

Modelling and Forecasting Daily Electricity Load via Curve Linear Regression

Haeran Cho, Yannig Goude, Xavier Brossat, and Qiwei Yao

Abstract In this paper, we discuss the problem of short-term electricity load forecasting by regarding electricity load on each day as a curve. The dependence between successive daily loads and other relevant factors such as temperature, is modelled via curve linear regression where both the response and the regressor are functional (curves). The key ingredient of the proposed method is the dimension reduction based on the singular value decomposition in a Hilbert space, which reduces the curve linear regression problem to several ordinary (i.e. scalar) linear regression problems. This method has previously been adopted in the hybrid approach proposed by Cho et al. (J Am Stat Assoc 108:7–21, 2013) for the same purpose, where the curve linear regression modelling was applied to the data after the trend and the seasonality were removed by a generalised additive model fitted at the weekly level. We show that classifying the successive daily loads prior to curve linear regression removes the necessity of such a two-stage approach as well as resulting in reducing the forecasting error by a great margin. The proposed methodology is illustrated using the electricity load dataset collected between 2007 and mid-2012, on which it is compared to the hybrid approach and other available competitors. Finally, various ways for improving the forecasting performance of the curve linear regression technique are discussed.

H. Cho (✉)

School of Mathematics, University of Bristol, Bristol, UK
e-mail: haeran.cho@bristol.ac.uk

Y. Goude • X. Brossat
Électricité de France, Paris, France

Q. Yao
Department of Statistics, London School of Economics, London, UK
Guanghua School of Management, Peking University, Beijing, China

1 Introduction

While there are means for storing and discharging electricity, they cause extra costs as well as being limited to a small capacity compared to the overall electric power consumption. Therefore, it is of great importance for electricity providers to model and forecast electricity loads accurately over short-term (from 1 h to 1 month ahead) and middle-term (from 1 month to 5 years ahead) horizons. The electricity load forecast is an essential entry to the optimisation tools adopted by many energy companies for power system scheduling, and a small improvement in load forecasting can bring in substantial benefits from reducing production costs. Besides, there are further advantages to be gained in the electricity trading market, especially during the peak periods.

The French energy company *Électricité de France* (EDF) manages a large panel of production units across Europe, which includes water dams, nuclear plants, wind turbines, coal and gas plants. Figure 1 shows the electricity consumption of their customers measured over half an hour intervals between 2007 and mid-2012. Note that for confidentiality, we only report the ratio between the load over each half-hour interval, and the maximum load during the period throughout the paper. Based on the vast knowledge on French electricity consumption patterns accumulated over 20 years, EDF has developed a forecasting model which consists of complex regression models based on past loads, temperature, date and calendar events, coupled with classical time series models such as the seasonal ARIMA (SARIMA) [4]. This operational model performs very well, attaining about 1.4 % mean absolute percentage error (see (8)) in forecasting the consumption of EDF customers over one day horizon. Due to its complexity, however, the model may not be well-adapted to constant changes in electricity consumption habits resulted from the opening of new electricity markets, technological innovations and social and economic changes, to name a few.

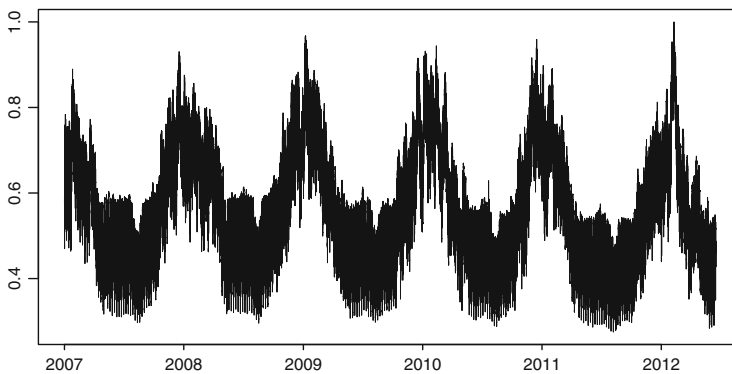


Fig. 1 Electricity consumption of the French customers of EDF measured every half an hour between 2007 and mid-2012

Cho et al. [6] recognised the strategic importance of a forecasting model which was more adaptive to ever-changing electricity consumption environment. Electricity loads exhibit several interesting features at more than one level, as can be seen in Fig. 1, and addressing such multi-level nature of the data, they proposed a hybrid approach which consisted of the following two building blocks:

- Modelling the overall trend and seasonality in the data by fitting a generalised additive model (GAM) to the *weekly* averages of the load, with meteorological factors (e.g., temperature and nebulosity) as explanatory variables;
- Modelling the dependence across successive, de-trended *daily* loads via curve linear regression, where both the response and the regressor are functional (curves), with the load curve on the next day as the response and that on the current day, jointly with the temperature forecast, as the regressor.

By regarding each daily load and temperature as a curve, the proposed curve linear regression modelling takes advantage of the continuity of the curve data in statistical modelling. Moreover, it embeds some nonstationary features, such as daily patterns of electricity loads (see Fig. 2), into a stationary framework in a functional space. Its key ingredient is the dimension reduction based on the singular value decomposition in a Hilbert space, which effectively reduces the curve linear regression problem to several ordinary linear regression problems. Compared to the EDF operational model, the hybrid method does not incorporate much of the data-specific knowledge, while maintaining competitive prediction accuracy when applied to the French electricity consumption data.

While the hybrid approach represents a determined effort in developing an adaptive and widely-applicable forecasting model, it is conceivable that the two-stage procedure may carry over the estimation and the forecasting errors from the

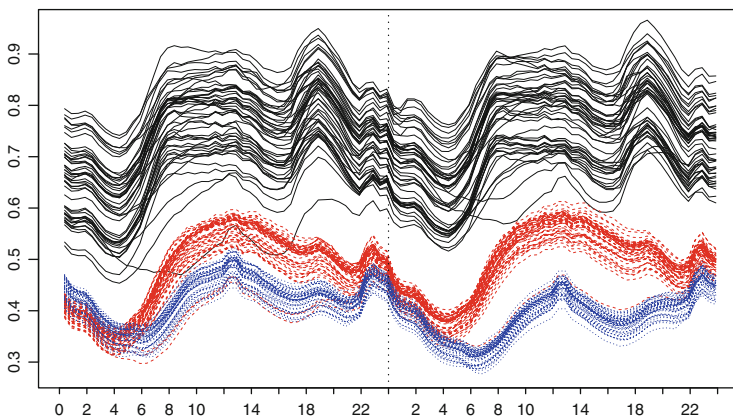


Fig. 2 Electricity loads on Mondays–Tuesdays in January and December (*solid*), Mondays–Tuesdays in June and July (*dashed*) and Saturdays–Sundays in June and July between 2007 and 2012 (*dotted*)

first stage to the next stage, and thus lead to greater forecasting errors. Besides, even after the trend and the seasonality are removed at the weekly level, the daily loads exhibit dependency on calendar variables, such as the corresponding days of a week and the months of a year, both in their profiles and the covariance structure between successive loads. As a solution, [6] proposed to classify the pairs of daily loads into (approximately) homogeneous sub-groups prior to fitting a curve linear regression model, which, as we show, renders the weekly level modelling unnecessary.

