Modelling and forecasting daily electricity load curves: a hybrid approach *

Haeran Cho[†] Yannig Goude[‡] Xavier Brossat[‡] Qiwei Yao[†,§]

Abstract

We propose a hybrid approach for the modelling and the short-term forecasting of electricity loads. Two building blocks of our approach are (i) modelling the overall trend, seasonality and the effect of meteorological factors by fitting a generalised additive model to the *weekly* averages of the loads, and (ii) modelling the dependence structure across consecutive *daily* loads via curve linear regression. For the latter, a new methodology is proposed for linear regression with both curve response and curve regressors. The key idea behind the proposed methodology is the dimension reduction based on a singular value decomposition in a Hilbert space, which reduces the curve regression problem to several ordinary (i.e. scalar) linear regression problems. We illustrate the hybrid method using the French electricity loads between 1996 and 2009, on which we also compare our method with other available models including the EDF operational model.

KEY WORDS: Curve regression; Correlation dimension; Dimension reduction; Forecasting; Electricity loads; Generalised additive models; Singular-value decomposition.

*Partially supported by the EPSRC research grant EP/G026874/1.

[†]Department of Statistics, London School of Economics, UK.
[‡]Électricité de France, France.
[§]Guanghua School of Management, Peking University, China

1 Introduction

As electricity can be stored or discharged only at extra costs, it is an important task for electricity providers to model and forecast electricity loads accurately over short-term (from one day to one month ahead) or middle-term (from one month to five years ahead) horizons. The electricity load is an essential entry of the optimisation tools adopted by energy companies for power system scheduling. A small improvement in the load forecasting can bring in substantial benefits in reducing the production costs as well as increasing the trading advantages, especially during the peak periods.

The French energy company Électricité de France (EDF) manages a large panel of production units in France and in Europe, which include water dams, nuclear plants, wind turbines, coal and gas plants. Over the years, EDF has developed a very accurate load forecasting model which consists of complex regression methods coupled with classical time series techniques such as the seasonal ARIMA (SARIMA) model. The model integrates a great deal of physical knowledge on the French electricity consumption patterns that has been accumulated over 20 years, such as the fact that the temperature felt indoors is more relevant than the real temperature in modelling the electricity load. Furthermore it includes exogenous information ranging from economic growth forecasts to different tariff options provided by the company. The forecasting model in operation performs very well at present, attaining about 1% mean absolute percentage error in forecasting over one day horizon. However, it has a drawback in terms of its poor capacity in adapting to the changes in electricity consumption habits which may occur due to the opening of new electricity markets, technological innovations, social and economic changes, to name a few. Hence it is strategically important to develop some new forecasting models which are more adaptive to ever-changing electricity consumption environment and the hybrid method proposed in this paper, designed for short-term forecasting for daily loads, represents a determined effort in this direction.

Electricity load exhibits interesting features at different levels. Figure 1 displays the

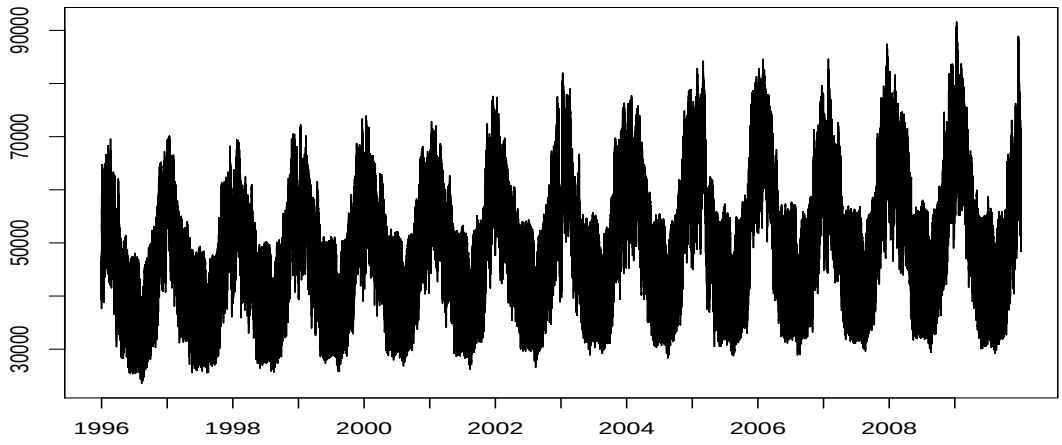

Figure 1: Electricity load from 1996 to 2009 in France.

electricity load in France measured every half an hour from 1996 to 2009. First of all, there is an overall increasing trend due to meteorological and economic factors. In addition, an annual seasonal pattern repeats itself every year, which can be explained by seasonal changes in temperature, day light duration and cloud cover. Engle et al. (1986) and Taylor and Buizza (2002) discussed the impact of meteorological factors on the electricity load, and singled out the temperature as the most important one, due to the large demand of electrical heating in cold weather. Further studies on the meteorological effect include Taylor and McSharry (2008), where seasonal patterns of electricity loads over 10 European countries were reported. Also, there exist daily patterns which, unfortunately, do not show off due to the large scale of Figure 1, attributed to varying demands for electricity in the different periods within a day. Figure 6 below provides an example of such daily patterns.

Based on the above observations, we propose to model the electricity loads at two different levels using different methods, hence the name *hybrid* approach. First assuming that the long-term trends do not vary greatly within a week, we extract those trends from weekly average loads using a generalised additive model, where temperature and

other meteorological factors are included as additional explanatory variables. After removing the long term trend component from the data, we view the daily loads as curves and model the dynamic dependence among the electricity loads of successive days via curve linear regression. For this, a new dimension-reduction technique based on a singular value decomposition in Hilbert space is proposed, which reduces the regression with a curve response and a curve regressor to several ordinary (i.e. scalar) linear regression models. Regarding the daily loads as curves, our approach takes advantage of the continuity of the consumption curves in statistical modelling, as well as embedding some nonstationary features (such as daily patterns) into a stationary framework in a functional space. Therefore, as we expected, the proposed method provides more accurate predictions than more conventional methods such as those based on seasonal ARIMA models or exponential smoothing. Although the EDF operation model provides more accurate predictions than our method, our model is considerably simpler and does not make use of the full subject knowledge which is accumulated over more than 20 years of the development at the EDF and which is not available in the public domain. Hence our approach is more adaptive to the changing electricity consumption environment while retaining a competitive prediction capacity, and can be adopted as a generic tool applicable to a wide range of problems including the electricity load forecasting in countries other than France. Furthermore it has the potential to serve as a building block for constructing a more effective operation forecasting model when incorporating the full EDF subject knowledge.

There is a growing body of literature devoted to electricity load forecasting models. Focusing on the main interest of this paper, we list below the recent papers in short-term load forecasting; see Bunn and Farmer (1985) for a more comprehensive overview. In the category of parametric approaches, Ramanathan et al. (1997) proposed linear regression models with autoregressive errors for each hour of a day. Univariate methods such as SARIMA models or exponential smoothing can be found in Hyndman et al.

(2002), Taylor et al. (2006) and Taylor (2010), and state-space models in Dordonnat et al. (2008) and Dordonnat et al. (2011). Among the nonparametric and semiparametric methods, Engle et al. (1986) proposed to include the temperature effect in the load modelling, and Harvey and Koopman (1993) proposed a time-varying spline model that captured both the temperature effect and the seasonal patterns in a semi-parametric way. Generalised additive models for electricity loads were studied in Fan and Hyndman (2011) and Pierrot and Goude (2011), where the semi-parametric approaches were shown to be well-adapted to non-linear behaviours of the electricity load signal. In Antoniadis et al. (2006), a forecasting model based on functional data analysis was proposed which treated the daily electricity loads as curves. This approach has been further developed in Cugliari (2011). On the other hand, Cottet and Smith (2003) proposed a Bayesian autoregressive model for short-term forecasts, where the meteorological effects were non-linear and estimated using semi-parametric regression methods. They obtained good forecasting results with New South Wales dataset.

The rest of the paper is organised as follows. In Section 2, we present the modelling of weekly average loads using a generalised additive model. Then Section 3 discusses the modelling of the dependence structure between daily loads in a curve linear regression framework. We conduct a comparison study in Section 4, where our new method as well as other competitors are applied to predict the French daily loads in 2009. Section 5 contains some conclusive remarks. All the proofs are relegated to a supplementary document.

2 Modelling weekly averages

Assuming that the overall trend and seasonality do not vary greatly within a week, we propose to model the long-term trends with the *weekly* averages, i.e. we treat the trend and seasonal component as being constant within each week. In this manner, we lose little from the gradual changes of the trends within each week, while preserving the

dependence structure across the electricity loads of different days. The weekly averages of the EDF loads from 1996 to 2008 are plotted in Figure 2. In the literature, it has

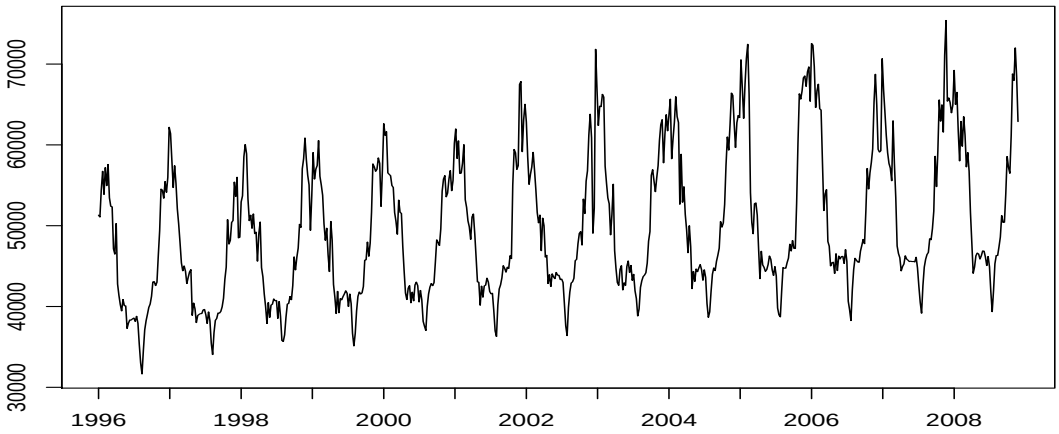

Figure 2: Weekly average electricity load in France from 1996 to 2008.

been noted that some meteorological factors, such as temperature and cloud cover, have a significant impact on the electricity consumption patterns. While there are other detrending techniques that have been proposed for removing long-term trends and seasonal cycles, we fit the weekly averages with a generalised additive model (GAM) for its ability to model implicit nonlinear relationships between response and explanatory variables without suffering from the so-called "curse of dimensionality"; see Hastie and Tibshirani (1990) and Wood (2006) for further details on the GAM and Pierrot et al. (2009) and Pierrot and Goude (2011) for its application in electricity load modelling. Denoting the time index representing each week by t, the explanatory variables considered in fitting the weekly average load process L_t are: O_t is the weekly median of the *offset* (a temporal variable determined by the experts at EDF to represent the seasonal trend in the data, taking values -3, -2, -1 and 0 to denote different winter holidays, 1 to denote spring, 2–6 to denote summer and summer holidays, and 7 to denote autumn), T_t is the weekly average of the temperature, C_t is the weekly average of the cloud cover, and I_t

is the weekly index ranging from 1 to 53.

