Short-Term Load Forecasting: The Similar Shape Functional Time-Series Predictor

Efstathios Paparoditis and Theofanis Sapatinas

***Abstract*—A novel functional time-series methodology for short-term load forecasting is introduced. The prediction is performed by means of a weighted average of past daily load segments, the shape of which is similar to the expected shape of the load segment to be predicted. The past load segments are identified from the available history of the observed load segments by means of their closeness to a so-called reference load segment. The latter is selected in a manner that captures the expected qualitative and quantitative characteristics of the load segment to be predicted. As an illustration, the suggested functional time-series forecasting methodology is applied to historical daily load data in Cyprus. Its performance is compared with some recently proposed alternative methodologies for short-term load forecasting.**

***Index Terms*—Functional kernel regression, short-term load forecasting, time series, wavelets.**

I. Introduction

LOAD forecasting is an integrable process in the design of power systems faced by electricity authorities worldwide. It involves accurate predictions of electric load over different time periods in the future. It can be broadly classified as short-term, medium-term, and long-term forecasting, in terms of planning time horizons. Different planning time horizons for these categories seem to exist in the literature. Herein, the load forecasting classification scheme of [1] has been adopted, that is, up to one day for short-term load forecasting (STLF), more than one day up to one year for medium-term load forecasting (MTLF), and more than one year up to 10 years for long-term load forecasting (LTLF).

Competition, the need of saving raw materials that are used in the production of electrical energy, the reduction of emissions, and the avoidance of money wasting in general, are the main reasons that force electricity authorities worldwide to proceed to a better planning for the production of electricity. The main characteristics of the programming are the quantity of electrical energy that needs to be produced and the type of machine that is going to be used. This can be achieved by requiring accurate STLF. This is an important category of load forecasting. It plays a major role in real-time control and security functions for designing larger power systems [2], [3]. Additionally, the particular characteristics of the electricity production process are essential for optimal planning of daily power generation. In connection with the fact that the produced electricity that is not consumed instantly is lost (electricity cannot be stored), it becomes obvious that the right planning of daily power generation must have as a result the avoidance of emergency situations as well as of producing much greater quantities of electricity than the ones needed. It must also offer the capability to electricity authorities worldwide to use as far as possible the low functionality cost machines for covering the electrical energy needed. As a result, it becomes easily understood that an accurate STLF, like the prediction of the consumption of electrical energy of the next day, is an important tool for good power planning. Over the years, a large number of methodologies have been developed to perform STLF. These methodologies are mainly emerged from two different paradigms: classical statistical techniques and computational intelligent techniques. The former techniques include, among others, regression models, ARIMA time series models, Kalman filtering and semi-parametric models [4]–[7], while the latter techniques include, among others, artificial neural networks and expert systems [8]–[10]. Extensive reviews on these techniques for STLF can be found in [11] and [12].

Here, statistical techniques for STLF are considered. Notice that STLF is commonly considered to be a difficult task because the daily load demand is influenced by many factors, viz., weather conditions, holidays, weekdays, weekends, economic conditions, and, last but not least, idiosyncratic and social habits of individuals. Moreover, daily load demand is commonly recorded at a finite number of equidistance time points, every half of each hour or every quarter of each hour. Thus, in order to forecast the load demand of the next day, one has to predict the load demand at 48 or 96, respectively, time points. Therefore, it seems convenient to think of the daily load demand recorded at these time points as a *segment*. Prediction is then performed for the whole segment of time points rather than forecasting the load demand at each one of these time points separately. Thus, the *functional* time series framework is adopted in our approach [13, Ch. 9], [14, Ch. 12].

In this paper, a novel functional time-series methodology for STLF is introduced. The prediction is performed by means of a weighted average of past daily load segments, the shape of which is similar to the expected shape of the load segment to be predicted. The past load segments are identified from the available history of the observed load segments by means of their closeness to a so-called reference load segment. The latter is selected in a manner that captures the expected qualitative and quantitative characteristics of the load segment to be predicted.

Manuscript received August 13, 2012; revised August 29, 2012, August 31, 2012, February 12, 2013, and June 24, 2013; accepted June 28, 2013. Date of publication July 26, 2013; date of current version October 17, 2013. This work was supported in part by the Electricity Authority of Cyprus (EAC) and the Transmission System Operator, Cyprus (TSO). Paper no. TPWRS-00869-2012.

The authors are with the Department of Mathematics and Statistics, University of Cyprus, CY 1678, Nicosia, Cyprus (e-mail: paparoditis@ucy.ac.cy; fanis@ucy.ac.cy).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2013.2272326

0885-8950 © 2013 IEEE

This paper is organized as follows. In Section II, some methodological background is provided and alternative functional time series methodologies that can be applied for STLF are reviewed. The suggested functional time series forecasting methodology is then described in detail. As an application, Section II illustrates its performance by applying it to the historical daily electricity load data in Cyprus. Comparisons with some recently proposed alternative methodologies for STLF are given. Some concluding remarks are provided in Section IV.

II. Functional Time Series Forecasting

Putting the above discussion in a statistical methodological context, one seeks information on the evolution of a (real-valued) continuous-time stochastic process $X = (X(t); t \in R)$ in the future, where $R = (-\infty, \infty)$. Given a trajectory (curve) of X observed on the interval $[0, T]$, one would like to predict the behavior of X on the entire interval $[T, T+\delta]$, where $\delta > 0$. An appropriate approach to this problem is to divide the interval $[0, T]$ into subintervals $[l\delta, (l+1)\delta]$, $l = 0, 1, \ldots, k-1$ with $k = T/\delta$, and to consider the (function-valued) discrete-time stochastic process $\mathcal{S} = (\mathcal{S}_n; n \in N)$, defined by

$$\mathcal{S}_n(t) = X\left(t + (n-1)\delta\right), \quad n \in N, \quad \forall t \in [0, \delta) \qquad (1)$$

where $N = \{1, 2, \ldots\}$. For the specific STLF application in mind, where the aim is one-day ahead prediction, 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 hour ($P = 24$), every half of each hour ($P = 48$) or every quarter of each hour ($P = 96$). Letting $S_n(t_i)$ be the observation at time point t_i, $i = 1, 2, \ldots, P$, within curve $\mathcal{S}_n$, $n \in N$, the segment containing the total number of observations of the nth curve $\mathcal{S}_n$ is denoted by