Therefore, we focus on the curve linear regression method and its application to the one-day ahead forecasting problem in conjunction with the daily load classification, and investigate whether this simplified approach improves the accuracy and the adaptivity of the forecasting model when compared to the hybrid approach. Besides, the ways of further enhancing its forecasting performance are discussed, such as aggregating several forecasting models resulting from varying choices for the curve regressor.

The rest of the paper is organised as follows. In Sect. 2, we describe the dimension-reduction based curve linear regression technique in a generic setting. Section 3 discusses the application of the proposed approach to electricity load modelling, including the problem of classifying the successive daily load curves. We conduct a comparative study in Sect. 4, where our method and other competitors are applied to predict the daily electricity consumption of EDF customers in France. Finally, we conclude the paper with some remarks on the future research.

2 Curve Linear Regression via Dimension Reduction

Every day at noon, EDF forecasts the half-hourly consumption of electricity for the next 24 h. Viewing that the 48 half-hourly loads are sampled from a curve, we may regard the loads for the next 24 h from the noon of day i as a curve response ($\equiv Y_i(\cdot)$), and let the curve regressor ($\equiv X_i(\cdot)$) contain information such as the loads observed up to the noon of the same day, as well as observed and predicted daily temperature. Then the following curve linear regression model can be adopted to model the dependence between such $Y_i(\cdot)$ and $X_i(\cdot)$:

$$Y_i(u) = \mu_Y(u) + \int_{\mathcal{S}_2} \{X_i(v) - \mu_X(v)\} \beta(u, v) dv + \varepsilon_i(u) \quad \text{for } u \in \mathcal{S}_1, \quad (1)$$

where $\mu_Y(u) = \mathbb{E}\{Y_i(u)\}$, $\mu_X(v) = \mathbb{E}\{X_i(v)\}$ and \mathcal{S}_1 and \mathcal{S}_2 denote the supports of $Y_i(\cdot)$ and $X_i(\cdot)$, respectively. The linear operator β is a regression coefficient function defined on $\mathcal{S}_1 \times \mathcal{S}_2$, and $\varepsilon_i(\cdot)$ is noise satisfying $\mathbb{E}\{\varepsilon_i(u)\} = 0$ for all $u \in \mathcal{S}_1$.

The conventional approach to the linear regression problem in (1) is based on decomposing $Y_i(\cdot)$ and $X_i(\cdot)$ using the Karhunen-Loève expansion, which has been featured predominantly in the functional data analysis literature for dimension reduction. Then the terms from such expansions are modelled using simple linear regression, which is equivalent to the dimension reduction based on principal

component analysis in multivariate analysis. For further references on functional linear models, see e.g. [20, 25] and [12].

Since the principal components do not necessarily represent the directions in which $X_i(\cdot)$ and $Y_i(\cdot)$ are most correlated, [6] presented a novel approach where the singular value decomposition (SVD) in a Hilbert space was adopted to single out the directions upon which the projections of $Y_i(\cdot)$ were most correlated with $X_i(\cdot)$. While closely related to the functional canonical regression method proposed in [15], this approach focuses on regressing $Y_i(\cdot)$ on $X_i(\cdot)$ and thus the two curves are not treated on an equal footing which is different from, and much simpler than, the latter method. In what follows, we lay out the details of the SVD-based curve linear regression method in a generic setting.

Let $\{Y_i(\cdot), X_i(\cdot)\}$, $i = 1, \dots, n$, be a random sample where $Y_i(\cdot) \in \mathcal{L}_2(\mathcal{I}_1)$, $X_i(\cdot) \in \mathcal{L}_2(\mathcal{I}_2)$, and let \mathcal{I}_1 and \mathcal{I}_2 be two compact subsets of \mathbb{R} . We denote by $\mathcal{L}_2(\mathcal{I})$ the Hilbert space consisting of all the square integrable curves defined on the set \mathcal{I} , which is equipped with the inner product $\langle f, g \rangle = \int_{\mathcal{I}} f(u)g(u)du$ for any $f, g \in \mathcal{L}_2(\mathcal{I})$. For now, it is assumed that $\mathbb{E}\{Y_i(u)\} = 0$ for all $u \in \mathcal{I}_1$ and $\mathbb{E}\{X_i(v)\} = 0$ for all $v \in \mathcal{I}_2$. The covariance function between $Y_i(\cdot)$ and $X_i(\cdot)$ is denoted by $\Sigma(u, v) = \text{cov}\{Y_i(u), X_i(v)\}$. Under the assumption

$$\int_{\mathcal{I}_1} \mathbb{E}\{Y_i(u)^2\}du + \int_{\mathcal{I}_2} \mathbb{E}\{X_i(v)^2\}dv < \infty, \quad (2)$$

Σ defines the following two bounded operators between $\mathcal{L}_2(\mathcal{I}_1)$ and $\mathcal{L}_2(\mathcal{I}_2)$,

$$f_1(u) \rightarrow \int_{\mathcal{I}_1} \Sigma(u, v)f_1(u)du \in \mathcal{L}_2(\mathcal{I}_2) \quad \text{and} \quad f_2(v) \rightarrow \int_{\mathcal{I}_2} \Sigma(u, v)f_2(v)dv \in \mathcal{L}_2(\mathcal{I}_1)$$

for any $f_l(\cdot) \in \mathcal{L}_2(\mathcal{I}_l)$, $l = 1, 2$.

Performing the SVD on Σ , we obtain a triple sequence $\{\lambda_j, \varphi_j(\cdot), \psi_j(\cdot)\}$, $j = 1, 2, \dots$ which satisfies

$$\Sigma(u, v) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \varphi_j(u) \psi_j(v), \quad (3)$$

where $\{\varphi_j(\cdot)\}$ is an orthonormal basis of $\mathcal{L}_2(\mathcal{I}_1)$, $\{\psi_j(\cdot)\}$ is that of $\mathcal{L}_2(\mathcal{I}_2)$, and the squared singular values $\{\lambda_j\}$ are ordered in a decreasing manner as

$$\lambda_1 \geq \lambda_2 \geq \dots \geq 0.$$

Further, it holds that for $u \in \mathcal{I}_1$, $v \in \mathcal{I}_2$ and $j = 1, 2, \dots$,

$$\int_{\mathcal{I}_1} M_1(u, u') \varphi_j(u') du' = \lambda_j \varphi_j(u), \quad \int_{\mathcal{I}_2} M_2(v, v') \psi_j(v') dv' = \lambda_j \psi_j(v),$$

where M_l , $l = 1, 2$ are non-negative operators defined on $\mathcal{L}_2(\mathcal{I}_l)$ as

$$M_1(u, u') = \int_{\mathcal{I}_2} \Sigma(u, w) \Sigma(u', w) dw, \quad M_2(v, v') = \int_{\mathcal{I}_1} \Sigma(w, v) \Sigma(w, v') dw.$$

It is clear from the above that λ_j is the j -th largest eigenvalue of M_1 and M_2 , with $\varphi_j(\cdot)$ and $\psi_j(\cdot)$ as the respective eigenfunctions. See [23] for further discussion on the SVD in a Hilbert space.