Our first attempt at taking into account the meteorological effects as well as the temporal trend is summarised in the following GAM with the Gaussian link function

$$L_t = f_1(t) + f_2(O_t) + f_3(L_{t-1}) + f_4(T_t) + f_5(T_{t-1}) + f_6(C_t), \tag{1}$$

where each f_j is a smooth function of the corresponding covariate with thin plate regression splines as a smoothing basis. We use the R package *mgcv* introduced in Wood (2006), where each smooth function f_j is estimated by penalised regression splines. In this implementation, the amount of penalisation is calibrated according to the generalised cross-validation (GCV) score, see Wood (2004) and Wood (2011) for details.

We note that the basis used to estimate f_1 has knots at each first week of September, which are imposed to model the time-varying trend in the electricity load at the yearly level. The boxplot of the residuals from fitting the above GAM to the weekly average load between 1996 and 2008 is provided in Figure 3, and the estimated curves for $f_1, \ldots, f_6$ in (1) are plotted in Figure 4, with shaded area representing the twice standard error bands below and above the estimate. The fitted curve explains 98.7% of the

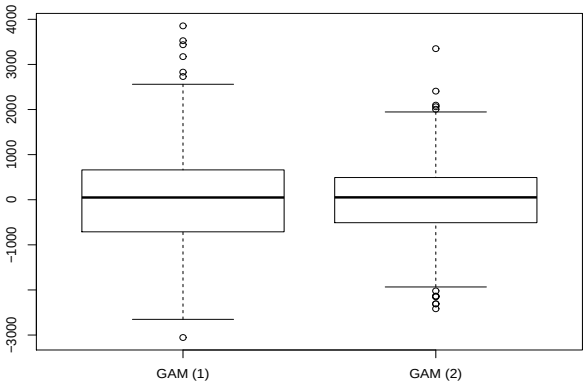

Figure 3: Boxplots of the residuals from fitting the weekly average load between 1996 and 2008 using the model (1) (left) and the model (3) (right).

data, and the mean absolute percentage error (MAPE) and the root mean square error (RMSE) from the estimated curve are 1.63% and 1014MW, respectively. The two error measures, MAPE and RMSE, are defined as

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\widehat{L}_t - L_t}{L_t}\right| \quad \text{and} \quad \text{RMSE} = \left\{\frac{1}{T}\sum_{t=1}^{T}(\widehat{L}_t - L_t)^2\right\}^{1/2}, \tag{2}$$

where $\widehat{L}_t$ denotes the estimated (or predicted) load in the week t.

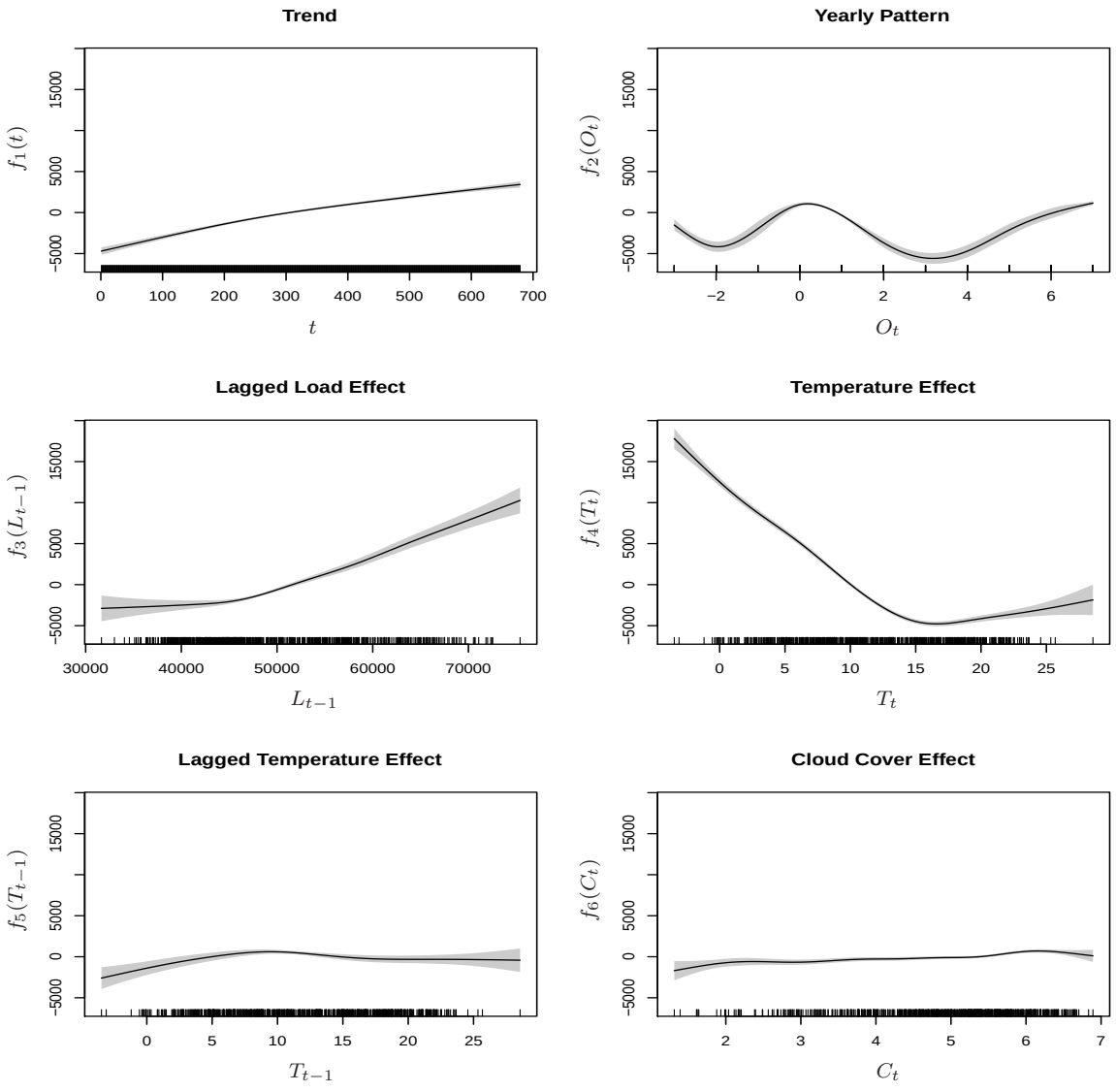

Figure 4: Estimated $f_1, \ldots, f_6$ from model (1); shaded regions represent the confidence bands.

We state below some observations based on the estimated functions in Figure 4. The top left panel shows that the electricity load increases over time t, and the trend is almost linear. The top right panel shows clearly the presence of the seasonality as the load is lower during holidays and in summer than that in winter. As for the lagged load effect, L_t increases with respect to its lagged value L_{t-1} (the second left panel) and the rate of increase is greater when $L_{t-1} > 5 \times 10^4$ approximately, which implies that the value 5×10^4 may be regarded as a "threshold" acting on the impact of L_{t-1} on L_t. Since the increase in the usage of electricity is closely related to the climate, which in turn is linked to the time of the year, we may include the joint effect of L_{t-1} and I_t in the model to accommodate the dependence between those two variables. Also, the impact of temperature is significant (the second right panel). The low temperatures lead to high electricity consumptions due to electrical heating, resulting the initial sharp decrease in $\widehat{f}_4$. Then as the temperature increases from about 17°C upwards, $\widehat{f}_4$ also increases slowly, which can be accounted by the use of cooling system in hot weather. As the meteorological changes within a year is closely related to the time index, we may include the joint effect of the variables T_t and I_t in the model. The bottom panels show that, although not as prominent as other terms, the lagged temperature and the cloud cover do have an impact on the weekly average load at large values of T_{t-1} and C_t. The effect of cloud cover is significantly different from 0 for large values of C_t, as heavy cloud cover induces the increasing use of lighting (the bottom right panel). We note that the estimated effect of the low cloud cover may be an artifact: there are only few observations available for low cloud cover and thus the variance of the fitted curve at such small values of C_t is large.

Based on the above observations, we propose another model

$$L_t = f_1(t) + f_2(O_t) + f_3(L_{t-1}, I_t) + f_4(T_t, I_t) + f_5(T_{t-1}, I_t) + f_6(C_t, I_t), \tag{3}$$

where $f_3, \ldots, f_6$ include the weekly index I_t as a covariate. To study the bivariate

effects, we plot the estimated f_3 and f_4 in Figure 5. The impact of the lagged load L_{t-1} on the load L_t is similar as previously described with the model (1) in the sense that, the rate of increase of L_t changes when L_{t-1} is greater than a threshold value. However, we also note that the relationship between L_{t-1} and L_t varies throughout a year with the weekly index I_t, and that the impact of L_{t-1} is far stronger in winter than in summer. As for the effect of temperature, there is a smooth transition observable throughout a year from the winter heating effect to the summer cooling effect.