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

Therefore, given a "sample" $S_1, S_2, \ldots, S_L$ of segments, the aim is to predict the whole next segment S_{L+1}, that is, to predict

$$S_{L+1} = [S_{L+1}(t_1), S_{L+2}(t_2), \ldots, S_{L+1}(t_P)].$$

A. Existing Approaches

Practically, all investigations today for the aforementioned functional time-series prediction problem are for the case where one assumes that the underlying stochastic process is driven by a Hilbert-valued, first-order, autoregressive processes. This implies that the best predictor, $\hat{\mathcal{S}}_{L+1}$, of curve $\mathcal{S}_{L+1}$ given its past history (the "sample" of curves) $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_L$, is the conditional mean of $\mathcal{S}_{L+1}$ given $\mathcal{S}_L$ [13, Ch. 3]. In practice, however, an appropriate version of $\hat{\mathcal{S}}_{L+1}$ is obtained using some regularization on the predictor $\hat{S}_{L+1}$ of segment S_{L+1}. In particular, projection, spline, and wavelet based regularization techniques have been developed [13, Ch. 4], [15]–[18].

An alternative approach to this prediction problem was recently suggested [19]. Theses authors developed a predictor via functional wavelet-kernel nonparametric regression estimation techniques, using a conditioning idea. In particular, prediction of segment S_{L+1} was obtained by kernel smoothing and conditioning on the last observed segment S_L. The resulting predictor was then expressed as a weighted average of the past segments, placing more weight on those segments, the preceding of which is similar to the last segment. This functional time-series forecasting methodology is rooted in the ability to find "similar" segments. Considering that segments can be sampled values of quite irregular curves, similarity matching was based on a distance metric on the discrete wavelet coefficients of a suitable wavelet decomposition of the available segments. For a similar approach, see [20] and [14, Ch. 11].

The basic implicit assumption in developing the above functional time series forecasting methodologies is that all relevant information for predicting segment S_{L+1} is essentially contained in the last observed segment, viz., segment S_L. However, it is more appropriate to assume that the profile of the daily electricity load demand depends, in a complicated and unknown way, on a number of quantitative and qualitative variables of the *day to be predicted*. For instance, amongst others, quantitative variables include daily temperature, daily humidity, and daily wind speed while qualitative variables include weekdays, weekends, holidays, and seasonal characteristics. Apparently, this information is not necessarily contained in the behavior of the last (observed) segment S_L. Thus, functional time-series forecasting approaches that are based on such conditioning ideas can ignore important information concerning the segment to be predicted.

A kernel regression estimator when both the response and the explanatory variables are functional was also recently considered [21]. Thinking of the explanatory variables being quantitative, such as, daily temperature, daily humidity, daily wind speed, etc., the resulting functional kernel regression approach could be used for STLF. However, this functional regression-based approach does not appropriately take into account a number of specific factors, the behavior of which turns out to be important for STLF. For instance, the same behavior of the daily temperature as an explanatory variable could lead to a different response, viz., daily load demand, depending on the seasonal characteristics and on other factors, such as, weekdays, weekends, and holidays. This suggests that, in order to perform accurate STLF, one needs to appropriately take into account not only the behavior of some quantitative variables but also qualitative characteristics of the segment to be predicted that *jointly* affect the daily load demand behavior.

As a result of the previous discussion, one identifies "typical" curves of daily load demand behavior which depend on a number of quantitative and qualitative variables. In order to perform accurate STLF, one then has to: 1) identify the appropriate curve of next day's load demand based on next day's behavior or expected behavior of these variables and 2) find in the entire time series history those curves that are similar to the identified one. Since the proposed STLF is looking at the entire past for "shapes" that are similar to the expected "shape" of the day to be predicted, the resulting predictor is called the *similar shape functional time series predictor* (SSP).

B. Similar Shape Functional Time Series Predictor

Based on the previous discussion, it is assumed herein that daily load demand depends on several quantitative and qualitative variables. Suppose that these characteristics result in a typical daily load demand shape, so that each curve $\mathcal{S}_n$ can be expressed as

$$\mathcal{S}_n(t) = \sum_{g \in \mathcal{G}} f_g\left(\mathcal{T}_n(t), \mathcal{H}_n(t), \mathcal{W}_n(t), \ldots\right) I(g_n = g) + \varepsilon_n(t)$$

for $n \in N, t \in [0, \delta)$, where $I(A)$ denotes the indicator function of the set A. The following terms are also defined.