Since $\{\varphi_j(\cdot)\}$ and $\{\psi_j(\cdot)\}$ are the orthonormal bases of $\mathcal{L}_2(\mathcal{I}_1)$ and $\mathcal{L}_2(\mathcal{I}_2)$, we may write

$$Y_i(u) = \sum_{j=1}^{\infty} \xi_{ij} \varphi_j(u), \quad X_i(v) = \sum_{k=1}^{\infty} \eta_{ik} \psi_k(v), \quad (4)$$

where ξ_{ij} and η_{ik} are random variables defined as $\xi_{ij} = \langle Y_i, \varphi_j \rangle$ and $\eta_{ik} = \langle X_i, \psi_k \rangle$. From (3), it is straightforward to derive that

$$\text{cov}(\xi_{ij}, \eta_{ik}) = \mathbb{E}(\xi_{ij} \eta_{ik}) = \begin{cases} \sqrt{\lambda_j} & \text{when } j = k, \\ 0 & \text{when } j \neq k. \end{cases} \quad (5)$$

The dimensionality of the functional data has been defined in various contexts, e.g. see [13] and [2]. A correlation dimension between the two curves $Y_i(\cdot)$ and $X_i(\cdot)$ was defined in [6] with the squared singular values λ_j .

Definition 1 If $\lambda_r > 0$ and $\lambda_{r+1} = 0$, the (linear) correlation between $Y_i(\cdot)$ and $X_i(\cdot)$ is r -dimensional.

When the correlation between $Y_i(\cdot)$ and $X_i(\cdot)$ is r -dimensional, it follows from (5) that $\text{cov}\{\xi_{ij}, X_i(v)\} = 0$ for all $j > r$ and $v \in \mathcal{I}_2$, from which we can conclude that the curve linear regression model (1) has an equivalent representation by r (scalar) linear regression models, as summarised in the following theorem.

Theorem 1 (Theorem 1 of [6]) *Let the linear correlation between $Y_i(\cdot)$ and $X_i(\cdot)$ be r -dimensional. Assume that*

- *The regression coefficient operator β is in the Hilbert space $\mathcal{L}_2(\mathcal{I}_1 \times \mathcal{I}_2)$, and*
- *$\varepsilon_i(\cdot)$ are i.i.d. with $\mathbb{E}\{\varepsilon_i(u)\} = 0$ and $\mathbb{E}\{X_i(v)\varepsilon_j(u)\} = 0$ for any $u \in \mathcal{I}_1$, $v \in \mathcal{I}_2$ and $i, j \geq 1$.*

Then the curve regression model (1) may be represented equivalently by

$$\begin{aligned} \xi_{ij} &= \sum_{k=1}^{\infty} \beta_{jk} \eta_{ik} + \varepsilon_{ij} & \text{for } j = 1, \dots, r, \\ \xi_{ij} &= \varepsilon_{ij} & \text{for } j = r + 1, r + 2, \dots, \end{aligned} \quad (6)$$

where $\varepsilon_{ij} = \int_{\mathcal{I}_1} \varphi_j(u) \varepsilon_i(u) du$, and $\beta_{jk} = \int_{\mathcal{I}_1 \times \mathcal{I}_2} \varphi_j(u) \psi_k(v) \beta(u, v) dudv$.

The above theorem implies that the SVD-based approach provides a framework to define and exploit the correlation dimension between a pair of curves, and to model the functional linear regression relationship between the pair using a finite number of ordinary (scalar) linear regression models. In this framework, as described in Sect. 3.2 below, the prediction is achieved directly from the estimated ordinary linear regression models.

Taking into account the fact that $\text{var}(\eta_{ik}) \rightarrow 0$ as $k \rightarrow \infty$ (see (2) and (4)), we may include only the first Q terms η_{ik} , $k = 1, \dots, Q$ in the r multiple linear regression models, and obtain the ordinary least squares (OLS) estimator of the finite number of linear coefficients. Note that, while the OLS estimator of β_{jk} is unbiased, its variance tends to increase with Q in finite sample performance. That is, if Q is selected too large, we may end up with a model which fits the data too closely but performs poorly in prediction.

As noted in [6], Theorem 1 holds for any valid expansion $X_i(v) = \sum_k \eta_{ik} \psi_k(v)$, provided $\{\xi_{ij}\}$ are obtained from the SVD. Let $X_i(\cdot)$ be of finite dimension in the sense that its Karhunen-Loève decomposition has q terms only, i.e. $X_i(v) = \sum_{k=1}^q \zeta_{ik} \gamma_k(v)$ where $q(\geq r)$ is a finite integer, $\{\gamma_k(\cdot)\}_{k=1}^q$ are q orthonormal functions in $\mathcal{L}_2(\mathcal{S}_2)$ and $\zeta_{i1}, \dots, \zeta_{iq}$ are uncorrelated random variables with $\text{var}(\zeta_{ik}) > 0$. Then, decomposing $X_i(\cdot)$ with respect to $\{\psi_k(\cdot)\}_{k=1}^q$ from the SVD of Σ , the corresponding $\{\eta_{ik}\}$ satisfy $\text{cov}(\eta_{ik}, \eta_{il}) = 0$ for any $k \neq l$. This, together with (5) and (6), implies that $\beta_{jk} = 0$ for all $j \neq k$ and thus (6) is reduced to r simple linear regression problems

$$\begin{aligned} \xi_{ij} &= \beta_{jj} \eta_{ij} + \varepsilon_{ij} && \text{for } j = 1, \dots, r, \\ \xi_{ij} &= \varepsilon_{ij} && \text{for } j = r+1, r+2, \dots \end{aligned}$$

2.1 Estimation

Given the observed pairs of curves $\{Y_i(\cdot), X_i(\cdot)\}$, $i = 1, \dots, n$, let

$$\hat{\Sigma}(u, v) = \frac{1}{n} \sum_{i=1}^n \{Y_i(u) - \bar{Y}(u)\} \{X_i(v) - \bar{X}(v)\},$$

where $\bar{Y}(u) = n^{-1} \sum_i Y_i(u)$ and $\bar{X}(v) = n^{-1} \sum_i X_i(v)$. Performing the SVD on $\hat{\Sigma}(u, v)$, we obtain the estimators $\{\hat{\lambda}_j, \hat{\phi}_j(\cdot), \hat{\psi}_j(\cdot)\}$ for $\{\lambda_j, \phi_j(\cdot), \psi_j(\cdot)\}$, $j = 1, 2, \dots$ in (3). Note that the SVD can be achieved by performing eigenanalysis on the non-negative operators

$$\hat{M}_1(u, u') = \int_{\mathcal{S}_2} \hat{\Sigma}(u, w) \hat{\Sigma}(u', w) dw \quad \text{and} \quad \hat{M}_2(v, v') = \int_{\mathcal{S}_1} \hat{\Sigma}(w, v) \hat{\Sigma}(w, v') dw,$$

which may be transformed into the eigenanalysis of non-negative definite matrices, see Section 2.2.2 of [2].

Adapting Theorem 1 of [2] to the current setting, we can show the consistency of $\hat{\lambda}_j$. We first assume that

- $\{Y_i(\cdot), X_i(\cdot)\}$ is strictly stationary and ψ -mixing with the mixing coefficients $\psi(k)$ satisfying the condition

$$\sum_{k \geq 1} k\psi(k)^{1/2} < \infty.$$

- $\mathbb{E}\{\int_{\mathcal{G}_1} Y_i(u)^2 du + \int_{\mathcal{G}_2} X_i(v)^2 dv\}^2 < \infty$.
- $\lambda_1 > \dots > \lambda_r > 0 = \lambda_{r+1} = \lambda_{r+2} = \dots$.