With the new model, there is an increase in the percentage of the data explained (99.2%), and both the MAPE (1.28%) and the RMSE (801MW) of the fitted trend have decreased. Further, the GCV score indicates that the new model is favourable (8.4×10^5) to the previous one (1.2×10^6). Also, when comparing the forecasts from the two models for the weekly average loads of 2009, (3) performed considerably better (MAPE 1.72%, RMSE 1250MW) than the model (1) (MAPE 2.15%, RMSE 1532MW). We note that the superior performance of the model (3) at the weekly level carries over to that at the daily electricity load forecasting; when applied to forecast the daily loads in 2009, the MAPE and RMSE from the model (3) were 1.35% and 869MW respectively, whereas the model (1) led to 1.41% and 901MW, see Section 4 for more details. From these observations and also from the residual boxplots in Figure 3, we choose model (3) over model (1).

3 Regression of daily load curves

Once the long-term trend is fitted as in Section 2 and removed, we regard the residuals on the i-th day as a curve $Y_i(\cdot)$ defined on the index set $\mathcal{I}_1$, and model the dependency among the daily loads via curve linear regression as

$$Y_i(u) = \int_{\mathcal{I}_2} X_i(v)\beta(u,v)dv + \varepsilon_i(u) \quad \text{for} \quad u \in \mathcal{I}_1, \tag{4}$$

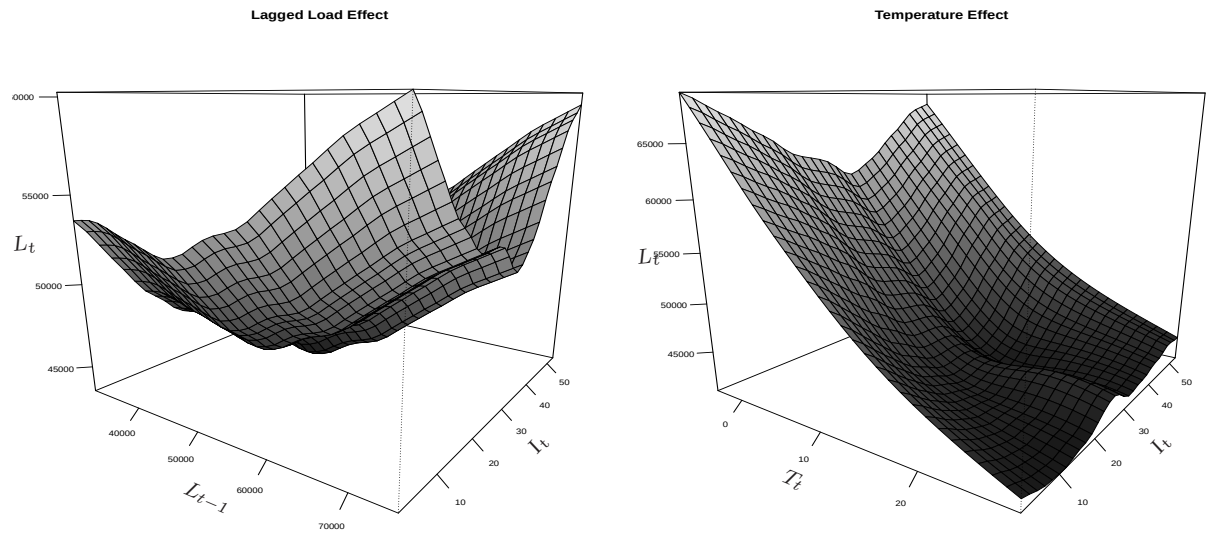

Figure 5: Estimated f_3 (left) and f_4 (right) from model (3).

where $X_i(\cdot)$ can be, for example, the residual curve on the $(i-1)$-th day (i.e. $Y_{i-1}(\cdot)$), or the curve joining $Y_{i-1}(\cdot)$ and the temperature curve on the i-th day. Therefore the index set of $X_i(\cdot)$, say $\mathcal{I}_2$, may be different from $\mathcal{I}_1$. In (4), β is a regression coefficient function defined on $\mathcal{I}_1 \times \mathcal{I}_2$, and $\varepsilon_i(\cdot)$ is noise with mean 0.

Linear regression with curves as both response and regressor, has been studied by, among others, Ramsay and Dalzell (1991), He et al. (2000), and Chiou et al. (2004) and Yao et al. (2005). The conventional approach is to apply the Karhunen-Loève decomposition to both $Y_i(\cdot)$ and $X_i(\cdot)$, and then to fit a regression model using the finite number of terms obtained from such decompositions. The Karhunen-Loève decomposition has featured predominantly in functional data analysis; see also Fan and Zhang (1998) and Hall and Horowitz (2007). This approach is identical to the dimension reduction based on principal component analysis in multivariate analysis. Since the principal components do not necessarily represent the directions in which $X_i(\cdot)$ and $Y_i(\cdot)$ are most correlated, we present below a novel approach where the singular value decomposition (SVD) is applied to single out the directions upon which the projections of $Y_i(\cdot)$ are most correlated with $X_i(\cdot)$. Our method is closely related to the canonical

correlation analysis yet we focus on regressing $Y_i(\cdot)$ on $X_i(\cdot)$, and thus $Y_i(\cdot)$ and $X_i(\cdot)$ are not treated on an equal footing, which is different from, and much simpler than, the canonical correlation analysis. The literature on functional canonical correlation analysis includes Hannan (1961), Silverman (1996), He et al. (2003), Cupidon et al. (2008), Eubank and Hsing (2008) and Yang et al. (2011).

3.1 Curve linear regression via dimension reduction

Let $\{Y_i(\cdot), X_i(\cdot)\}$, $i = 1, \ldots, n$, 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 $f, g \in \mathcal{L}_2(\mathcal{I})$. We assume that $\mathbb{E}\{Y_i(u)\} = 0$ for all $u \in \mathcal{I}_1$ and $\mathbb{E}\{X_i(v)\} = 0$ for all $v \in \mathcal{I}_2$, and denote the covariance function between $Y_i(\cdot)$ and $X_i(\cdot)$ by $\Sigma(u, v) = \mathrm{cov}\{Y_i(u),\, X_i(v)\}$. Under the assumption

$$\int_{\mathcal{I}_1} \mathbb{E}\{Y_i(u)^2\}du + \int_{\mathcal{I}_2} \mathbb{E}\{X_i(v)^2\}dv < \infty, \tag{5}$$

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

$$f_1(u) \to \int_{\mathcal{I}_1} \Sigma(u, v)f_1(u)du \in \mathcal{L}_2(\mathcal{I}_2), \qquad f_2(v) \to \int_{\mathcal{I}_2} \Sigma(u, v)f_2(v)dv \in \mathcal{L}_2(\mathcal{I}_1)$$

for any $f_i \in \mathcal{L}_2(\mathcal{I}_i)$. Based on the SVD, there exists a triple sequence $\{(\varphi_j, \psi_j, \lambda_j),\ j = 1, 2, \ldots\}$ for which

$$\Sigma(u, v) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, \varphi_j(u)\, \psi_j(v), \tag{6}$$

where $\{\varphi_j\}$ is an orthonormal basis of $\mathcal{L}_2(\mathcal{I}_1)$, $\{\psi_j\}$ is an orthonormal basis of $\mathcal{L}_2(\mathcal{I}_2)$, and $\{\lambda_j\}$ are ordered such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq 0. \tag{7}$$

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

$$\int_{\mathcal{I}_1} M_1(u, z)\, \varphi_j(z)\, dz = \lambda_j\, \varphi_j(u), \qquad \int_{\mathcal{I}_2} M_2(v, z)\, \psi_j(z)\, dz = \lambda_j\, \psi_j(v), \tag{8}$$

where M_i is a non-negative operator defined on $\mathcal{L}_2(\mathcal{I}_i)$ as

$$M_1(u, u') = \int_{\mathcal{I}_2} \Sigma(u, z)\, \Sigma(u', z)\, dz, \qquad M_2(v, v') = \int_{\mathcal{I}_1} \Sigma(z, v)\, \Sigma(z, v')\, dz.$$

It is clear from (8) that λ_j is the j-th largest eigenvalue of M_1 and M_2 with φ_j and ψ_j as the corresponding eigenfunctions, respectively. Since $\{\varphi_j\}$ and $\{\psi_j\}$ are the orthonormal basis of $\mathcal{L}_2(\mathcal{I}_1)$ and $\mathcal{L}_2(\mathcal{I}_2)$, we may write

$$Y_i(u) = \sum_{j=1}^{\infty} \xi_{ij} \varphi_j(u), \qquad X_i(v) = \sum_{j=1}^{\infty} \eta_{ij} \psi_j(v), \tag{9}$$

where ξ_{ij} and η_{ij} are random variables defined as

$$\xi_{ij} = \int_{\mathcal{I}_1} Y_i(u) \varphi_j(u) du, \qquad \eta_{ij} = \int_{\mathcal{I}_2} X_i(v) \psi_j(v) dv. \tag{10}$$

It follows from (6) that

$$\mathrm{cov}(\xi_{ij},\, \eta_{ik}) = \mathbb{E}(\xi_{ij} \eta_{ik}) = \begin{cases} \sqrt{\lambda_j} & \text{for } \ j = k, \\ 0 & \text{for } \ j \neq k. \end{cases} \tag{11}$$

We refer to Smithies (1937) for further details on the SVD in a Hilbert space.

Now, we are ready to introduce the notion of the correlation dimension between the two curves. See Hall and Vial (2006) and Bathia et al. (2010) for the definitions of curve dimensionality in different contexts.

Definition 1. The correlation between curves $Y_i(\cdot)$ and $X_i(\cdot)$ is r-dimensional if $\lambda_r > 0$ and $\lambda_{r+1} = 0$ in (7).