- $\mathcal{G} = \{\mathcal{G}_{1,m_1}, \mathcal{G}_{2,m_2}, \ldots : m_j \in N, j = 1, 2, \ldots\}$ denotes a group of *deterministic* variables, denoting qualitative characteristics while m_j refers to the number of choices the jth qualitative variable allows for.
  $\mathcal{G}_{1,m_1}$ refers to the m_1 different choices that the variable *group membership* allows to take (e.g., weekdays, weekends, or holidays).
  $\mathcal{G}_{2,m_2}$ refers to the m_2 different values that the variable *season membership* allows to take.
  Notice that, in order to identify the seasonal characteristics, it is not only taken into account the "global seasonality" (i.e., autumn, winter, spring, and summer), but also the "local seasonality," which refers to the weather conditions during the very recent past and which affect the behavior of the daily load demand. This is important since, for instance, a period of warm days in the winter causes a different behavior of the daily load demand compared to the one caused by a period of similar warm days in the spring or in the summer. This local seasonality aspects are also one of the reasons why regression based approaches are not very appropriate in this context. For instance, the same value of the daily temperature (the explanatory variable) may cause a different daily load demand (the response variable), depending on these local seasonal characteristics.
- $g_n = (g_{1,n}, g_{2,n}, \ldots)$ denotes the particular value that the array of qualitative characteristics takes for curve $\mathcal{S}_n$, $n \in N$, e.g., $g_{1,n}$ is the value of the variable *group membership* $\mathcal{G}_{1,m_1}$ for curve $\mathcal{S}_n$, $g_{2,n}$ is the value of the variable *season membership* $\mathcal{G}_{2,m_2}$ for the same curve.
- $\mathcal{T}_n(t), \mathcal{H}_n(t), \mathcal{W}_n(t), \ldots$ are exogenous random variables, denoting quantitative characteristics, e.g., $\mathcal{T}_n(t)$ is the *daily temperature* curve, $\mathcal{H}_n(t)$ is the *daily humidity* curve, and $\mathcal{W}_n(t)$ is the *daily wind speed* curve.
  Notice that these curves are function-valued random variables, usually called *functional* random variables in the statistical literature.
- $(\varepsilon_n(t), n \in N)$ is a sequence of independent and identically distributed (i.i.d) $C([0, \delta))$-valued Gaussian random variables, where $C([0, \delta))$ denotes the space of continuous functions defined on the interval $[0, \delta)$. Furthermore, it is assumed that they have zero mean and finite variance. (Note that, in this case, the errors $\epsilon_n(t_i)$, $i = 1, 2, \ldots, P$, forms a sequence of i.i.d. Gaussian random variables with zero mean and finite variance.) It is also assumed that the noise $\varepsilon_n(t)$ is independent of $\mathcal{T}_s(t), \mathcal{H}_s(t), \mathcal{W}_s(t), \ldots$ for all $s \leq n$, $n \in N$.

In what follows, for simplicity and data availability issues, attention is restricted to the daily temperature curve $\mathcal{T}_n$, $n \in N$, which is one of the main exogenous functional random variables affecting daily load demand, that is, the following model is considered:

$$\mathcal{S}_n(t) = \sum_{g \in \mathcal{G}} f_g\left(\mathcal{T}_n(t)\right) I(g_n = g) + \varepsilon_n(t) \tag{2}$$

for $n \in N$, $t \in [0, \delta)$. (Notice that the suggested methodology can be straightforwardly modified to any available number of exogenous functional random variables.) Letting $T_{L+1}(t_i)$ be the observation at time point t_i, $i = 1, 2, \ldots, P$, within curve $\mathcal{T}_{L+1}$, that is

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

is the segment of the total number of observations of the $(L+1)$th curve $\mathcal{T}_{L+1}$.

Given model (2) and based on the "sample" $S_1, S_2, \ldots, S_L$ of segments, the following algorithm describes in more detail how to construct the suggested SSP, $\hat{S}_{L+1}$, of segment S_{L+1}. In particular, steps 1) and 2) sort out the problem of identifying the appropriate shape of the segment to be predicted while step 3) refers to the calculation of the predictor.

Step 1) For segment S_{L+1}, specify the values of the *group membership* $\mathcal{G}_{1,m_1}$ and *season membership* $\mathcal{G}_{2,m_2}$. While specifying the group membership is easily done based on the particular day to be predicted (weekdays, weekends, holidays), the specification of the "local seasonality" is more difficult and rather arbitrary. This is so, since, as explained before, local seasonality depends on the specific seasonal and weather characteristics of the very recent past of the time series and their stability. In the SLP approach, local seasonality is essential and is taken into account by selecting a small number n_L of past segments that are further considered for selecting what it is called the typical shape or *reference segment*. This is done in step 2) of the algorithm.

Step 2) Determine a relevant load profile by specifying a so-called reference segment, $S^{(Re)}$, as follows.

- Find among the last n_L segments those belonging to the same *group membership* as segment S_{L+1}. Let $\mathcal{C}_{L+1}$ be the set of the selected segments, viz.,

$$\mathcal{C}_{L+1} = \{S_l : g_{1,l} = g_{1,L+1}, 1 \leq l \leq n_L\}. \tag{3}$$

  Notice that the length of local seasonality and its stability is controlled by the parameter n_L which determines how far in the past one goes to select a possible set of segments for specifying the reference segment $S^{(Re)}$.
- Let $\widehat{T}_{L+1}$ be a predictor of segment T_{L+1}.
- Let $\mathcal{D}$ be any of the (equivalent) *distances* in R^P. Then, for each $i = 1, 2, \ldots, P$, the *reference segment* $S^{(Re)}$ is obtained as

$$S^{(Re)}(t_i) = \frac{1}{|C^*|} \sum_{S_l \in C^*} S_l(t_i) \tag{4}$$

where

$$C^* = \left\{ S_l \in \mathcal{C}_{L+1} \text{ and } I\left(\mathcal{D}(T_l, \widehat{T}_{L+1}) \le \delta\right) \right\}$$

and

$$\delta := \delta(L) \ge \min_{\{S_l \in \mathcal{C}_{L+1}\}} \left\{ \mathcal{D}(T_l, \widehat{T}_{L+1}) \right\}. \qquad (5)$$

Notice that the reference segment $S^{(Re)}$ is obtained as a *simple average* of selected segments S_l from the set $\mathcal{C}_{L+1}$. These segments belong to the same group and have the same local seasonal characteristics, while their temperature segments T_l, $l = 1, 2, \ldots, n_L$, are sufficiently close to the predicted temperature $\widehat{T}_{L+1}$ of the segment S_{L+1} to be predicted. This "closeness" is controlled by the parameter δ. It is also pointed out (and this is important for practical applications) that it is not necessary to have the prediction $\widehat{T}_{L+1}$ on the entire set of time points $\{t_1, t_2, \ldots, t_P\}$ in order to specify the set C^*. In fact, one can compare the temperature segments T_l and $\widehat{T}_{L+1}$ using only the subset of time points on which the predictions for the segment T_{L+1} are available (or provided by other sources).