Then we have $|\hat{\lambda}_k - \lambda_k| = O_p(n^{-1/2})$ for $1 \leq k \leq r$, and $|\hat{\lambda}_k| = O_p(n^{-1})$ for $k > r$, as $n \rightarrow \infty$.

This result implies that the ratios $\hat{\lambda}_{j+1}/\hat{\lambda}_j$ for $j < r$ are asymptotically bounded away from 0, while $\hat{\lambda}_{r+1}/\hat{\lambda}_r \rightarrow 0$ in probability. Therefore, one way of determining the correlation dimensionality is to employ the following ratio-based estimator

$$\hat{r} = \arg \max_{1 \leq j \leq d} \hat{\lambda}_j / \hat{\lambda}_{j+1},$$

where d is a pre-specified upper bound on r . However, this estimator should be used with caution as different components of the SVD can have different degrees of “strength” in the sense that, there may exist some $k < r$ for which non-zero $\lambda_j \neq 0$, $j > k$ are considerably smaller than $\lambda_{j'}$, $j' \leq k$. Further discussion on this point in the framework of factor analysis can be found in [17]. Heuristically, we may estimate r as

$$\hat{r} = \max\{1 \leq j \leq d : \hat{\lambda}_j / \hat{\lambda}_{j+1} > M\}, \quad (7)$$

for sufficiently chosen M to avoid neglecting such smaller non-zero eigenvalues.

Alternatively, [6] proposed the following information criterion based on the estimated eigenvalues, which extended the information criterion introduced in [14] for high-dimensional time series analysis:

$$IC(q) = \log \left(c_* + \frac{1}{d^2} \sum_{k=q+1}^d \hat{\lambda}_k \right) + \tau q \cdot g(n),$$

where $c_*, \tau > 0$ are constants and $g(\cdot)$ is a function of n satisfying $n \cdot g(n) \rightarrow \infty$ and $g(n) \rightarrow 0$ as $n \rightarrow \infty$. While $IC(\cdot)$ was shown to be consistent in identifying r asymptotically, the choices of τ and $g(\cdot)$ played a significant role in finite sample performance. Therefore, it was proposed to fix $g(n)$ as $g(n) = n^{-1/2}$, obtain $q^* = \arg \min_q IC(q; \tau)$ over a grid of values for τ , and choose r as the most frequently

returned among q^* . For the full description of this majority voting scheme, see Section 3.2 of [6].

3 Application to Electricity Load Modelling

In this section, we discuss applying the proposed curve linear regression method to electricity load modelling and forecasting. The load data example (plotted in Fig. 1) contains electricity loads consumed by the French customers of EDF between 2007 and mid-2012. We first highlight some time-varying features exhibited by the daily electricity load curves, which makes it difficult to assume that the entire data can be modelled as being stationary. Then, we introduce a simple classification rule which divides the pairs of load curves into homogeneous sub-groups, such that the curve linear regression modelling is applicable to each sub-group separately. Finally, the combined procedure of classification and curve linear regression is illustrated using a real electricity load forecasting example.

3.1 Classification of Daily Electricity Load Curves

In electricity load data, there exist systematic discrepancies in the profiles and the variability of daily load curves observed on different days of a week or in different months. Figure 2 shows that, while successive daily loads on Mondays–Tuesdays in June and July behave similarly, they are distinctively different from those observed on Saturdays–Sundays in June and July, and also from those observed on Mondays–Tuesdays in January and December. Those profile discrepancies are reflected predominantly in the locations and magnitudes of daily peaks. Typically in France, daily peaks occur at noon in summer and in the evening in winter, due to economic cycle as well as the usage of electrical heating or cooling and lighting. Hence, the profiles and (presumably) the dynamic structure of successive daily curves vary over different days within a week, and also over different months within a year. It has been noted that these systematic discrepancies persist even after the weekly level de-trending step of the hybrid approach (see Section 4.1 of [6]), which implies that the classification of daily loads is an essential step prior to curve linear regression modelling with or without the weekly level modelling.

According to the experts at EDF, in the case of French electricity consumption data, load curves on the same day of a week tend to have similar profiles. Therefore it is reasonable to assign a day type (DT) to each daily load as summarised in Table 1.

Table 1 Daily classification of daily load curves

Day type	0	1	2	3	4	5	6
Day of a week	Mon	Tue	Wed	Thu	Fri	Sat	Sun

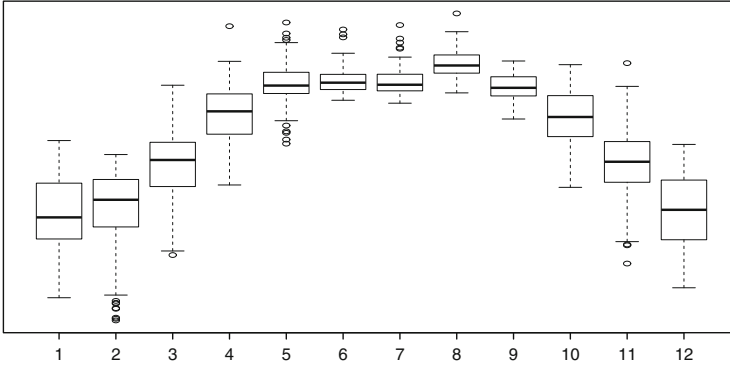


Fig. 3 Boxplots of $\hat{\zeta}_{i1}$ from different months

Table 2 Seasonal classification of daily load curves

Seasonal class	1	2	3	4	5	6	7	8
Month	Jan–Feb, Nov–Dec	Mar	Apr	May	Jun–Jul	Aug	Sep	Oct

To gain an insight into the possible seasonal variation present in the covariance between successive daily loads, as well as in their profiles, we decompose the daily load curves (denoted by $Z_i(\cdot)$ for the loads on the i -th day) as follows. Performing the SVD on the sample covariance function between successive daily curves $Z_{i+1}(\cdot)$ and $Z_i(\cdot)$, we obtain the first left singular function $\hat{\gamma}_1(\cdot)$ and decompose each $Z_{i+1}(\cdot)$ as $\hat{\zeta}_{i1} = \langle Z_{i+1}, \hat{\gamma}_1 \rangle$; see Fig. 3. We note that each $Z_i(\cdot)$ has been de-measured with the mean curve obtained by averaging out all the observations of the same DT. If the dependence structure between the pairs of curves undergoes seasonal changes, we expect such seasonality to be reflected in the behaviour of $\hat{\zeta}_{i1}$ over the span of 1 year. Indeed, this is the case as observable in the boxplots of $\hat{\zeta}_{i1}$ from different months and based on this, we choose to create 8 seasonal classes (SC) as in Table 2.

Combining the two classification rules, we classify each pair of successive daily loads into sub-groups of (approximately) homogeneous dependence structure, according to the corresponding DTs and SCs. While it lacks rigorous statistical ground, the forecasting models estimated within such sub-groups perform well as demonstrated in Sect. 4. Besides, the problem of classifying electricity load curves and functional data in general can stand alone as an independent research problem, and it has attracted considerable attention, see e.g. [5, 21, 22] and [16] for functional data clustering, and [1] for that in the context of electricity loads classification.