When the correlation between $Y_i(\cdot)$ and $X_i(\cdot)$ is r-dimensional, it follows from (11) that $\text{cov}\{\xi_{ij},\, X_i(v)\} = 0$ for all $j > r$ and $v \in \mathcal{I}_2$. Moreover, the linear curve regression (4) admits an equivalent representation with r (scalar) linear regression models; see Theorem 1 below. Before presenting the theorem, we further assume that the regression coefficient $\beta(u, v)$ is in the Hilbert space $\mathcal{L}_2(\mathcal{I}_1 \times \mathcal{I}_2)$, and that $\varepsilon_i(\cdot)$ 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 $i, j \geq 1$.

Theorem 1. Let the linear correlation between $Y_i(\cdot)$ and $X_i(\cdot)$ be r-dimensional. Then the curve regression (4) may be represented equivalently by

$$\begin{aligned} &\xi_{ij} = \textstyle\sum_{k=1}^{\infty} \beta_{jk}\eta_{ik} + \varepsilon_{ij} && \text{for } \ j = 1, \ldots, r, \\ &\xi_{ij} = \varepsilon_{ij} && \text{for } \ j = r+1, r+2, \ldots, \end{aligned} \tag{12}$$

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 proof of Theorem 1 is provided in the supplementary document. Some remarks are listed in order.

(a) For each $j = 1, \ldots, r$, we may apply model selection criteria such as the AIC, to select among $\{\eta_{ik},\, k \geq 1\}$ the variables to be included in the first linear regression model of (12), noting $\text{var}(\eta_{ik}) \to 0$ as $k \to \infty$; see (5) and (9). We also note that $\{\varphi_j(u)\psi_k(v)\}_{j,k}$ form an orthonormal basis of $\mathcal{L}_2(\mathcal{I}_1 \times \mathcal{I}_2)$. Since $\beta(u, v) \in \mathcal{L}_2(\mathcal{I}_1 \times \mathcal{I}_2)$, it holds that $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \beta_{jk}^2 = \int_{\mathcal{I}_1 \times \mathcal{I}_2} \beta(u, v)^2 dudv < \infty$.

(b) In fact, Theorem 1 holds for any valid expansion of $X_i(v)$ as $X_i(v) = \sum_k \eta_{ik}\psi_k(v)$, as long as $\{\xi_{ij}\}$ are obtained from the SVD. For example, we may use the Karhunen-Loève decomposition of $X_i(\cdot)$. Then resulting η_{ik} is the projection of $X_i(\cdot)$ on the k-th principal direction, and those $\{\eta_{ik}\}$ are uncorrelated with each other.

(c) Let $X_i(\cdot)$ be of finite dimension in the sense that its Karhunen-Loève decomposition has q terms only as $X_i(v) = \sum_{k=1}^{q} \zeta_{ik}\gamma_k(v)$, where $q(\geq r)$ is a finite integer, $\{\gamma_k(\cdot)\}_{k=1}^{q}$ are q orthonormal functions in $\mathcal{L}_2(\mathcal{I}_2)$, and $\zeta_{i1}, \ldots, \zeta_{iq}$ are uncorrelated

with $\mathrm{var}(\zeta_{ik}) > 0$ for all $k = 1, \ldots, q$. Without loss of generality, we may assume that $\mathrm{var}(\zeta_{ik}) = 1$, which can be achieved by replacing $X_i(v)$ by its linear transformation $\int_{\mathcal{I}_2} \Gamma(v, w) X_i(w) dw$, where $\Gamma(v, w) = \sum_{k=1}^{q} \gamma_k(v)\gamma_k(w) \big/ \sqrt{\mathrm{var}(\zeta_{ik})}$. Then for such $X_i(\cdot)$, the second equation in (9) is reduced to $X_i(v) = \sum_{k=1}^{q} \eta_{ik}\psi_k(v)$ with $\{\eta_{ik}\}$ satisfying $\mathrm{var}(\eta_{ik}) = 1$ and $\mathrm{cov}(\eta_{ik}, \eta_{ij}) = 0$ for any $k \neq j$. This, together with (11) and (12), implies that $\beta_{jk} = 0$ in (12) for all $j \neq k$. Hence (12) is reduced to

$$\begin{aligned} \xi_{ij} &= \beta_{jj}\eta_{ij} + \varepsilon_{ij} \qquad && \text{for} \quad j = 1, \ldots, r, \\ \xi_{ij} &= \varepsilon_{ij} && \text{for} \quad j = r+1, r+2, \ldots, \end{aligned} \tag{13}$$

i.e. under the additional condition on the dimensionality of $X_i(\cdot)$, the curve regression (4) is reduced to r simple linear regression problems.

(d) We provide a recap of the above results in the context of vector regression. Let $\mathbf{y}_i$ and $\mathbf{x}_i$ be, respectively, $p \times 1$ and $q \times 1$ vectors. Suppose that $\mathrm{rk}(\mathbf{\Sigma}_{yx}) = r$, where $\mathbf{\Sigma}_{yx} = \mathrm{cov}(\mathbf{y}_i, \mathbf{x}_i)$. Then the multiple linear regression problem $\mathbf{y}_i = \mathbf{B}\mathbf{x}_i + \boldsymbol{\varepsilon}_i$ may be reduced to the r scalar linear regression problems:

$$u_{ij} = \mathbf{v}_i'\boldsymbol{\beta}_j + \epsilon_{ij}, \qquad j = 1, \cdots, r. \tag{14}$$

Here, $(u_{i1}, \cdots, u_{ip})' = \mathbf{U}'\mathbf{y}_i$ and $\mathbf{v}_i = (v_{i1}, \cdots, v_{iq})' = \mathbf{V}'\mathbf{x}_i$. Also, $\mathbf{\Sigma}_{yx} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}'$ is the SVD of $\mathbf{\Sigma}_{yx}$ with $\mathbf{U}\mathbf{U}' = \mathbf{I}_p$, $\mathbf{V}\mathbf{V}' = \mathbf{I}_q$ and $\mathbf{\Lambda}$ is a $p \times q$ diagonal matrix with only the first $r(\leq \min(p, q))$ main diagonal elements being nonzero. If $\mathrm{var}(\mathbf{x}_i) = \sigma^2\mathbf{I}_q$ is satisfied in addition, (14) reduces to r simple regression models $u_{ij} = v_{ij}\beta_j + \varepsilon_{ij}$ for $j = 1, \cdots, r$.

3.2 Estimation

We assume the availability of observed curves $\{Y_i(\cdot),\, X_i(\cdot)\}$ for $i = 1, \cdots, n$. Recalling $\Sigma(u, v) = \mathrm{cov}\{Y_i(u), X_i(v)\}$, let

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

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 $\widehat{\Sigma}(u, v)$, we obtain the estimators $(\widehat{\lambda}_j,\, \widehat{\varphi}_j,\, \widehat{\psi}_j)$ for $(\lambda_j,\, \varphi_j,\, \psi_j)$ as defined in (6). Note that this SVD is effectively an eigenanalysis of the non-negative operator

$$\widehat{M}_1(u, u') = \int_{\mathcal{I}_2} \widehat{\Sigma}(u, v)\widehat{\Sigma}(u', v)dv, \tag{15}$$

which may be transformed into an eigenanalysis of a non-negative definite matrix. Furthermore $\widehat{\varphi}_j(\cdot)$ and $\widehat{\psi}_j(\cdot)$ may be taken as linear combinations of, respectively, the observed curves $Y_i(\cdot)$ and $X_i(\cdot)$. See, for example, Section 2.2.2 of Bathia et al. (2010). Proposition 1 below presents the asymptotic properties for the estimators $\widehat{\lambda}_j$. Its proof is similar to that of Theorem 1 of Bathia et al. (2010) and is thus omitted.

Proposition 1. Suppose that $\{Y_i(\cdot), X_i(\cdot)\}$ is strictly stationary and ψ-mixing with the mixing coefficients $\psi(k)$ satisfying the condition $\sum_{k\geq 1} k\psi(k)^{1/2} < \infty$. Further, assume $\mathbb{E}\{\int_{\mathcal{I}_1} Y_i(u)^2 du + \int_{\mathcal{I}_2} X_i(v)^2 dv\}^2 < \infty$ and let $\lambda_1 > \cdots > \lambda_r > 0 = \lambda_{r+1} = \lambda_{r+2} = \cdots$. Then as $n \to \infty$,

(i) $|\widehat{\lambda}_k - \lambda_k| = O_p(n^{-1/2})$ for $1 \leq k \leq r$, and

(ii) $|\widehat{\lambda}_k| = O_p(n^{-1})$ for $k > r$.

We refer to Section 2.6 of Fan and Yao (2003) for the further details on mixing conditions. The fast convergence for the zero-eigenvalues λ_j with $j > r$ is due to the quadratic form in (15), see the relevant discussion in Bathia et al. (2010) and Lam and Yao (2011). It follows from Proposition 1 that the ratios $\widehat{\lambda}_{j+1}/\widehat{\lambda}_j$ for $j < r$ are asymp-

totically bounded away from 0, and $\widehat{\lambda}_{r+1}/\widehat{\lambda}_r \to 0$ in probability. This motivates the following ratio-based estimator. In Lam and Yao (2011), a more elaborate investigation of this estimator can be found in a different context.

The ratio-based estimator for the correlation dimension r:

$\widehat{r} = \arg\min_{1\le j\le d} \widehat{\lambda}_{j+1}/\widehat{\lambda}_j$, where $d > r$ is a fixed and pre-specified integer.

One alternative is to use properly defined information criteria for estimating r as in, e.g. Hallin and Liška (2007), where a similar idea was adopted for high-dimensional time series analysis. To this end, we define

$$IC_1(q) = \frac{1}{d^2}\sum_{k=q+1}^{d} \widehat{\lambda}_k + \tau_1 q \cdot g(n), \quad \text{and} \quad IC_2(q) = \log\left(c_* + \frac{1}{d^2}\sum_{k=q+1}^{d} \widehat{\lambda}_k\right) + \tau_2 q \cdot g(n),$$

where $c_*, \tau_1, \tau_2 > 0$ are constants, $d > r$ is a pre-specified integer, and $g(n) > 0$ satisfies

$$n \cdot g(n) \to \infty \quad \text{and} \quad g(n) \to 0, \quad \text{as } n \to \infty. \tag{16}$$

Theorem 2 below shows that $\widehat{r} \equiv \arg\min_{0\le q<d} IC_i(q)$ is a consistent estimator of r for both $i = 1, 2$. The proof is given in the supplementary document.