Step 3) Finally, the SSP $\hat{S}_{L+1}$ of the segment S_{L+1} is obtained as

$$\hat{S}_{L+1}(t_i) = \sum_{r=1}^{L} w_r S_r(t_i), \qquad i = 1, 2, \ldots, P \qquad (6)$$

where the weights $w_r = w(S_r, S^{(Re)})$, $r = 1, 2, \ldots, L$, satisfy $w_r \ge 0$, $r = 1, 2, \ldots, L$, and $\sum_{r=1}^{L} w_r = 1$. Following the nonparametric literature, the weights w_r, $r = 1, 2, \ldots, L$, are chosen as

$$w_r = \frac{K_h\left(\mathcal{D}\left(S_r, S^{(Re)}\right)\right)}{\sum_{l=1}^{L} K_h\left(\mathcal{D}\left(S_l, S^{(Re)}\right)\right)}, \qquad r = 1, 2, \ldots, L \qquad (7)$$

with $K_h(\cdot) = h^{-1}K(\cdot/h)$ for some kernel function K, bandwidth $h := h_L$ and distance measure $\mathcal{D}$ between segments.

Notice that the SSP $\hat{S}_{L+1}$ is obtained as a weighted average of past segments, where more weight w_r, $r = 1, 2, \ldots, L$, is placed on the segment the shape of which is similar (in terms of the particular distance $\mathcal{D}$ used) to the shape of the reference segment $S^{(Re)}$. This clarifies the differences between the suggested approach and several other approaches proposed in the literature that are based on conditioning ideas. In the case of conditioning on the last observed segment S_L, the predictor is obtained as a weighted average of past segments, where the weight given to a segment depends on its closeness to the conditioning segment S_L [19]. For the suggested approach, the role of the conditioning segment is taken over by the reference segment $S^{(Re)}$. This reference segment comprehensively contains all relevant information regarding the shape of the daily load demand to be predicted. Thus, the selection of the reference segment $S^{(Re)}$ is essential for the quality of the predictor obtained.

In order to investigate the asymptotic behavior of the SSP predictor, the following set of assumptions 1)–4) (referred to categorically as Assumption 1.1 in the Appendix) are imposed. As $L \to \infty$, we make the following assumptions.

1) $n_L \to \infty$ and $|C^*| \to \infty$.
2) $\delta \to 0$ and $|C^*|\delta \to \infty$.
3) $h \to 0$ and $hL \to \infty$.

Moreover,

4) K is a compactly supported bounded symmetric density.

Requirements 3) and 4) are standard for weak consistency in nonparametric kernel estimation. Assumption 1) implies that the number of segments taken into account to calculate the reference segment $S^{(Re)}$, that is, the number of segments belonging to the set C^*, grows as the sample size increases. Finally, assumption 2) requires that the bandwidth δ, used for obtaining the reference segment $S^{(Re)}$, goes to zero in such a way that the number of segments effectively used in calculating $S^{(Re)}$, viz., $|C^*|\delta$, increases to infinity. This is also a standard assumption for weak consistency in nonparametric kernel estimation.

Now, under the validity of (2), assumptions 1)–4) above and for $\hat{S}_{L+1}$ defined as in (6), if

$$S_{L+1} = f_g(T_{L+1}) + \varepsilon_{L+1}$$

for some $g \in \mathcal{G}$, where

$$f_g(T_{L+1}) = \left[f_g\left(T_{L+1}(t_1)\right), \ldots, f_g\left(T_{L+1}(t_P)\right)\right]$$
$$\varepsilon_{L+1} = \left[\varepsilon_{L+1}(t_1), \ldots, \varepsilon_{L+1}(t_P)\right].$$

Then

$$\mathcal{D}(\hat{S}_{L+1}, f_g(T_{L+1})) \xrightarrow{\mathcal{P}} 0, \text{ as } L \to \infty$$

where $\xrightarrow{\mathcal{P}}$ denotes convergence in probability. This result establishes consistency of the suggested SSP. (The precise mathematical statement and its proof can be found in the Appendix.)

III. Application: Electricity Authority of Cyprus (EAC) Daily Load Data

A. Description of the Data Set

EAC is the organization that is responsible for the generation, transmission, and distribution of electricity in Cyprus. The target of EAC is to provide Cypriots with high quality of safe and reliable services and activities at competitive prices. EAC uses two types of machines to produce electricity. The first type of machine is a *steam turbine* that uses dynamic pressure generated by expanding steam to turn the blades of a turbine. Almost all large nonhydro plants use this system. About 80% of all electric power produced in the world is by use of steam turbines. The advantages of using such a type of machine are the high overall cogeneration efficiencies of up to 80%, the wide range of possible fuels, including waste fuel and biomass, the production of high-temperature/pressure steam, and the established technology. On the contrary, we can mark the low electrical efficiencies, the slow start up times, the poor part load performance, and especially the need for expensive high-pressure

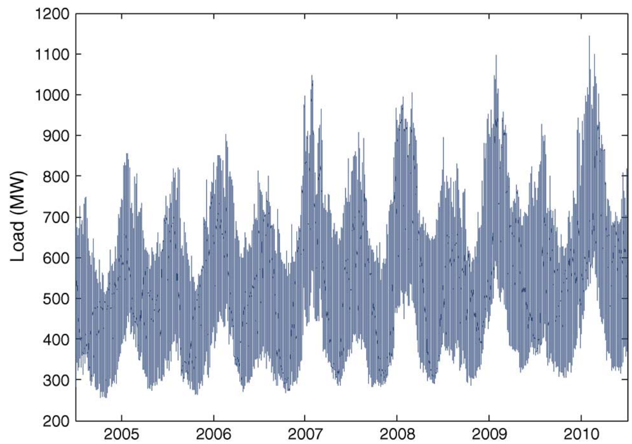

Fig. 1. Electrical power consumption in Cyprus between 1st January 2005 and 1st January 2010, recorded every 15 min.

boilers and other equipment. The second type of machine is a *diesel engine* which uses an electrical generator. Diesel generating sets are used in places without connection to the power grid, as emergency power supply if the grid fails. Of course, they are widely used not only for emergency power but also many of them have a secondary function of feeding power to utility grids either during peak periods or during periods with a shortage of large power generators, although we may say that the cost of their functionality is forbidding.