3.2 An Illustration

We illustrate the application of curve linear regression with an example where our aim is to predict the electricity load curve for the next 24 h (48 half-hours), denoted by $Y(\cdot)$, at the noon of Tuesday 12 June 2012. Note that in (1), $Y_i(\cdot)$ and $X_i(\cdot)$ are allowed to have different supports as \mathcal{S}_1 and \mathcal{S}_2 , such that we have flexibility in the choice of the curve regressor. Therefore we consider the following three choices:

- $X^{(1)}(\cdot)$: load curve for the 24 h up to the midday of 12 June 2012.
- $X^{(2)}(\cdot)$: $X^{(1)}(\cdot)$ joined with the temperature forecast ($\equiv T^F(\cdot)$) for the next 24 h.
- $X^{(3)}(\cdot)$: $X^{(2)}(\cdot)$ joined with the temperature curve ($\equiv T(\cdot)$) observed over the same 24 h interval as $X^{(1)}(\cdot)$.

We have used the temperature forecasts from meteoFrance in our study. As discussed in Sect. 3.1, $\{Y_i(\cdot), X_i^{(m)}(\cdot)\}$, $m = 1, 2, 3$ are collected as all the observed pairs of curves corresponding to $\{(DT 1, SC 5), (DT 0, SC 5)\}$ between 1 January 2007 and the midday 12 June 2012. In total, there are $n = 38$ observations, which are plotted in Fig. 4 along with their respective mean curves. It may be noted that, due to the classification step, the regressor curves $\{X_i^{(1)}(\cdot), i = 1, \dots, n\}$ and the response curves $\{Y_i(\cdot), i = 1, \dots, n\}$ do not satisfy the relationship

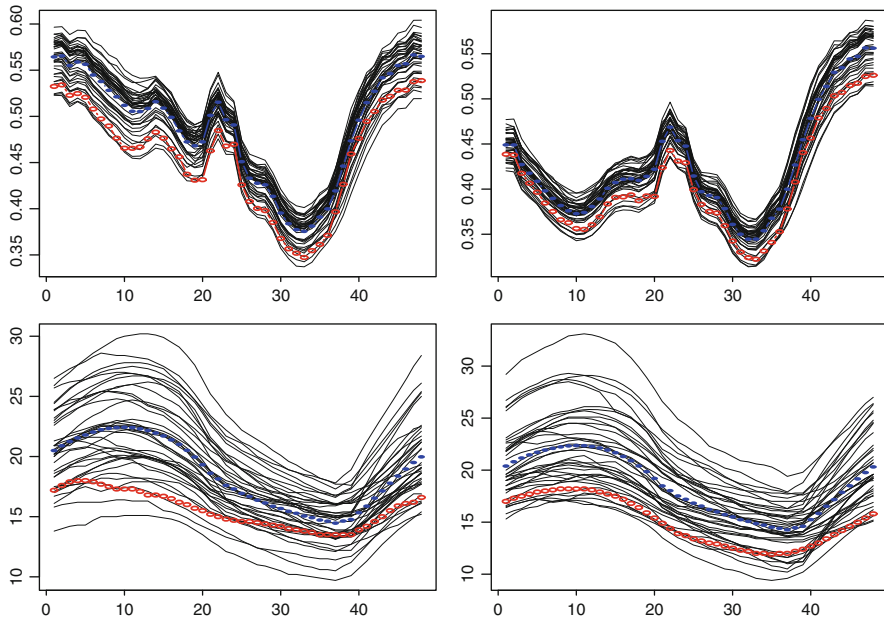


Fig. 4 The n curve observations $Y_i(\cdot)$ (top left), $X_i^{(1)}(\cdot)$ (top right), $T_i(\cdot)$ (bottom left) and $T_i^F(\cdot)$ (bottom right), together with their respective mean curves (filled circle) as well as $Y(\cdot)$, $X^{(1)}(\cdot)$, $T(\cdot)$ and $T^F(\cdot)$ (empty circle)

$X_{i+1}^{(1)}(\cdot) = Y_i(\cdot)$. Hence, even with $X_i(\cdot) \equiv X_i^{(1)}(\cdot)$, the curve linear regression model (1) is distinguished from the autoregressive Hilbertian process of order 1 (ARH(1)) proposed in [3].

Note that for $X_i^{(2)}(\cdot)$ and $X_i^{(3)}(\cdot)$ which join the observed loads with the temperature, different components have different scales since $X_i^{(1)}(\cdot)$ range in tens of thousands (MW), while $T_i(\cdot)$ and $T_i^F(\cdot)$ range in a far smaller scale between 6 and 33 ($^{\circ}\text{C}$). Since the SVD-based method is not scale-invariant, we apply a simple standardisation step to have different components of the regressor curves in a similar scale.

From the observed curves, we estimate the sample covariance function

$$\hat{\Sigma}^{(m)}(u, v) = \frac{1}{n} \sum_{i=1}^n \{Y_i(u) - \bar{Y}(u)\} \{X_i^{(m)}(v) - \bar{X}^{(m)}(v)\}, \quad m = 1, 2, 3,$$

and perform the SVD on $\hat{\Sigma}^{(m)}(u, v)$ to obtain $\{\hat{\lambda}_j^{(m)}, \hat{\phi}_j^{(m)}(\cdot), \hat{\psi}_j^{(m)}(\cdot)\}$, $j = 1, 2, \dots$. Applying (7) to the estimated eigenvalues with $M = 5$, the correlation dimensions are estimated as $\hat{r}^{(m)} = 4$ for all $m = 1, 2, 3$. Defining $\hat{\xi}_{ij}^{(m)} = \langle Y_i - \bar{Y}, \hat{\phi}_j^{(m)} \rangle$ and $\hat{\eta}_{ik}^{(m)} = \langle X_i^{(m)} - \bar{X}^{(m)}, \hat{\psi}_k^{(m)} \rangle$ analogously as $\hat{\xi}_{ij}$ and $\hat{\eta}_{ik}$, the next step is to estimate the linear coefficients $\beta_{jk}^{(m)}$ in the following scalar linear regression models

$$\hat{\xi}_{ij}^{(m)} = \sum_{k=1}^Q \beta_{jk}^{(m)} \hat{\eta}_{ik}^{(m)} + \varepsilon_{ij}^{(m)}$$

for $m = 1, 2, 3$. We set $Q = 15$ to preserve the prediction accuracy by having sufficient number of terms, while attaining the numerical stability of the OLS estimator of $\beta_{jk}^{(m)}$. Then the predictor of $Y(u)$ takes the following form

$$\hat{Y}^{(m)}(u) = \bar{Y}(u) + \sum_{j=1}^{\hat{r}^{(m)}} \hat{\xi}_j^{(m)} \hat{\phi}_j^{(m)}(u),$$

where $\hat{\xi}_j^{(m)}$ are predicted as

$$\hat{\xi}_j^{(m)} = \sum_{k=1}^Q \hat{\beta}_{jk}^{(m)} \hat{\eta}_k^{(m)}, \quad j = 1, \dots, \hat{r}^{(m)},$$

with $\hat{\eta}_k^{(m)} = \langle X^{(m)} - \bar{X}^{(m)}, \hat{\psi}_k^{(m)} \rangle$.