Theorem 2. Let the conditions of Proposition 1 hold and both r and d be fixed as $n \to \infty$. Then, for both $i = 1, 2$, we have $P\{IC_i(r) < IC_i(q)\} \to 1$ for any $0 \le q < d$ and $q \ne r$.

The choice of c_* is not critical as it is introduced to ensure that the term inside the logarithm is positive. The proof of Theorem 2 indicates that the consistency holds for any constants τ_1 and τ_2. However, they affect the finite sample performance of the method and therefore in practice, the choice of the tuning parameters and τ_1 and τ_2 and the penalty function $g(n)$ requires more care. In our data analysis reported below, we set $g(n) = n^{-1/2}$ and elaborate the choice of τ_i using the following majority voting scheme.

We start with two values τ_* and τ^* such that $IC_i(q)$ is minimised at $q = d$ for any $\tau_i \leq \tau_*$, and at $q = 0$ for any $\tau_i \geq \tau^*$. Over the interval $[\tau_*, \tau^*]$, the function $h(\tau) \equiv \arg\min_q IC_i(q)$ is non-increasing in τ. Then, assigning a grid of values from $[\tau_*, \tau^*]$ as τ_i, we look for the q that is returned over the longest interval of τ_i within $[\tau_*, \tau^*]$, and set such q as the estimate of r. Figure 8 below shows an example of applying $IC_2(q)$ for the selection of r, where $IC_2(q)$ is computed over $q = 1, \ldots, 20$ for 100 different values of τ_2. In this example, $q = 4$ was returned most frequently as the minimiser of $IC_2(q)$. We have further conducted a simulation study to check whether the proposed scheme worked well on simulated datasets of varying dimensionalities, and the results have confirmed its good performance over a range of r.

3.3 An illustration

We illustrate the hybrid approach by predicting the load curve on 2 April 2009, which is denoted by $Z(\cdot)$. Unfortunately, even after removing the long-term trend estimated in Section 2, there exist some systematic discrepancies among the profiles of daily load curves over different days in a week and different months in a year. Figure 6 shows that, while the daily loads on Tuesdays in July are similar to each other, they are distinctively different from those on Saturdays in July, and also from those on Tuesdays in December. Those profile differences 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 the economic cycle as well as the usage of electrical heating and lighting. Hence, the daily curves and presumably their dynamic structure vary over different days within a week, and also over different months in a year; further elaboration on those features is provided in Section 4 below.

To forecast the load curve on Wednesday, 2 April 2009, we take the joined curve of the de-trended curve on Tuesday, 1 April 2009 ($= X^{\mathrm{L}}(\cdot)$) and the temperature curve on 2 April 2009 ($= X^{\mathrm{T}}(\cdot)$) as the regressor, i.e. $X(\cdot) = (X^{\mathrm{L}}(\cdot), X^{\mathrm{T}}(\cdot))$. We use all the pairs

of curves on Tuesday and Wednesday in April from 1996 to 2008 as our observations to fit a curve regression model, and the total number of observations is $n = 53$. Those 53 pair curves $\{X_i(\cdot), Y_i(\cdot)\}$ are plotted in Figure 7 together with their de-meaned and standardised counterparts. From those de-meaned curves, we form a sample covariance matrix

$$\widehat{\Sigma}(u, v) = \frac{1}{53} \sum_{i=1}^{53} \{Y_i(u) - \bar{Y}(u)\}\{X_i(v) - \bar{X}(v)\}, \tag{17}$$

where $\bar{Y}(u) = \frac{1}{53} \sum_{1 \le i \le 53} Y_i(u)$ is the average of all the de-trended daily curves on Wednesdays in April between 1996 and 2008, and $\bar{X}(v)$ is obtained analogously. Applying the SVD to $\widehat{\Sigma}(u, v)$, we obtain the estimators $(\widehat{\lambda}_k, \widehat{\varphi}_k, \widehat{\psi}_k)$. To determine the correlation dimension, we apply the information criterion $IC_2(q)$ with 100 different values of τ_2, as discussed towards the end of Section 3.2. Figure 8 shows $IC_2(q)$ against q for each of the 100 τ_2-values. With this set of data, $q = 4$ minimises $IC_2(q)$ over the longest interval of τ_2, which leads to the estimator $\widehat{r} = 4$.

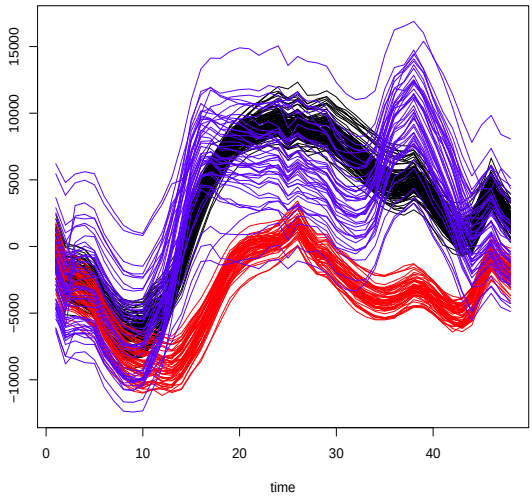

Figure 6: De-trended daily curves for Tuesdays in July (black), Saturdays in July (red) and Tuesdays in December (blue) between 1996 and 2008.

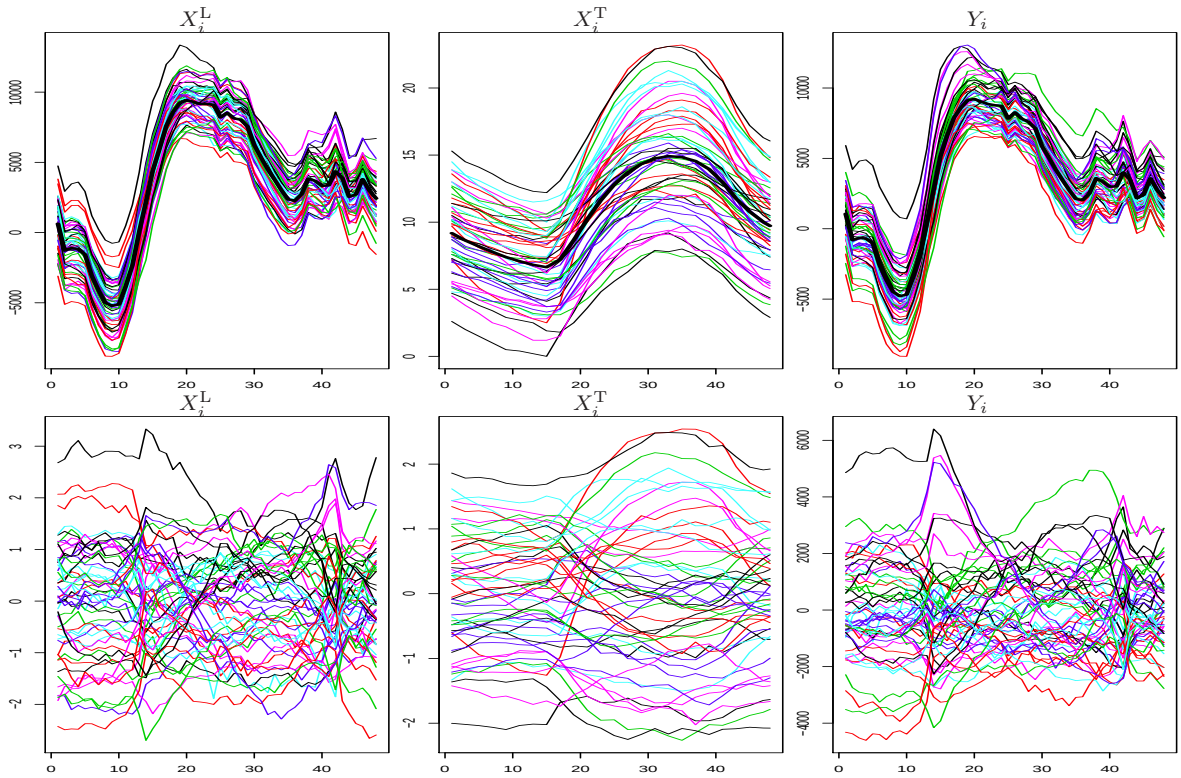

Figure 7: The 53 curves $X_i^{\mathrm{L}}(\cdot)$ (top-left), $X_i^{\mathrm{T}}(\cdot)$ (top-middle) and $Y_i(\cdot)$ (top-bottom), together with their respective mean curves plotted together in black. The de-meaned and standardised curves are plotted in the bottom panels.

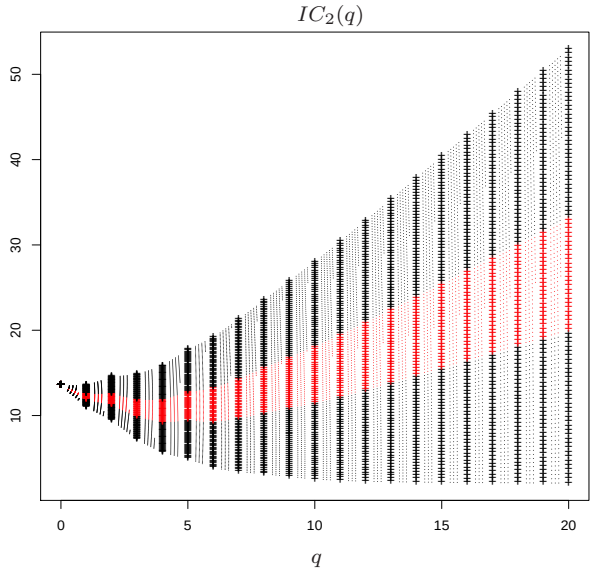

Figure 8: Plots of $IC_2(q)$ against q for 100 different values of τ_2. The curves with the minimum attained at $q = 4$ are highlighted in red.