The best planning for EAC is to avoid these emergency situations and use machines which have low functionality cost and high electrical efficiencies for electricity generation and distribution, which would cover the needs of the whole island. The most important characteristic of the right plan should be the fact that the quantity of electricity produced must not be greater than one needs since the additional electricity produced can not be stored and is lost. An extremely useful tool for the right plan is the accurate prediction of the consumption of the electrical energy of the next day.

Below, the suggested SSP methodology proposed is applied to a set of daily load data, provided by EAC and the Transmission System Operator, Cyprus (TSO), concerning the electrical energy consumption, in megawatts (MW), per 15-min intervals, viz., $P = 96$, for the period from 01/01/2005 to 31/12/2010. This dataset is displayed in Fig. 1.

From this figure, a slightly upward trend can be observed along with a strong periodic component within each year. It is also evident that the electrical energy consumption slightly increases every year and during the summer months attains its maximum. Since the goal is to predict the daily shape of electrical energy consumption, and not the overall trend, the daily curves are rescaled by dividing them by their maximum value. This leads to daily shape curves that vary between zero and one. The aim is then to produce accurate predictions of the shape of the rescaled consumption of next day's electrical energy. The predictor is then transformed to the original scale, by multiplying the resulting SSP by the predicted maximum value of the electrical load of the day to be predicted, provided by TSO. In the next section, we demonstrate how the SSP methodology, proposed above, can be implemented to fulfill this aim.

B. Implementation of the SSP

In our context, the curves $\mathcal{S}_n$, $n = 1, 2, \ldots, L$, that are derived from (1) coincide with the calendar days from the 1st of January 2005 up to the last day $\mathcal{S}_L$, from which observations are available. Hence, $\mathcal{S}_{L+1}$ coincide with the day for which prediction is required. Based on the "sample" $S_1, S_2, \ldots, S_L$ of segments, the goal is to specify the SSP $\hat{S}_{L+1}$. To this end, the following steps are taken.

Step 1) The value of the group membership $\mathcal{G}_{1,m_1}$ for segment S_{L+1} is specified. Feedback from the EAC, have shown that an appropriate grouping of days, with similar shape behavior based on some national characteristics, is the following:

- Group I: Monday, Tuesday, Thursday, Friday.
- Group II: Wednesday.
- Group III: Saturday.
- Group IV: Sunday.

Notice that the work schedule of the public sector is 07:30–18:00 and the commercial shops are open 09:00–13:00 on Wednesdays. On the other hand, the corresponding periods on Mondays, Tuesdays, Thursdays and Fridays are 07:30–14:30 and 09:00–18:00. As a result, Wednesday is grouped separately.

Step 2) To determine the reference segment $S^{(Re)}$, as mentioned previously, our attention has been restricted only to the exogenous random variable $\mathcal{T}_n(t)$, that denotes the daily temperature segment. All days have been found that belong to same group membership $\mathcal{G}_{1,m_1}$ as the segment S_{L+1} to be predicted. The parameter n_L is set equal to $n_L = 14$ if the segment to be predicted corresponds to Monday, Tuesday, Thursday, or Friday and is set equal to $n_L = 28$ if the segment to be predicted corresponds to Wednesday, Saturday, or Sunday. This seems to be appropriate in order to have sufficient information to select the reference segment while at the same time to retain local seasonality. Notice that, in looking back, the procedure selects only days belonging to the same group as the day to be predicted in order to determine the reference segment. Regarding the exogenous random variable $\mathcal{T}_n(t), n \in N$, we use a predictor $\widehat{T}_{L+1}$ of T_{L+1}. As mentioned earlier, it is not necessary to have the prediction $\widehat{T}_{L+1}$ on the entire set of time points $\{t_1, t_2, \ldots, t_P\}$, $P = 96$, to specify the set C^*. In fact, one can compare the temperature segments $T_l, l = 1, 2, \ldots, n_L$, and $\widehat{T}_{L+1}$ using only a subset of time points on which the predictions for the segment T_{L+1} are available (or provided by other sources). More specifically, we have used predictions of next day's temperature at only four time points, that is, those corresponding to 08:00, 12:00, 16:00, and 20:00. The actual temperatures predictions of segment T_{L+1} at these time points were taken from weather forecast Web sites.

Step 3) To further simplify the selection of the reference segment $S^{(Re)}$, the parameter δ in (5) has been set equal to

$$\delta = \min_{\{S_l \in \mathcal{C}_{L+1}\}} \left\{ \mathcal{D}(T_l, \widehat{T}_{L+1}) \right\}$$

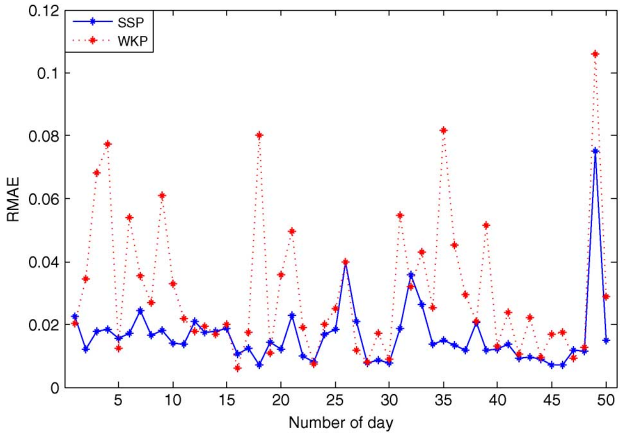

Fig. 2. RMAE for out-of-sample predictions for SSP (solid line) and WKP (dotted line) for the 50 randomly selected days within the year 2010.