For each m , we obtain two other predictors besides $\hat{Y}^{(m)}(\cdot)$, the oracle and the base predictors. The oracle predictor is of the form

$$\tilde{Y}^{(m)}(u) = \bar{Y}(u) + \sum_{j=1}^{\hat{r}^{(m)}} \tilde{\xi}_j^{(m)} \hat{\phi}_j^{(m)}(u),$$

which is similar to $\hat{Y}^{(m)}(u)$ except that $\hat{\xi}_j^{(m)}$ are replaced by $\tilde{\xi}_j^{(m)} = \langle Y - \bar{Y}, \hat{\phi}_j^{(m)} \rangle$. We use the term ‘‘oracle’’ to emphasise the fact that $\tilde{\xi}_j^{(m)}$ require the prior knowledge of $Y(\cdot)$ and thus are unavailable in practice. The base predictor is set simply as $\bar{Y}^{(m)}(\cdot) = \bar{Y}(\cdot)$, ignoring the dynamic dependence between the response and the regressor curves.

To evaluate the performance of different predictors, we employ the following two error measures

$$\text{RMSE} = \left\{ \frac{1}{N} \sum_{t=1}^N (\hat{f}_t - f_t)^2 \right\}^{1/2} \quad \text{and} \quad \text{MAPE} = \frac{1}{N} \sum_{t=1}^N \left| \frac{\hat{f}_t - f_t}{f_t} \right|, \quad (8)$$

where \hat{f}_t and f_t denote the predicted and the true loads in the t -th half-hour interval and N denotes the forecasting horizon ($N = 48$ in this case). The MAPE and RMSE for the above predictors are reported in Table 3.

As expected, the oracle predictors return smaller prediction errors than the SVD-based predictors, and the base predictor returns the largest error. Based on this, we can conclude that (a) there is much to be accounted for by the dependence between the regressor and the response curves, as observable from the poor performance of $\bar{Y}(\cdot)$, and (b) the reduced dimension captures such dependence structure well, as demonstrated by the superior performance of $\tilde{Y}^{(m)}(\cdot)$.

$\hat{Y}^{(m)}(\cdot)$ perform as competitively as $\tilde{Y}^{(m)}(\cdot)$, attaining RMSE as small as 292 MW without any prior knowledge on the true load $Y(\cdot)$. This fact is also confirmed in Fig. 5, where all $\hat{Y}^{(m)}(\cdot)$ and $\tilde{Y}^{(m)}(\cdot)$ are quite close to $Y(\cdot)$ throughout the forecasting horizon. Among the three $\hat{Y}^{(m)}(\cdot)$, $m = 1, 2, 3$, the choice of $X^{(2)}(\cdot)$ returns the best forecast.

Finally, when the aim is to produce multi-step ahead forecasts, we simply replace the curve regressor $X^{(1)}(\cdot)$ by one of the forecasts $\hat{Y}^{(m)}(\cdot)$ and repeatedly apply the above procedure until the desired multi-step ahead prediction is achieved. Note that the corresponding multi-step ahead temperature forecast may not be available and in such a case, $X^{(1)}(\cdot)$ is the only possible choice as a regressor curve. Thus-produced

Table 3 RMSE and MAPE of the different predictors

Predictor	$\hat{Y}^{(1)}$	$\hat{Y}^{(2)}$	$\hat{Y}^{(3)}$	$\tilde{Y}^{(1)}$	$\tilde{Y}^{(2)}$	$\tilde{Y}^{(3)}$	\bar{Y}
RMSE (MW)	361	292	327	189	218	220	2,440
MAPE (%)	0.77	0.64	0.73	0.41	0.46	0.46	6.65

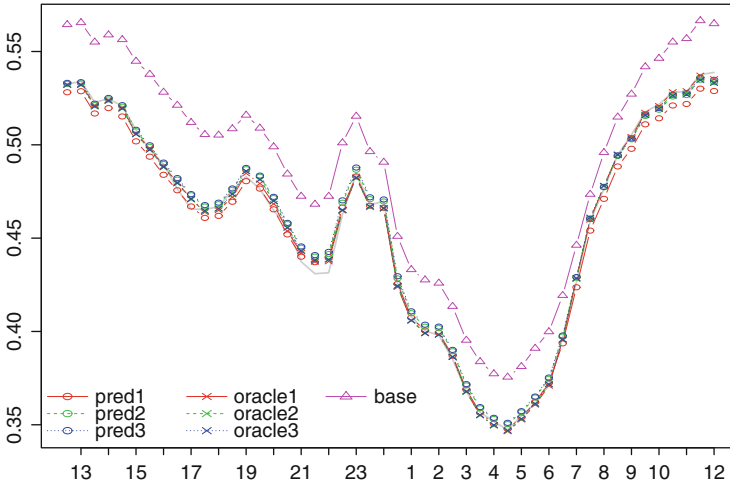


Fig. 5 Different predictors against the true load curve (*grey, solid*) for the next 24 h at noon 12 June 2012

two-day ahead predictor at the noon of 12 June 2012 attains 464 MW RMSE and 1.20 % MAPE, with $X^{(1)}(\cdot)$ replaced by $\hat{Y}^{(2)}(\cdot)$.

4 Forecasting Daily Electricity Consumption of EDF Customers

In this section, we perform one-day ahead forecasting for daily electricity loads consumed by the French customers of EDF from 1 September 2011 to 15 June 2012. As with the example in Sect. 3.2, the forecast is produced every day at noon. Hence, when forecasting the load curve for the next 24 h on day t , we assume the accessibility of the load and the temperature observations from 1 January 2007 up to the noon of day t , as well as the temperature forecast for the next 24 h. During this period, there are certain days (e.g., bank holidays) on which the load observations have not been validated and excluding such days, load forecasts are produced for 234 days in total. Also, when the temperature forecast ($T_i^F(\cdot)$) is not available, we assume that the true one-day ahead temperature ($T_{i+1}(\cdot)$) is known for convenience.

Recalling the notations from Sect. 3.2, we denote the forecasting models with the three regressors $X^{(m)}(\cdot)$, $m = 1, 2, 3$ by P1–P3, respectively, and the corresponding oracle forecasting models by O1–O3. We also consider the predictors from the hybrid approach ([6], H1–H3), where we employ the same regressors at the curve linear regression stage. At the weekly GAM stage, the explanatory variables are lagged weekly average load, weekly average temperature, weekly average cloud cover and two calendar variables representing the yearly trend and the seasonality.