Our predicted load curve is of the form

$$\widehat{Z}(u) = \widehat{L}_w + \bar{Y}(u) + \sum_{j=1}^{4} \widehat{\xi}_j \widehat{\varphi}_j(u), \tag{18}$$

where $\widehat{L}_w$ is the predicted weekly trend for the week containing 2 April 2009 from the GAM method in Section 2, $\bar{Y}(u)$ is the mean curve as in (17), and $\widehat{\xi}_j$, $j = 1, \ldots, 4$ are the predictors based on linear regression models defined as follows. Based on Theorem 1, the curve linear regression $Y_i(\cdot)$ on $X_i(\cdot)$ may be recast into $\widehat{r} = 4$ ordinary regression models

$$\widehat{\xi}_{ij} = \sum_{k=1}^{10} \beta_{jk} \widehat{\eta}_{ik} + \varepsilon_{ij}, \quad i = 1, \cdots, 53, \quad j = 1, \cdots, 4, \tag{19}$$

where

$$\widehat{\xi}_{ij} = \int_{\mathcal{I}_1} \{Y_i(u) - \bar{Y}(u)\} \widehat{\varphi}_j(u) du, \qquad \widehat{\eta}_{ik} = \int_{\mathcal{I}_2} \{X_i(v) - \bar{X}(v)\} \widehat{\psi}_k(v) dv,$$

see (10). In (19) we choose to use the first 10 singular value components of the regressor only, see Section 4.2 below. Based on the least squares estimators $\widehat{\beta}_{jk}$ from the regression models (19), we obtain the predictors $\widehat{\xi}_j$ as $\widehat{\xi}_j = \sum_{k=1}^{10} \widehat{\beta}_{jk} \widehat{\eta}_k$, where $\widehat{\eta}_k = \int_{\mathcal{I}_2} \{X(v) - \bar{X}(v)\} \widehat{\psi}_k(v) dv$.

We compare our method with the two alternative predictors: the oracle and the baseline predictors. The oracle predictor is of the form

$$\widetilde{Z}(u) = \widehat{L}_w + \bar{Y}(u) + \sum_{j=1}^{4} \widetilde{\xi}_j \widehat{\varphi}_j(u), \tag{20}$$

which is defined similarly as our predictor (18) except with $\widehat{\xi}_j$ being replaced by $\widetilde{\xi}_j \equiv \langle Y(\cdot) - \bar{Y}(\cdot), \widehat{\varphi}_j \rangle$, where $Y(\cdot)$ denotes the de-trended load curve on 2 April 2009. Note that $Y(\cdot)$ is unavailable in practice, and hence the name "oracle". The baseline predictor

is defined as

$$\bar{Z}(u) = \widehat{L}_w + \bar{Y}(u), \tag{21}$$

which is the sum of the first two terms in our predictor, ignoring the dynamic dependence between days. We compare the performance of the three predictors in terms of the following two error measures

$$\text{MAPE} = \frac{1}{48}\sum_{j=1}^{48} |(\widehat{f}_j - f_j)/f_j| \quad \text{and} \quad \text{RMSE} = \{\frac{1}{48}\sum_{j=1}^{48}(\widehat{f}_j - f_j)^2\}^{1/2},$$

where $\widehat{f}_j$ and f_j denote the predicted and the true loads in the j-th half-hour interval. The MAPE and RMSE for our predictor $\widehat{Z}(\cdot)$, the oracle predictor $\widetilde{Z}(\cdot)$ and the baseline predictor $\bar{Z}(\cdot)$ are (0.91%, 634MW), (6.00%, 420MW) and (3.14%, 1911MW), respectively. The three predicted curves are plotted in Figure 9 together with the true curve. Our predictor $\widehat{Z}(\cdot)$, making good use of the dynamic dependence across different days, is a significant improvement from the baseline predictor $\bar{Z}(\cdot)$. While the oracle predictor $\widetilde{Z}(\cdot)$ is impractical as $\widetilde{\xi}_j$ is unavailable in practice, its superior performance in terms of both MAPE and RMSE indicates that the dimension reduction achieved via SVD retains the relevant dynamic information in the system.

Finally, we discuss the extension to multi-step ahead predictions using the hybrid approach, which straightforwardly translates to producing multi-step ahead predictions from the GAM at the weekly level, and from the ordinary (scalar) linear regression steps at the daily level. Specifically, if the corresponding week of the multi-step ahead forecast is different from that of the one-step ahead forecast, the forecast is obtained by plugging the average temperature and cloud cover of the week into the fitted GAM. At the daily level modelling, the forecast of the next day's load replaces (part of) the regressor curve to produce that of the following day, and this is repeated until the desired multi-step ahead prediction is achieved. In the above example, when making a two-day ahead prediction for Thursday, 3 April 2009 on 1 April 2009, the first part of the regres-

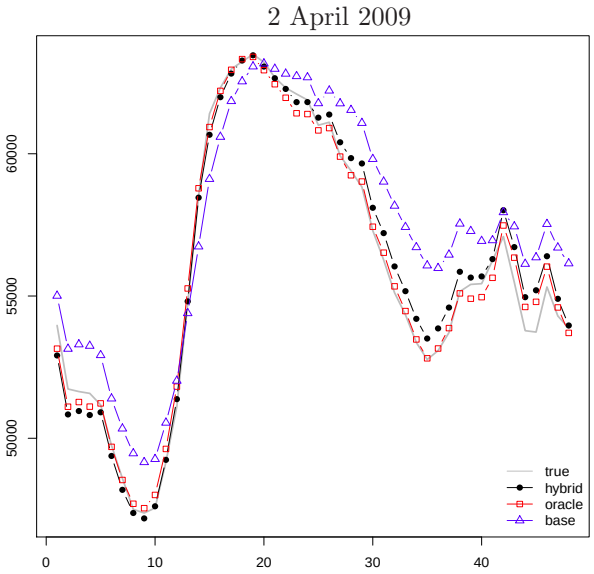

Figure 9: The true daily load curve (grey, solid) of 2 April 2009, together with its predicted curves by our method (black, filled circle), the oracle method (red, empty square) and the base-line (blue, empty triangle).

sor curve becomes $X^{\mathrm{L}}(\cdot) = \widetilde{Z}(u)$, while the second part is the daily temperature curve on 3 April 2009. Following the identical steps described in this section, the two-step ahead forecast achieves MAPE 1.06% and RMSE 657MW. In general, the performance of multi-step ahead forecasts is worse than that of one-day ahead forecasts as the errors in the latter carry over to the errors in the former.

4 Predicting daily loads in 2009

To compare different predictive models more systematically, and to gain further appreciation of the performance of our method over different periods of a year, we predict the daily load curves for all days in 2009. For each day in 2009, we use the data from 1 January 1996 to its previous day to build the prediction models in the same manner as described in Section 3.3, i.e. first the trend component (i.e. as $\widehat{L}_w$ in (18)) is predicted by the GAM model in (3), and then the residual process is divided into daily curves.

Table 1: Day types furnished by the EDF experts.

index	0	1	2	3	4	5	6	7
day type	Mon	Tue–Thu	Fri	Sat	Sun (rest)	Sun (Jun-Jul)	Sun (Aug)	Sun (Dec)

4.1 Classification of daily curves

Discussions in Section 3.3 indicate that we need to treat the daily residual curves on each day of a week differently. For the French electricity load dataset, we are furnished with the *day type* of each day, i.e. a classification of the daily curves determined by the experts at EDF. The day type is defined with respect to different days of a week, and bank holidays are assigned to separate day types according to their profiles. See Table 1 for the summary of day types. Furthermore, to take into account the seasonal changes which may be present in the shapes ($\mathbb{E}\{Y_i(\cdot)\}$ and $\mathbb{E}\{X_i(\cdot)\}$) as well as the dependence structure ($\Sigma(u,v) = \mathrm{cov}(Y_i(u), X_i(v))$) of daily curves, we divide one year into 9 seasonal segments: January to February, March, April, May, June to July, August to September, October, November, and December. This segmentation was determined by inspecting the decomposition of electricity loads with respect to adaptively chosen orthonormal functions. More precisely, we performed principal component analysis on the pool of de-meaned daily curves (according to the day type) and decomposed them with respect to the first principal direction. By examining the changes in the decomposition over a year (see Figure 10) we obtained the segmentation of a year as provided above.

While the above classification lacks a rigorous statistical ground, the prediction model based on this classification performs well in practice. Besides, classification of electricity load curves can stand alone as an independent research problem which has attracted considerable attention, see e.g. Chiou and Li (2007), Ray and Mallick (2006), Serban and Wasserman (2005) and James and Sugar (2003) for functional clustering and Antoniadis et al. (2010) in the context of electricity loads classification. In summary, each daily curve is classified according to the day of a week and the season of a year, and there are about 67 pairs of classes for any two consecutive days. For each pair of classes,

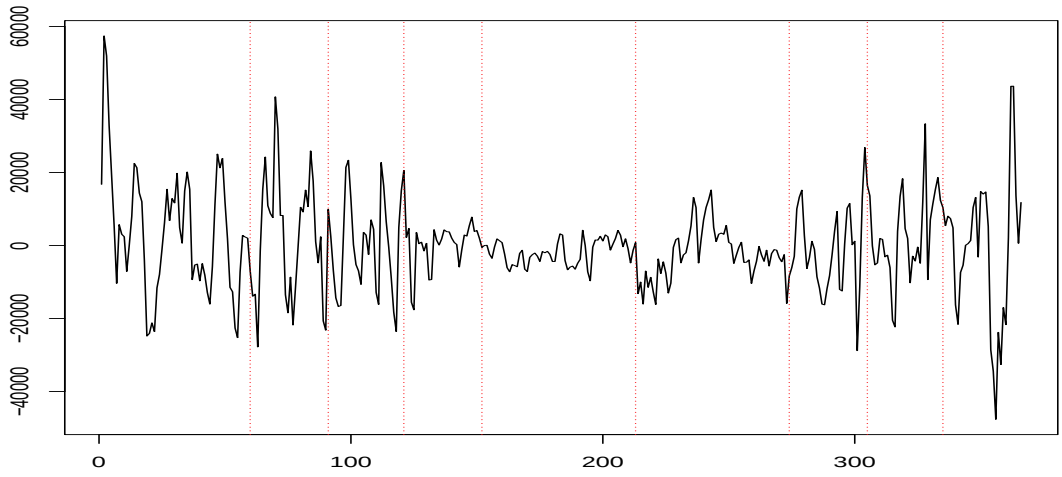

Figure 10: Decomposition of the daily curves from 2008 with respect to the first principal component estimated from the pooled daily curves between 1996 and 2008: seasonal segments are denoted by dotted, red lines.

we fit a prediction model separately in the same manner as described in Section 3.3.

