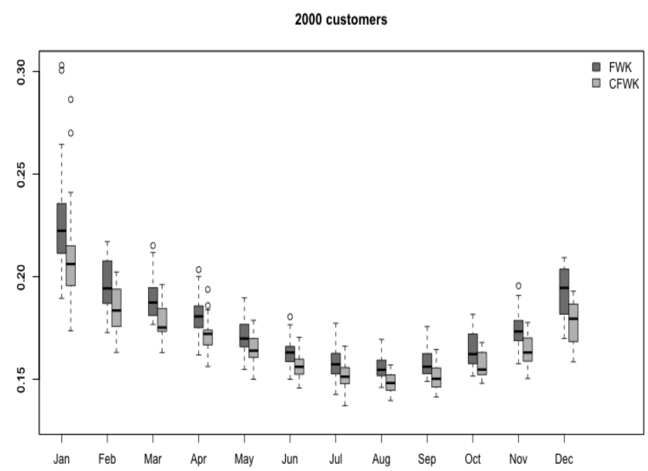


Figure 10: Distribution (by month) of the Sample Daily Median Absolute Errors (SDMAE) obtained by CFWK and FWK.

4 Conclusion

In this paper, a new approach for forecasting functional time series has been proposed. An application to short-term intra-day household-level load curve forecasting has shown the performance of the proposed methodology. The idea behind the use of a classification step is mainly to get a reasonable assumption of stationarity for our time series. Moreover, because the intra-day individual load curve shape is mainly affected by the consumption behaviour of the customer and there is no evidence to identify a common pattern between days we used an unsupervised classification method to find

Table 2: Distribution (by month) of the Sample Daily Median Absolute Errors (SDMAE) obtained by CFWK and FWK.

	CFWK				FWK			
	Mean	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	Mean	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$
Jan.	0.209	0.195	0.206	0.215	0.226	0.211	0.222	0.235
Feb.	0.183	0.175	0.183	0.193	0.196	0.187	0.193	0.205
Mar.	0.177	0.173	0.174	0.184	0.189	0.181	0.187	0.194
Apr.	0.171	0.166	0.172	0.174	0.181	0.175	0.180	0.188
May	0.164	0.160	0.163	0.167	0.171	0.165	0.169	0.175
Jun.	0.156	0.151	0.155	0.159	0.163	0.157	0.162	0.166
Jul.	0.151	0.147	0.151	0.156	0.158	0.152	0.158	0.164
Aug.	0.148	0.144	0.148	0.151	0.155	0.151	0.154	0.158
Sep.	0.151	0.148	0.151	0.156	0.158	0.153	0.156	0.162
Oct.	0.156	0.152	0.153	0.163	0.164	0.157	0.162	0.172
Nov.	0.164	0.158	0.163	0.170	0.174	0.169	0.173	0.178
Dec.	0.178	0.169	0.180	0.186	0.193	0.182	0.198	0.203

similar segments. The numerical results obtained showed that the clustering based approach works very satisfactorily and outperforms the functional wavelet-kernel time series predictor. We note the proposed methodology might be improved by using some daily exogenous functional random variables, like internal/external daily temperature and sunshine curves. Other discrete variables, such as surface of the property, number of electric appliances and number of occupants can also be taken into account which might affect daily individual load demand.

Acknowledgment

The author would like to thank Scottish and Southern Power Distribution SSEPD for support and funding via the New Thames Valley Vision Project (SSET203 - New Thames Valley Vision) - funded through the Low Carbon Network Fund[3].

[3]See http://www.ofgem.gov.uk/networks/elecdist/Icnf/pages/Icnf.aspx

References

[1] K. Clement-Nyns, E. Haesen and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid", *IEEE Trans. Power. Syst.*, vol. 25, no. 1, pp. 371-380, 2010.

[2] S. D. Jenkins, J. R. Rossmaier and M. Ferdowsi, "Utilization and effect of plug-in hybrid electric vehicles in the United States power grid", *IEEE Vehicile Power and Propulsion Conference (VPPC)*, Harbin, China, Sept. 3-5, pp. 1-5, 2008.

[3] B. M. J. Vonk, P. H. Nguyen, M. O. W. Grond, J. G. Slootweg and W. L. Kling, "Improving Short-term load forecasting for a local energy storage system", *Universities Power Engineering Conference (UPEC)*, Sept. 4-7, pp. 1-6, 2012.

[4] B. Guannan, L. Chao, Y. Zhichang and L. Zhigang, "Battery energy storage system load shifting control based on real time load forecasting and dynamic programming", *8th IEEE International Conference on Automation Science and Engineering*, Aug. 20-24, Seoul, Korea, 2012.

[5] J. Nowicka-Zagrajek and R. Weron, "Modeling electricity loads in California: ARMA models with hyperbolic noise", *Signal Processing*, vol. 82, no. 12, pp. 1903-1915, 2002.

[6] S. J. Huang and K. R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations", *IEEE Trans. Power Syst.*, vol. 18, pp. 673-679, 2003.

[7] R. J. Hyndman, A. B. Koehler, R. D. Snyder and S. Grose, "A comparison of univariate methods for fore-casting electricity demand up to a day ahead", *Inter. J. of Forecast.*, vol. 22, pp. 1-16, 2006.

[8] J. W. Taylor, L. M. Menezes and P. E. McSharry, "A state space framework for automatic forecasting using exponential smoothing methods", *Inter. J. of Forecast.*, vol. 18, pp. 439-454, 2002.

[9] J. W. Taylor, "Triple seasonal methods for short-term electricity demand fore- casting", *European. J. of Operat. Research*, vol. 204, pp. 139-152, 2010.