Table 4 RMSE and MAPE of the daily electricity load forecasts between 1 September 2011 and 15 June 2012

	P1	P2	P3	P4	O1	O2	O3	Base	H1	H2	H3	GAM
RMSE (MW)	1,250	853	872	804	336	312	312	6,164	1,917	1,812	1,813	832
MAPE (%)	1.97	1.47	1.50	1.37	0.53	0.50	0.51	10.75	2.91	2.72	2.75	1.40

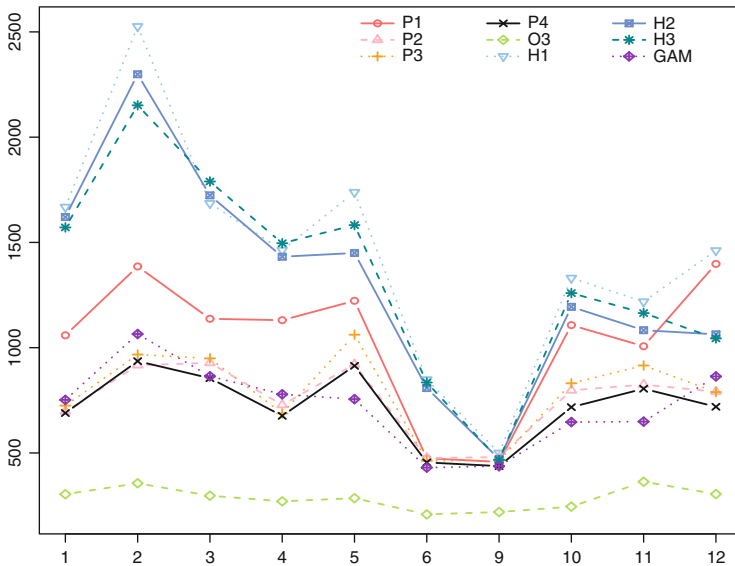


Fig. 6 RMSE from P1–P4, O3, H1–H3 and GAM with respect to different months

Finally, the results from the GAM model provided by the EDF R&D department are presented (“GAM”) for comprehensive comparative study.

Additive models for short-term electricity load forecasting have been studied e.g., in [19] and [11], where the proposed models were shown to be well-adapted to non-linear behaviour of the electricity load. The GAM included in our study models the relationship between each half-hour interval load and several explanatory variables such as the lagged load, calendar events, temperature and cloud cover forecasts. For further information, see [24] and [18]. The EDF operational model is not included in our study. In practice, the true consumption of the EDF customers is not known in real time unlike our assumption above, and therefore the operational model cannot be compared with other models on an equal footing.

The RMSE and the MAPE from different models are reported in Table 4, and Fig. 6 shows the plot of RMSE averaged within each month. For brevity and better representation, only O3 is included among the oracle predictors and the base predictor is omitted.

Overall, the forecasting performance of any model considered, including the oracle predictors, is better in summer than in winter as can be seen in Fig. 6. The

relative difficulty of forecasting French electricity loads in winter has been noted in [9, 10] and [7]. This may be accounted for by higher variability among the daily loads in winter, which is markedly greater than that in summer as demonstrated in Fig. 2.

Also, it is observable from Fig. 6 that among P1–P3, different models outperform the others in different months. For instance, in June and September, P1 performs as well as P2 and P3 or even slightly better, but its performance is considerably worse during colder seasons. In general, the efficacy of having temperature included in the regressor is likely to depend on the homogeneity of the observed temperature curves within each class and the quality of the temperature forecasts. Therefore, we may achieve improved forecasting performance by combining these predictors in an adaptive way, either by selecting one predictor out of the three, or by assigning some data-driven weights to the three predictors on each day. Indeed, by selecting the best forecast out of the three a posteriori (i.e. assuming that the true future load is known), we can reduce the overall RMSE to 660 MW.

Without attempting to be theoretically rigorous, we produced a new predictor (P4) by averaging two out of the three each day, where the two predictors were chosen as those two closest to each other. This additional step can be achieved without any prior knowledge of the future load, yet succeeds in reducing prediction errors by a considerable margin as reported in Table 4. Also, P4 universally outperforms P1–P3 in terms of RMSE in any month of a year. We note that there is a growing interest in the problem of aggregating multiple expert advices in the context of short-term electricity load forecasting. For example, [8] investigate this problem by sequentially updating the convex weights applied to various forecasting models based on the past performance.

The performance of hybrid approaches (H1–H3) is substantially worse than that of their simplified counterparts (P1–P3). It can be explained by the fact that the errors from fitting and predicting the weekly average loads at the weekly level modelling (see Fig. 7), are carried over to the daily level curve linear regression modelling. We note that the electricity load dataset studied here covers the consumption of the customers of EDF only, rather than that of the entire French population as in [6]. Therefore its weekly average loads are more prone to digress from the overall trend or the seasonality estimated from the past observations due to e.g., the departure and the arrival of customers. This leads to greater variance in modelling the linear relationship between $\hat{\xi}_{ij}$ and $\hat{\eta}_{ik}$, $k = 1, \dots, Q$ (see Fig. 8), even when the same classification rule has been applied to the daily loads, and thus to worse prediction models.

The superior performance of P1–P3 to H1–H3 indicates that the classification of successive daily loads effectively handles the dependency of the trend and the seasonality of electricity load data on the calendar variables. While we have used a simple classification rule combining the DT and the SC in this study, existing functional data clustering methods such as [5] may be applied to divide the successive daily loads into sub-groups of homogeneous profiles and covariance

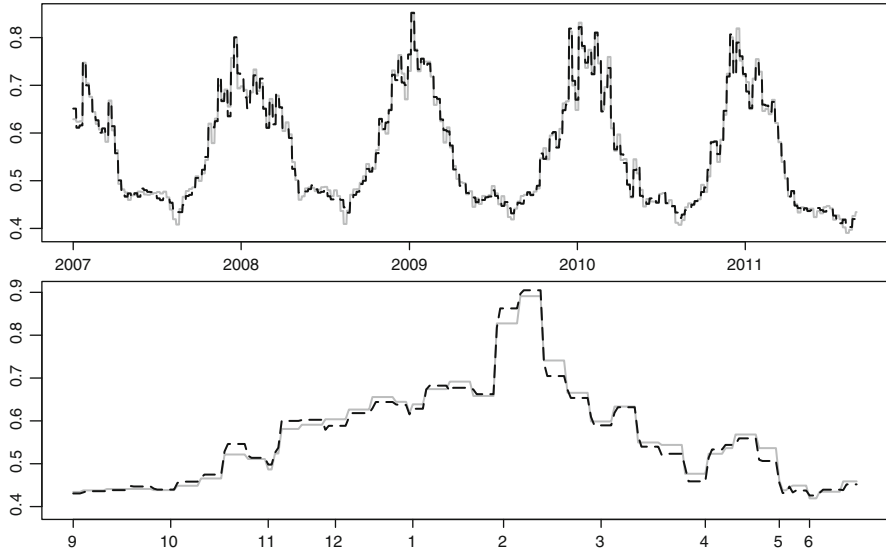


Fig. 7 Weekly average loads (*grey, bold*) between 1 January 2007 and 31 August 2011 and their fitted values (*black, dashed*) (*top*); weekly average loads between 1 September 2011 and 15 June 2012 and their forecasts (*bottom*)

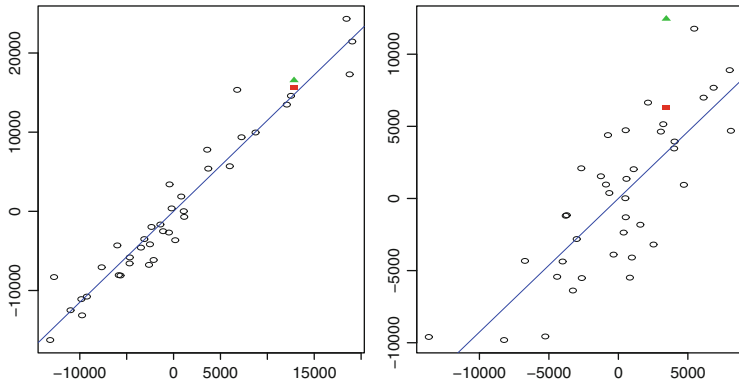


Fig. 8 Relationship between $\hat{\eta}_{i1}$ and $\hat{\xi}_{i1}$ obtained from P2 (*left*) and H2 (*right*) at noon 12 June 2012, along with the respective ($\hat{\eta}_1, \hat{\xi}_1$) (predicted, *filled square*) and ($\hat{\eta}_1, \hat{\xi}_1$) (oracle, *filled triangle*)

structure in a more data-driven way, rather than relying on any prior knowledge on electricity consumption patterns which may change over time.