4.2 Prediction comparisons

In applying the proposed hybrid method, we consider four different versions H1–H4 depending on the choice of regressor. H1 uses the load curve on the current day as the regressor (i.e. $X(\cdot) = X^{\mathrm{L}}(\cdot)$). H2 uses the joined curve of the load curve on the current day and the temperature curve on the next day (i.e. $X(\cdot) = (X^{\mathrm{L}}(\cdot), X^{\mathrm{T}}(\cdot))$), as it has been practiced in Section 3.3. H3 adopts the same regressor as H1 but with a half-day curve such that, if we are forecasting the electricity load from 00:30 to 12:00 on the next day, the load curve on the current day from 12:30 to 24:00 is used as the regressor curve; if we are forecasting the curve from 12:30 to 24:00, the regressor curve is the load curve from 00:30 to 12:00 on the same day. Similarly, H4 employs the same regressor as H2 but also with half-day curves. To facilitate a more comprehensive comparison, we predict the daily load curves by our proposed hybrid method (18), the oracle method (with the same regressor as in H2, (20)) and the baseline method (21). We also include in the

comparison study, the prediction results from the EDF operational model, the seasonal ARIMA model (denoted as SARIMA) as in Taylor and McSharry (2008), a combination of GAM and SARIMA (GSARIMA) method, and the exponential smoothing technique (EST) discussed in Taylor (2010). In total, there are 10 different models used in our comparison study.

Denote the number of observations for each class by n. Since we impose an upper bound of 10 on the correlation dimension r, we include those classes with n greater than 15 in our comparison study. Also, only the first 10 η_{ik}s are used in the scalar linear regression models (12), as having more than 10 terms does not improve the results dramatically while n is allowed to be as small as 15. We further note that it is considered a more challenging task to forecast electricity loads for holidays than those for working days, and often additional prior information is used for holidays in practice. Instead of making the whole exposition over complicated, we focus on the forecasting for the working days only. There are 315 days in total where all the conditions stated above are satisfied.

Note that in the hybrid approach, we require the forecasts of the average temperature of the following week, as well as the temperature curve of the next day. As such information can easily be furnished by Météo-France for this particular dataset, we may assume that the forecast of the next day's temperature has been provided in the form of a curve, and the weekly average temperature of the following week can be replaced by the mean of such a forecast (in accordance with the assumption that the long-term trend to vary little within each week). Since the resulting MAPE (1.38%) and RMSE (891MW) from H2 are only slightly worse than those obtained with the true temperature values (MAPE 1.35%, RMSE 869MW), we report in what follows the results obtained assuming that all the necessary information is available. Forecasting errors are measured by the MAPE and RMSE, and summarised in Table 2. We also present the errors with respect to different seasons and day types in Figures 11–12.

The prediction based on any model considered is more accurate in summer than in

winter, see Figure 11. The relative difficulty of load forecasting in winter has been noted for the French dataset in Dordonnat et al. (2008), Dordonnat et al. (2011) and Cugliari (2011). SARIMA and GSARIMA are consistently outperformed by other methods by a large margin, and between the two, GSARIMA achieves smaller forecasting errors. Between H1 (H3) and H2 (H4), the latter attains considerably smaller forecasting errors as it makes use of more information on the temperature, although Figure 11 shows that this observation is not held consistently throughout the year. We note that the performance of our approach may further be improved by making an adaptive choice of regressor curve dependent on the level of temperature.

From Figure 11, it is interesting to observe that the half-day based approaches, H3 or H4, achieve better forecasting performance than H1 or H2 in some colder months (February–April, October–November), while the opposite is true in warmer months. This may be understood in relation with the variability among the curves, which is considerably greater in winter than in summer (see e.g. Figure 6). On a similar note, while forecasting errors from the EDF operational model are smaller than those from hybrid approaches on average, the difference is noticeably reduced from May to September. Indeed, H1 and H2 return errors which are comparable to or even smaller than those from the operational model in June, July and September. In terms of day type, the forecasting errors from the hybrid methods are larger on Mondays than for the rest of a week on average (see Figure 12), which may also be due to the greater variability in the relationship between the curves from Sundays and Mondays. The oracle predictor attains the minimum errors throughout the year except for in December, which suggests that there is the scope for improvement in the hybrid approach by improving the linear regression fit at the daily level.

There are certain factors which are known to have substantial influence on daily electricity loads yet have not been incorporated into our hybrid modelling. For example, from November to March, EDF offers special tariff days to large businesses as financial

Table 2: Summary of MAPE and RMSE of the electricity load forecasts for 01/01/2009–31/12/2009 from our hybrid modelling (H1, H2, H3, H4), oracle, base, SARIMA, GSARIMA, EST and operational model.

	H1	H2	H3	H4	oracle	base	SARIMA	GSARIMA	EST	operation
MAPE (%)	1.54	1.35	1.37	1.20	0.46	3.05	2.55	2.49	1.97	0.93
RMSE (MW)	1018	869	918	787	317	1882	1607	1586	1330	625

incentives, which are activated to cut heavy electricity consumption in winter. Since the scheme is known to affect not only the daily loads on the special tariff days but also on the days before and after those days, we expect that including prior information on such days, e.g. by creating new classes, can further improve the quality of the forecasts especially in winter.

5 Conclusions

In this paper, we proposed a hybrid approach to electricity load modelling with the aim of forecasting daily electricity loads. In the hybrid procedure, we model the overall and seasonal trends of the electricity load data at the weekly level, by fitting a GAM with temporal and meteorological factors as explanatory variables. At the daily level, the serial dependence among the daily load curves is modelled under the assumption that the curves from two successive days have a linear relationship, and we propose a framework which effectively reduces the curve linear regression to a finite number of scalar linear regression problems. To the best of our knowledge, it has not been explored elsewhere to model the multi-layered features of electricity load dataset at multiple levels separately. Compared to the current operational model at EDF, our proposed method is more model-centred and developed without much of the specific knowledge that have been included in the former, while it still retains a competitive prediction capacity. We also note that our approach has the potential to be more adaptive to changing electricity consumption environment, as well as being applicable to a wider range of problems without much human intervention.

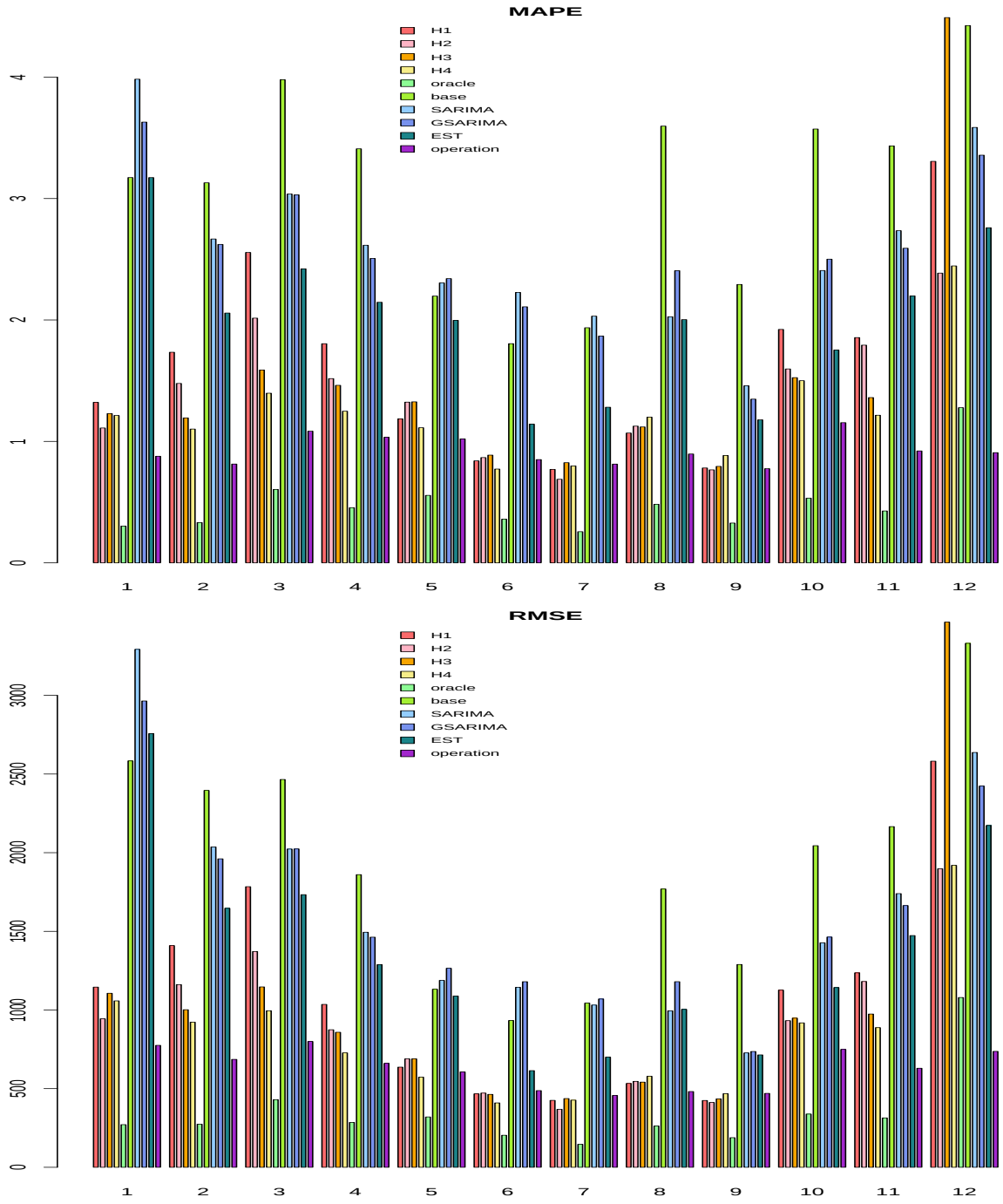

Figure 11: Bar plots of MAPE (top) and RMSE (bottom) with respect to months.