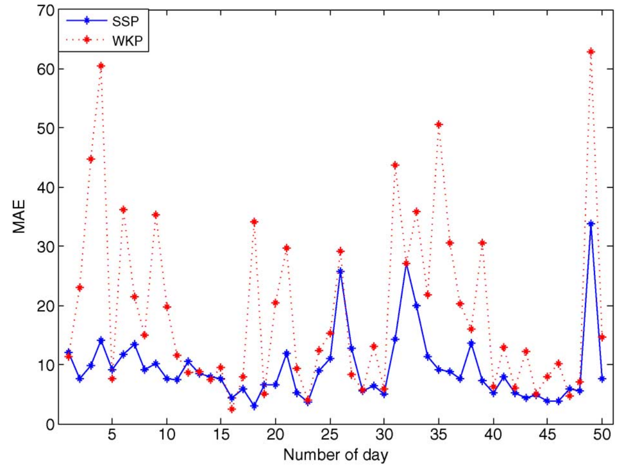

Fig. 3. MAE for out-of-sample predictions for SSP (solid line) and WKP (dotted line) for the 50 randomly selected days within the year 2010.

that is, the reference segment $S^{(Re)}$ is obtained as

$$S^{(Re)} = \arg\min_{\{S_l \in \mathcal{C}_{L+1}\}} \left\{ \mathcal{D}(T_l, \widehat{T}_{L+1}) \right\} \tag{8}$$

where $\mathcal{C}_{L+1}$ is given by (3).

Step 4) The SSP $\hat{S}_{L+1}$ is obtained by (6), where the weights w_r, $r = 1, 2, \ldots, L$, are determined by (7) where the kernel function K is the Gaussian kernel and the bandwidth h are selected by the empirical risk of prediction methodology [22].

Finally, the computational algorithm related to the above implementation as well as the overall numerical study presented above has been carried out in the MATLAB 7.7.0 programming environment.

C. Numerical Results

Based on the above implementation, the SSP functional time-series forecasting methodology is compared with some other forecasting methodologies for STLF.

First, the SSP methodology is compared to the recently established wavelet-kernel functional time series methodology (WKP) [19]. The comparison is restricted to this forecasting methodology, since as it has been demonstrated by these authors in a number of simulated and real-data examples, in terms of functional time series forecasting, WKP outperforms many well-established forecasting methods. This includes a wavelet regularization method, a smoothing spline method, the classical SARIMA model and the Holt–Winters forecasting procedure; see [19]. To do the comparison, both forecasting methods are applied to a randomly selected number of days within the year 2010, from the dataset displayed in Fig. 1. The quality of both SSP and WKP are measured by the *relative mean-absolute error* (RMAE) and the *mean-absolute error* (MAE).

It is evident from the analysis (see also Figs. 2–4) that the SSP clearly outperforms the WKP. It terms of RMAE and MAE, only in eight out of the 50 randomly selected days the SSP performs slightly worse than the WKP. Furthermore, in a large number of days, both the RMAE and MAE of the WKP considerably exceeds the RMAE and MAE of the SSP, as is clearly seen in Figs. 2 and 3. Looking at each day separately, the SSP curves are quite close to the actual load curve. At the same time,

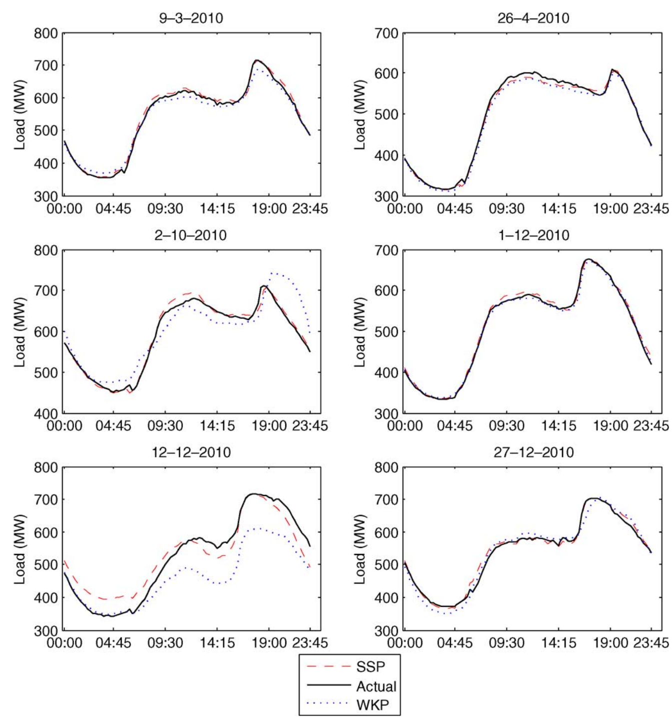

Fig. 4. Actual (solid line) and predicted load using SSP (dashed line) and WKP (dotted line) for six randomly selected days within the year 2010.

there are days where the WKP fails to appropriately capture even the overall behavior of the latter curve (see, e.g., 2 Oct 2010 and 12 Dec 2010).