Also, there are certain factors which are known to have substantial influence on daily electricity loads yet have not been incorporated into our forecasting framework. An example of such factors is the special tariff options offered by EDF to large businesses on certain days in January–March and November–December, with the purpose of reducing heavy electricity consumption during winter. This

scheme is known to affect not only the daily consumption on the special tariff days but also that on the days preceding and following. A data-driven classification tool may be able to identify such influence of the scheme without being furnished with the exact dates or any other information on the load patterns on the relevant days, and thus further improve the quality of forecasts.

According to the overall prediction errors, GAM performs better than P2 and worse than P4 by a small margin, and the breakdown of RMSE with respect to different months in Fig. 6 does not reveal any patterns so as to the relative performance of our method and GAM in different months. The oracle predictors attain the minimum errors throughout the year, which further validates the previous statement that the SVD-based dimension reduction method is successful in capturing the dependence between the regressor and the response curves. It also supports our observations that there is a scope for improvement, e.g. via adaptive aggregation of different forecasting models and data-dependent classification of successive daily loads.

5 Conclusions

In this article, we addressed the problem of daily electricity load forecasting via curve linear regression, with emphasis on the adaptivity of the proposed method to ever-changing electricity consumption environment. The curve linear regression technique was introduced in a generic setting, where the singular value decomposition in a Hilbert space reduced the curve linear regression model to a finite number of scalar linear regression models.

Although it had previously been proposed by [6] as the second stage of the hybrid method for daily load forecasting, we showed that the curve linear regression technique could be applied directly to the data without any preliminary trend and seasonality modelling, based on the following rationale.

- The trend and the seasonality depend on the calendar variables which can be used as classification criteria, and when equipped with such a classification step, the weekly level modelling is redundant.
- In the hybrid approach, the prediction error from the first stage is carried over to the second stage, which leads to the increased variance in curve linear regression modelling and thus to significantly deteriorated prediction performance.

Also, the reduced approach requires less human intervention and is more adaptive to the time-varying nature of the data, and its superior prediction performance has been demonstrated with a real data example. Besides, within the reduced framework, it is more straightforward to carry out further statistical analysis such as obtaining a prediction interval around the forecast. By focusing exclusively on curve linear regression, some interesting topics for further improving the methodology have been made clearer throughout the real data analysis.

Firstly, as seen in Sect. 3.1, clustering the daily loads into homogeneous sub-groups, in terms of both their profiles and dependence structure, plays a key role in

electricity load data analysis. Data-driven classification of the successive daily loads can greatly improve the forecasting results, as well as providing interesting insights on the data itself. There is an active interest on developing functional data clustering techniques, and adapting these methods to electricity load data is a problem which requires our immediate attention.

Further, since the curve linear regression framework allows flexible choice of regressor, we can have a number of forecasting models with different regressors. Therefore, it is of interest to see whether we can achieve improved forecasting performance by adaptively aggregating multiple forecasts. As briefly explored in Sect. 4, a simple adjustment in this direction can enhance the prediction performance substantially. Also on a more general note, an automatic selection of the regressor in curve linear regression may be widely adopted as a functional data analysis tool beyond the context of electricity load forecasting.

References

1. Antoniadis, A., Brossat, X., Cugliari, J., & Poggi, J. M. (2013). Functional clustering using wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, 11, 1350003(30 pages).
2. Bathia, N., Yao, Q., & Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *Annals of Statistics*, 38, 3352–3386.
3. Bosq, D. (2000). *Linear processes in function spaces: Theory and applications* (Lecture Notes in Statistics, Vol 149). New York: Springer.
4. Bruhns, A., Deurveilher, G., & Roy, J. S. (2005). A non linear regression model for mid-term load forecasting and improvements in seasonality. In *Proceedings of the 15th power systems computation conference*, Liège (pp. 22–26).
5. Chiou, J. M., & Li, P. L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B*, 69, 679–699.
6. Cho, H., Goude, Y., Brossat, X., & Yao, Q. (2013). Modelling and forecasting daily electricity load curves: A hybrid approach. *Journal of the American Statistical Association*, 108, 7–21.
7. Cugliari, J. (2011). Prévission non paramétrique de processus à valeurs fonctionnelles: Application à la consommation d'électricité. PhD thesis. University Paris XI, Paris, France.
8. Devaine, M., Gaillard, P., Goude, Y., & Stoltz, G. (2013). Forecasting electricity consumption by aggregating specialized experts. *Machine Learning*, 90, 231–260.
9. Dordonnat, V., Koopman, S. J., Ooms, M., Dessertaine, A., Collet, J. (2008). An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, 24, 566–587.
10. Dordonnat, V., Koopman, S. J., Ooms, M., Dessertaine, A., & Collet, J. (2012). Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling. *Computational Statistics and Data Analysis*, 56, 3134–3152.
11. Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27, 134–141.
12. Hall, P., & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35, 70–91.
13. Hall, P., & Vial, C. (2006). Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society: Series B*, 68, 689–705.
14. Hallin, M., & Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102, 603–617.

15. He, G., Müller, H. G., Wang, J. L., & Yang, W. (2010). Functional linear regression via canonical analysis. *Bernoulli*, *16*, 705–729.
16. James, G. M., & Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, *98*, 397–408.
17. Lam, C., & Yao, Q. (2012). Factor modelling for high-dimensional time series: inference for the number of factors. *Annals of Statistics*, *40*, 694–726.
18. Nedellec, R., Cugliari, J., & Goude, Y. (2014). Gefcom2012: Electricity load forecasting and backcasting with semi-parametric models. *International Journal of Forecasting*, *30*(2), 375–381.
19. Pierrot, A., & Goude, Y. (2011). Short-term electricity load forecasting with generalized additive models. In *Proceedings of the 16th international conference on intelligent system application to power systems*, Hersonissos (pp. 22–26)
20. Ramsay, J. O., & Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B*, *53*, 539–572.
21. Ray, S., & Mallick, B. (2006). Functional clustering by bayesian wavelet methods. *Journal of the Royal Statistical Society: Series B*, *68*, 305–332.
22. Serban, N., & Wasserman, L. (2005). CATS: Clustering after transformation and smoothing. *Journal of the American Statistical Association*, *100*, 990–999.
23. Smithies, F. (1937). The eigenvalues and singular values of integral equations. *Proceedings of the London mathematical society*, *43*(2), 255–279.
24. Wood, S., Goude, Y., & Shaw, S. (2015). Generalized additive models for large datasets. *Journal of Royal Statistical Society: Series C*, *64*(1), 139–155.
25. Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, *33*, 2873–2903.