[10] V. Dordonnat, S. J. Koopman, M. Ooms, "Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling", *Comput. Stat. Data Anal.*, vol. 56, no. 11, pp. 3134-3152, 2012.

[11] R. Lamedica, A. Prudenzi, M. Sforna, M. Caciotta and V. Cencelli, "A neural network based technique for short-term forecasting of anomalous load periods", *IEEE Trans. Power Syst.*, vol. 11, no. 4, pp. 1749-1756, 1996.

[12] S. R. Abbas and M. Arif, "Electric load forecasting using support vector machines optimized by genetic algorithm", *INMIC'06 IEEE Multitopic Conference 2006*, Dec. 2006, pp. 395-399.

[13] J. Fidalgo and M. A. Matos, "Forecasting Portugal global load with artificial neural networks", *ICANN2007-International Congress on Artificial Neural Networks 2007*, pp. 728-737.

[14] J. M. Poggi, "Prévision non paramétrique de la consommation électrique", *Rev. de Stat. Appliquée*, vol.12, no.4, pp. 83-98, 1994.

[15] V. Lefieu, "Modèles semi-paramétriques appliqués à la prévision des séries temporelles. Cas de la consommation d'électricité", *Ph.D. dissertation*, 2007.

[16] Y. Xia, H. Tong and W. K., Li, "An adaptive estimation of dimension reduction space", *Jour. of Roy. Stat. Soc. B*, vol. 64, no. 3, pp. 363-410, 2002.

[17] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model", *IEEE Trans. Power Syst.*, vol. 27, pp. 134-141, 2012.

[18] H. Cho, Y. Goude, X. Brossat and Q. Yao, "Modelling and forecasting daily electricity load curves: a hybrid approach", *Preprint*, 2012.

[19] H. K. Alfares and M. Nazeeruddin, "Electric load forecasting: literature survey and classification of methods", *Inter. J. Syst. Sci.*, vol. 33, pp. 23-34, 2002.

[20] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: an evaluation based on European data", *IEEE Trans. Power Syst.*, vol. 22, pp. 2213-2219, 2008.

[21] G. Aneiros-Pérez and P. Vieu, "Nonparametric time series prediction: A semi-functional partial linear modeling", *J. Multivar. Analys.*, vol. 99, no. 5, pp. 834-857, 2008.

[22] A. Antoniadis and T. Sapatinas, "Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes", *J. Multivar. Anal.*, vol.87, pp. 133-158, 2003.

[23] F. Ferraty and Ph. Vieu, *Nonparametric Functional Data Analysis. Theory and Practice*, Springer-Verlag, New-York, 2006.

[24] J. M. Vilar, R. Cao and G. Aneiros, "Forecasting next-day electricity demand and price using nonparametric functional methods", *Inter. J. Electri. Power and Energ. Syst.*, vol. 39, no. 1, pp. 48-55, 2012.

[25] A. Goia, C. May and G. Fused, "Functional clustering and linear regression for peak load forecasting", *Inter. J. Forecast.*, vol. 26, no. 4, pp. 700-711, 2010.

[26] J. Antoch, L. Prchal, M. R. De Rosa and P. Sarda, "Electricity consumption prediction with functional linear regression using spline estimators", *J. Appl. Stat.*, vol. 37, no. 12, pp. 2027-2041, 2010.

[27] A. Antoniadis, E. Paparoditis and T. Sapatinas, "A functional wavelet-kernel approach for time series prediction", *J. Roy. Stat. Soc. B*, vol. 68, pp. 837-857, 2006.

[28] E. Paparoditis and T. Sapatinas, "Short-Term Load Forecasting: The Similar Shape Functional Time Series Predictor", *Preprint*, 2012.

[29] D. Bosq, "Linear processes in functional spaces: theory and applications", Springer, Berlin, 2000.

[30] J. Damon and S. Guillas, "The inclusion of exogenous variables in functional autoregressive ozone forecasting", *Environmetrics*, vol. 13, pp. 759-774, 2002.

[31] D. Didericksen, P. Kokoszka and X. Zhang, "Empirical properties of forecasts with functional autoregressive model", *Comput. Stat.*, vol. 27, no. 2, pp. 285-298, 2012.

[32] A. Antoniadis, X. Brossat, J. Cugliari and J. M. Poggi, "Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. Application à la consommation d'électricité", *J. Soc. Fr. Stat.*, vol. 153, no. 2, pp. 52-78, 2012.

[33] C. Abraham, P. Cornillon, E. Matzner-Löber and N. Molinari, "Unsupervised curves classification using *B*-splines", *Scand. J. Stat.*, vol. 30, pp. 581-595, 2003.

[34] T. Tarpey and K. Kinateder, "Clustering functional data", *J. Classification*, vol. 20, pp. 93-114, 2003.

[35] J. Cuesta-Albertos and R. Fraiman, "Impartial trimmed *k*-means and classification rules for functional data", *Comput. Stat. Data Anal.*, vol. 51, no. 10, pp. 4864-4877, 2007.

[36] S. Dabo-Niang, F. Ferraty and P. Vieu, "Mode estimation for functional random variable and its application for curves classification", *Far East J. Theoret. Statist*, vol. 18, no. 1, pp. 93-119, 2006.

[37] F. Ferraty and P. Vieu, "Curves discrimination: a nonparametric functional approach", *Comput. Stat. Data Anal.*, vol. 44, pp. 161-173, 2003.

[38] P. Hall and M. Hosseini-Nasab, "On properties of functional principal components analysis", *J. Roy. Stat. Soc.*, vol. 68, no. 1, pp. 109-126, 2006.

[39] A. Antoniadis, E. Paparoditis and T. Sapatinas, "Bandwidth selection for functional time series prediction", *Stat. Probab. Letters*, vol. 79, pp. 733-740, 2009.