Next, the SSP is compared with an artificial neural network (ANN) forecasting methodology based on multiple multilayer perceptrons (MMLP) [10]. Notice that the ANN-MMLP methodology has been mainly developed for application to the EAC time-series data, and it concentrates on forecasting only the weekdays (Monday to Friday). Furthermore, the available software program allows the use of the EAC time-series data only for the years 2005 and 2006 in order to make predictions for the year 2007. To do the comparison, the ANN-MMLP predictor has been calculated for all 55 days of the year 2007

TABLE I
RMAE AND MAE FOR OUT-OF-SAMPLE PREDICTIONS FROM FEBRUARY TO DECEMBER 2007

	SSP		ANN-MMLP	
Month	RMAE	MAE	RMAE	MAE
Feb	0.0191	10.162	0.0280	15.569
Mar	0.0510	21.770	0.0519	25.797
Apr	0.0321	13.152	0.0314	15.106
May	0.0171	7.344	0.0176	8.513
Jun	0.0131	7.912	0.0182	11.491
Jul	0.0100	6.761	0.0217	15.038
Aug	0.0164	10.523	0.0268	18.495
Sept	0.0137	7.925	0.0161	9.655
Oct	0.0279	14.578	0.0291	14.589
Nov	0.0226	9.923	0.0144	7.021
Dec	0.0211	10.901	0.0292	16.893
Average	0.0222	10.996	0.0259	14.379

to which the available software program can be applied (predicting five consecutive days in each month, from February to December). The SSP predictor has been calculated for the same consecutive number of days and using the same historical data as the ANN-MMLP predictor. Table I gives the 5-day average RMAE and MAE for each month from February to December 2007. It can be seen that the forecasting performances in the months other than November are generally better.

IV. CONCLUSION

The major contribution of this paper is to introduce a novel functional time-series methodology for short-term load forecasting, named the *functional similar shape time-series predictor*. The predictor was performed by means of a weighted average of past daily load segments, the shape of which is similar to the expected shape of the load segment to be predicted. To quantify this similarity, the notion of *reference segment* was introduced which captures the expected qualitative and quantitative characteristics of the load segment to be predicted. The functional similar shape predictor was theoretically justified by proving a weak consistency property. Furthermore, its usefulness for short-term load forecasting was demonstrated by applying it to historical daily load data in Cyprus. The numerical results obtained showed that the functional similar shape predictor works very satisfactory. It outperforms other statistical methods like the functional wavelet-kernel time series predictor or ANN methods like the predictor based on multiple multilayer perceptrons. Notice that, for simplicity and data availability issues, our attention was restricted to the daily temperature, which is one of the main exogenous functional random variables. The suggested functional time-series methodology for short-term load forecasting, however, can also be modified to take into account other daily exogenous functional random variables. These include humidity, wind speed, sunshine, and market factors, which might affect daily load demand. Although the above modification is straightforward from a theoretical point of view, its practical implementation depends on the availability of the required time series data.

APPENDIX
WEAK CONSISTENCY OF SSP

Theorem 1: Assume that model (3) is true and that Assumption 1.1 is satisfied. Let $\hat{S}_{L+1}$ be defined by (7) and assume that

$$S_{L+1} = f_g(T_{L+1}) + \varepsilon_{L+1} \tag{9}$$

for some $g \in \mathcal{G}$, where

$$\begin{aligned} f_g(T_{L+1}) &= [f_g(T_{L+1}(t_1)), \ldots, f_g(T_{L+1}(t_P))] \\ \varepsilon_{L+1} &= [\varepsilon_{L+1}(t_1), \ldots, \varepsilon_{L+1}(t_P)]. \end{aligned}$$

Then

$$\mathcal{D}\left(\hat{S}_{L+1}, f_g(T_{L+1})\right) \xrightarrow{\mathcal{P}} 0, \text{ as } L \to \infty. \tag{10}$$

Here, $\xrightarrow{\mathcal{P}}$ denotes "convergence in probability" and, below, O_p denotes "bounded in probability".

To prove the above result, we first prove the following auxiliary result, showing that weak consistency of the suggested SSP $\hat{S}_{L+1}$, defined by (7), is equivalent to the weak consistency of the selected reference segment $S^{(Re)}$, defined by (5).

Theorem 2: Assume model (3), that Assumption 1.1 [assumption 3)] is satisfied, and that (9) holds true. Let $S^{(Re)}$ and $\hat{S}_{L+1}$ be defined by (5) and (7), respectively. Then, (10) holds true if and only if

$$\mathcal{D}\left(S^{(Re)}, f_g(T_{L+1})\right) \xrightarrow{\mathcal{P}} 0, \text{ as } L \to \infty. \tag{11}$$

Proof of Theorem 2: We first prove the "if" part of the theorem. Assume that (11) holds true. Since

$$\begin{aligned} \mathcal{D}\left(\hat{S}_{L+1}, f_g(T_{L+1})\right) &\le \mathcal{D}\left(\hat{S}_{L+1}, S^{(Re)}\right) + \mathcal{D}\left(S^{(Re)}, f_g(T_{L+1})\right) \\ &:= A_L + B_L. \end{aligned}$$

Since

$$A_L \le \sum_{r=1}^{L} w_r \mathcal{D}\left(S_r, S^{(Re)}\right) = O_p(h) \to 0, \text{ as } L \to \infty,$$

and

$$B_L \xrightarrow{\mathcal{P}} 0, \text{ as } L \to \infty,$$

the "if" part of the theorem follows.

We now prove the "only if" part. Assume that (10) holds true. Note that, for each $t \in \{t_1, t_2, \ldots, t_P\}$, we have

$$\begin{aligned} \left|\hat{S}_{L+1}(t) - S^{(Re)}(t)\right| &\le \sum_{r=1}^{L} w_r \left|S_r(t) - S^{(Re)}(t)\right| \\ &= O_p(h), \text{ as } L \to \infty. \end{aligned}$$

Furthermore

$$\begin{aligned} \mathcal{D}\left(S^{(Re)}, f_g(T_{L+1})\right) &\le \mathcal{D}\left(S^{(Re)}, \hat{S}_{L+1}\right) \\ &\quad + \mathcal{D}\left(\hat{S}_{L+1}, f_g(T_{L+1})\right) \\ &:= C_L + D_L. \end{aligned}$$

Since

$$C_L = O_p(h) \to 0, \quad \text{as} \quad L \to \infty$$
$$D_L \xrightarrow{\mathcal{P}} 0, \quad \text{as} \quad L \to \infty$$

the desired result follows. ■

Proof of Theorem 1: In view of Theorem 2, it suffices to prove (11). From (5), it is easily seen that $S^{(Re)}$ has the expression of a nonparametric estimator with a uniform kernel. Thus, in view of Assumption 1.1 [assumptions 1), 2), and 4)], (10) holds true by standard weak consistency arguments for nonparametric kernel estimators [23, Ch. 3]. Hence, the theorem follows. ■

ACKNOWLEDGMENT

The authors would like to thank A. Poullikkas, K. Varnavas (EAC), and S. Stavrinos (TSO) for supplying the data used in this work and valuable discussions, Prof. E. Kyriakides and Prof. M. Polykarpou (Department of Electrical and Computer Engineering, University of Cyprus) for their predictions, and M. Frangeskou (Department of Mathematics and Statistics, University of Cyprus) for her help with some of the numerical results. The authors are also grateful to the editor and the three reviewers for many useful comments and suggestions.