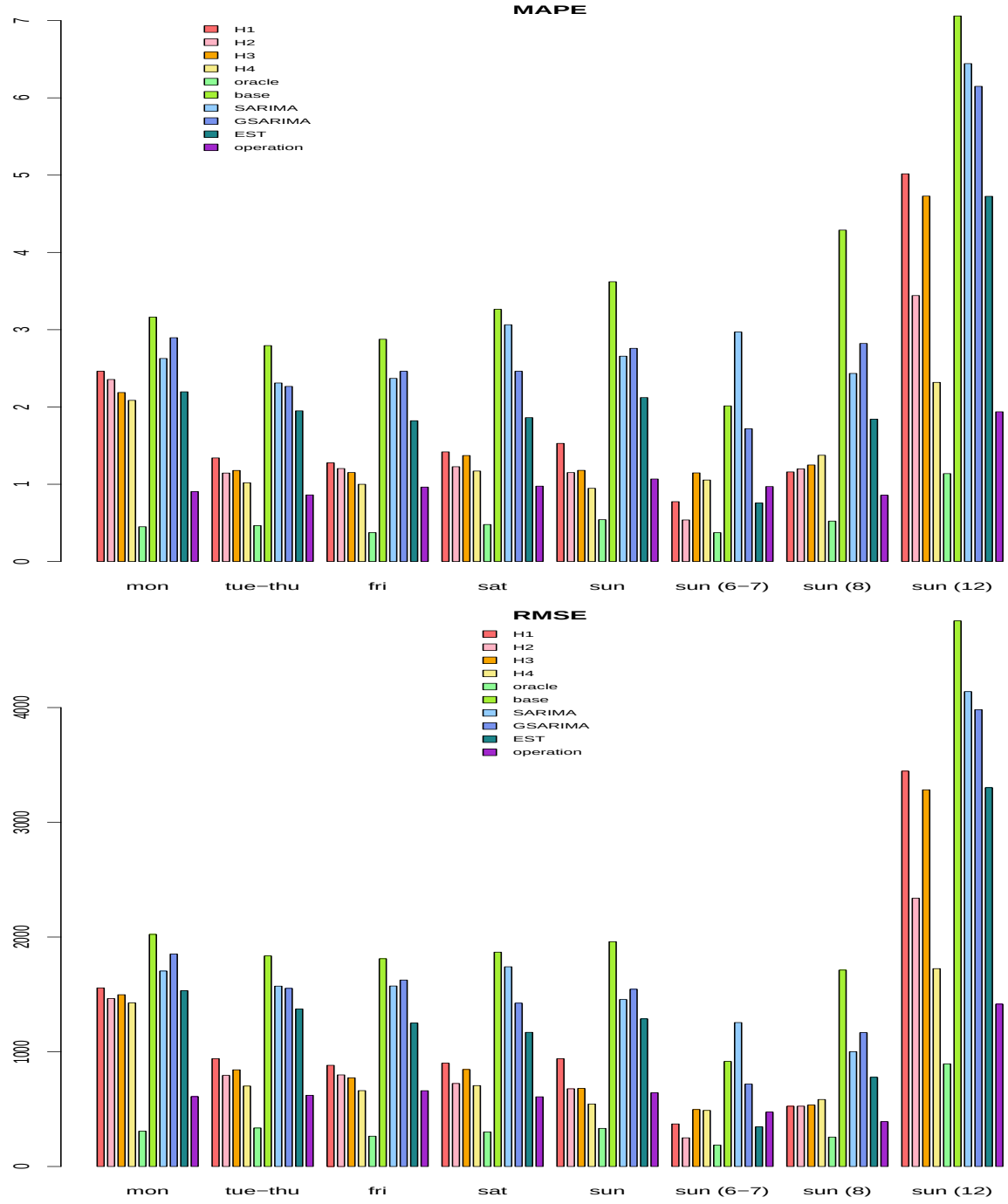

Figure 12: Bar plots of MAPE (top) and RMSE (bottom) with respect to the day type determined by experts; from left to right: Mondays, Tuesdays–Thursdays, Fridays, Saturdays, Sundays (except for June–August and December), Sundays in June–July, Sundays in August and Sundays in December.

When applying the hybrid approach to real-life dataset in Section 4.2, some factors which may have substantial influence over daily electricity loads have not been taken into account. This could have resulted in worsening the performance of our method for winter days when compared to the operational model, and it remains as a task to incorporate such relevant information in our method for practical applications. Also as briefly mentioned in Section 4.2, an adaptive choice of the regressor curve, depending e.g. on the level of temperature, may lead to better results in daily load forecasting. Indeed, an automatic selection of the regressor in the curve linear regression framework may benefit the prediction performance as a generic tool beyond the electricity load forecasting, and we leave the problem for future research.

References

Antoniadis, A., Brossat, X., Cugliari, J., and Poggi, J. M. (2010). Clustering functional data using wavelets. In *Proceedings of the Nineteenth International Conference on Computational Statistics (COMPSTAT)*.

Antoniadis, A., Paparoditis, E., , and Sapatinas, T. (2006). A functional wavelet kernel approach for time series prediction. *Journal of the Royal Statistical Society: Series B*, 68:837–857.

Bathia, N., Yao, Q., and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *Annals of Statistics*, 38:3352–3386.

Bunn, D. W. and Farmer, E. D. (1985). *Comparative Models for Electrical Load Forecasting*. Wiley: New York.

Chiou, J. M. and Li, P. L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B*, 69:679–699.

Chiou, J. M., Müller, H. G., and Wang, J. L. (2004). Functional response models. *Statistica Sinica*, 14:659–677.

Cottet, R. and Smith, M. (2003). Bayesian modeling and forecasting of intraday electricity load. *Journal of the American Statistical Association*, 98:839–849.

Cugliari, J. (2011). *Prévision non paramétrique de processus à valeurs fonctionnelles: Application à la consommation délectricité*. PhD Thesis.

Cupidon, J., Eubank, R., Gilliam, D., and Ruymgaart, F. (2008). Some properties of canonical correlations and variates in infinite dimensions. *Journal of Multivariate Analysis*, 99:1083–1104.

Dordonnat, V., Koopman, S. J., Ooms, M., Dessertaine, A., and Collet, J. (2008). An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, 24:566–587.

Dordonnat, V., Koopman, S. J., Ooms, M., Dessertaine, A., and Collet, J. (2011). Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling. *Computational Statistics and Data Analysis (In press)*.

Engle, R. F., Granger, C. W. J., Rice, J., and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81:310–320.

Eubank, R. and Hsing, T. (2008). Canonical correlation for stochastic processes. *Stochastic Processes and their Applications*, 118:1634–1661.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series*. Springer, New York.

Fan, J. and Zhang, J. T. (1998). Functional linear models for longitudinal data. *Journal of the Royal Statistical Society: Series B*, 39:254–261.

Fan, S. and Hyndman, R. J. (2011). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems (In press)*.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35:70–91.

Hall, P. and Vial, C. (2006). Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society: Series B*, 68:689–705.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102:603–617.

Hannan, E. J. (1961). The general theory of canonical correlation and its relation to functional analysis. *Journal of the Australian Mathematical Society*, 2:229–242.

Harvey, A. and Koopman, S. (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, 88:1228–1253.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Bodels*. Chapman & Hall/CRC.

He, G., Müller, H. G., and Wang, J. L. (2000). Extending correlation and regression from multivariate to functional data. In Puri, M., editor, *Asymptotics in Statistics and Probability*, pages 197–210.

He, G., Muller, H. G., and Wang, J. L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, 85:54–77.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18:439–454.

James, G. M. and Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, 98:397–408.

Lam, C. and Yao, Q. (2011). Factor modelling for high-dimensional time series: inference for the number of factors. *Submitted*.

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

Pierrot, A., Laluque, N., and Goude, Y. (2009). Short-term electricity load forecasting with generalized additive models. In *Proceedings of the Third International Conference on Computational and Financial Econometrics (CFE)*.

Ramanathan, R., Engle, R., Granger, C. W. J., Vahid-Araghi, F., and Brace, C. (1997). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, 13:161–174.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis (with discussions). *Journal of the Royal Statistical Society: Series B*, 53:539–572.

Ray, S. and Mallick, B. (2006). Functional clustering by bayesian wavelet methods. *Journal of the Royal Statistical Society: Series B*, 68:305–332.

Serban, N. and Wasserman, L. (2005). Cats: Clustering after transformation and smoothing. *Journal of the American Statistical Association*, 100:990–999.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, 24:1–24.

Smithies, F. (1937). The eigenvalues and singular values of integral equations. In *Proceedings of the London Mathematical Society*, pages 255–279.

Taylor, J., de Menezes, L. M., and McSharry, P. E. (2006). A comparison of univariate methods for fore-casting electricity demand up to a day ahead. *International Journal of Forecasting*, 22:1–16.

Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204:139–152.

Taylor, J. W. and Buizza, R. (2002). Neural network load forecasting with weather ensemble predictions. *IEEE Transactions on Power Systems*, 17:626–632.

Taylor, J. W. and McSharry, P. E. (2008). Short-term load forecasting methods: an evaluation based on european data. *IEEE Transactions on Power Systems*, 22:2213–2219.

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

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. CRC Press.

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

Yang, W., Müller, H. G., and Stadtmüller, U. (2011). Functional singular component analysis. *Journal of the Royal Statistical Society: Series B*, 73:303–324.

Yao, F., Müller, H. G., and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100:577–590.