REFERENCES

[1] D. Srinivasan and M. A. Lee, "Survey of hybrid fuzzy neural approaches to electric load forecasting," in *Proc. IEEE Intern. Conf. Syst., Man Cybern.*, Vancouver, BC, Canada, 1995, pp. 4004–4008, Pt. 5.

[2] D. Srinivasan, "Evolving artificial neural networks for short-term load forecasting," *Neurocomputing*, vol. 23, no. 1–3, pp. 265–276, Dec. 1998.

[3] P. Laurent, E. Fock, R. N. Randrianarivony, and J.-F. Manicom-Ramsamy, "Bayesian neural network approach to short time load forecasting," *Ener. Conver. Manag.*, vol. 49, no. 5, pp. 1156–1166, May 2007.

[4] O. Hyde and P. F. Hodnett, "An adaptable automated procedure for short-term electricity load forecasting," *IEEE Trans. Power Syst.*, vol. 12, no. 1, pp. 84–94, Feb. 1997.

[5] S. Sargunaraj, D. P. S. Gupta, and S. Devi, "Short-term load forecasting for demand side management," *Proc. Inst. Electr. Eng.—Gen., Transm. Distrib.*, vol. 144, no. 1, pp. 68–74, Jan. 1997.

[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, no. 2, pp. 673–679, May 2003.

[7] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive models," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 134–141, Feb. 2012.

[8] S. Rahman and O. Hazim, "A generalized knowledge-based short-term load-forecasting technique," *IEEE Trans. Power Syst.*, vol. 8, no. 2, pp. 508–514, May 1993.

[9] A. S. AlFuhaid, M. A. El-Sayed, and M. S. Mahmoud, "Cascaded artificial neural networks for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1524–1529, Nov. 1997.

[10] M. Markou, E. Kyriakides, and M. Polykarpou, "24-hour ahead short term load forecasting using multiple MLP," presented at the Int. Workshop on Deregulated Electricity Market Issues in South-Eastern Europe, Nicosia, Cyprus, Sep. 2008, paper no. 137, pp. 1–6, unpublished.

[11] E. Kyriakides and M. Polycarpou, "Short term electric load forecasting: A tutorial," in *Trends in Neural Computing, Studies in Computational Intelligence*, K. Chen and L. Wang, Eds. Berlin, Germany: Springer-Verlag, 2007, vol. 35, ch. 16, pp. 391–418.

[12] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: An evaluation based on European data," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2213–2219, Nov. 2007.

[13] D. Bosq, *Linear Processes in Function Spaces*. New-York, NY, USA: Springer-Verlag, 2000, vol. 149, Lecture Notes in Statistics.

[14] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis*. New York, NY, USA: Springer-Verlag, 2006.

[15] P. C. Besse and H. Cardot, "Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1," *Canad. J. Stat.*, vol. 24, pp. 467–487, 1996.

[16] B. Pumo, "Prediction of continuous time processes by $C[0,1]$-valued autoregressive processes," *Stat. Infer. Stoch. Process.*, vol. 3, pp. 297–309, 1998.

[17] P. C. Besse and D. B. Stephenson, "Autoregressive forecasting of some functional climatic variations," *Scand. J. Stat.*, vol. 27, pp. 673–687, 2000.

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

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

[20] F. Ferraty, A. Goia, and P. Vieu, "Functional nonparametric model for time series: A fractal approach for dimension reduction," *Test*, vol. 11, pp. 317–344, 2002.

[21] F. Ferraty, A. Laksaci, A. Tadj, and P. Vieu, "Kernel regression with functional response," *Electron. J. Stat.*, vol. 5, pp. 159–171, 2011.

[22] A. Antoniadis, E. Paparoditis, and T. Sapatinas, "Bandwidth selection for functional time series prediction," *Stat. Probabil. Let.*, vol. 79, pp. 733–740, 2009.

[23] D. Bosq, *Nonparametric Statistics for Stochastic Processes*, 2nd ed. New York, NY, USA: Springer-Verlag, 1998, vol. 110, Lecture Notes in Statistics.

Efstathios Paparoditis received the Diploma degree in mathematical economics and Ph.D. degree in mathematical statistics from the Freie Universität Berlin, Berlin, Germany, in 1985 and 1990, respectively.

He is a Professor of statistics with the University of Cyprus, Nicosia, Cyprus. He is a former associate editor of the *Journal of Nonparametric Statistics* and is currently an associate editor of *Journal of Statistical Planning and Inference*, *Journal of Time Series Econometrics*, and *Metrika*. His research interests include time-series analysis, nonparametric statistics, functional data analysis, and bootstrap methods.

Theofanis Sapatinas received the B.Sc. degree in mathematics from the University of Athens, Athens, Greece, in 1989, and the M.Sc. and Ph.D. degrees in statistics from the University of Sheffield, Sheffield, U.K., in 1991 and 1994, respectively.

He is a Professor of statistics with the University of Cyprus, Nicosia, Cyprus. He is a former associate editor of the *Annals of Statistics* and is currently an associate editor of *Electronic Journal of Statistics* and *Journal of Statistical Planning and Inference*. His research interests include nonparametric statistical inference, functional data analysis, and structural properties of probability distributions.