Electricity Demand Forecasting Using a Functional State Space Model

Komi Nagbe, Jairo Cugliari, Julien Jacques

Electricity Demand Forecasting Using a Functional State Space Model

Komi Nagbe[a,b,*], Jairo Cugliari[b], Julien Jacques[b]

[a] *Enercoop*
[b] *Université de Lyon, Lyon 2, ERIC EA3083, Lyon, France*

Abstract

In the last past years the liberalization of the electricity supply, the increase variability of electric appliances and their use, and the need to respond to the electricity demand in the real time had made electricity demand forecasting a challenge. To this challenge, many solutions are being proposed. The electricity demand involves many sources such as economic activities, household need and weather sources. All this sources make hard electricity demand forecasting. To forecast the electricity demand, some proposed parametric methods that integrate main variables that are sources of electricity demand. Others proposed non parametric method such as pattern recognition methods. In this paper we propose to take only the past electricity consumption information embedded in a functional vector autoregressive state space model to forecast the future electricity demand. To estimate the parameters of this model we use likelihood maximization, spline smoothing, functional principal components analysis and Kalman filtering. The principal advantage of this model is to forecast electricity demand without taking into account exogenous variables in case of stationary We have seen that in that case the results of the model are competitive and not competitive for non stationary case. But in case of non stationary, this model allows to integrate exogenous variables.

Keywords: Electricity demand forecasting; Functional state space model; Kalman filtering; Functional data; Spline smoothing; Functional principal components analysis.

1. Introduction

Important recent changes in electricity markets make the electricity demand and production forecast a current challenge for the industries. Market liberalization, increasing use of electronic appliances and the penetration of renewable electricity sources are just a few of the numerous causes that explains the current challenges [6]. On the other side, new sources of data are becoming available notably with the deployment of smart meters. However, access to these individual consumers data is not always possible (when available) and so aggregated data is used to anticipate the load of the system.

While only recorded at some time points (e.g. each hour, half-hour or quarter-hour), the electricity load of the system is a continuum. From this, one may consider mathematically the load curve as a function of time with some regularity properties. In fact, electrical engineers and forecaster usually represent the load curve as a function instead of a sequence of discrete measures. Then, one may study the electrical load as a sequence of functions. Recently, attention has been paid to this kind of setting which is naturally called functional time series (FTS). A nice theoretical framework to cope with FTS is within the autoregressive Hilbertian processes, defined through families of random variables taking values on a Hilbert space [4, 15]. These processes are strictly stationary and linear which are two constrictive assumptions to model the electrical load. An alternative to linearity was proposed in [2] where the prediction of a function is obtained as a linear combination of past observed segments, using the weights induced by a notion of similarity

*Corresponding author. komi.nagbe@enercoop.org

between curves. Although the stationary assumption of the full time series is still too strong for the electrical load data [1], corrections can be made in order to render the hypothesis more reasonable. First, one may consider that the mean level of the curves presents some kind of evolution. Second, the calendar structure creates on the data at least two different regimes: workable and non workable days. Of course the specific form of the corrections needed should depend on the nature of the model used to obtain the predictions.

State-space models (SSM) are an interesting alternative connected to non linearity and non stationary patterns of the electrical data. Let us mention some references where SSM have been used to forecast load demand. In [7], the authors propose to describe the hourly load curve as a set of 24 individual regression models that share trends, seasons at different levels, short-term dynamics and weather effects including non linear functions for heating effects. The equations represent 24 univariate stochastically time-varying processes which should be estimated simultaneously within a multivariate linear Gaussian state space framework using the celebrated Kalman filter [9]. However, the cumbersome of the computational burden is a drawback. A second work circumvents the problem by using a dimension reduction approach which reasonably resizes the problem into a handy number of dimension which render the use of the Kalman filter practicable [8]. Some successful uses of SSM to cope with functional data (not necessarily time series) are reported in literature. For instance by using common dynamical factor as in [14] to model electricity price and load, or as in [11] to predict yield curves of some financial assets. Also [19] where railway supervision is performed thanks to a new online clustering approach over functional time series using SSM.

Inspired by these ideas, we push forward the model in [8] to describe now a completely functional autoregressive process whose parameter may eventually vary on time. Indeed, at each time point (days in our practical case) the whole functional structure (load curve) is described through the projection coefficients on a spline basis. Then, using a functional version of principal components analysis, the dimension of the representation is reduced. The vector of spline coefficients is then used as a multivariate autoregressive process, as in [13]. Thus, our approach is completely endogenous but with the ability of incorporating exogenous information (available at the time of the forecast) as covariates.

This paper will be structured as follow. In Section 2 we describe the model we propose for forecasting electricity demand. We present the functional data, functional data representation in splines basis, the state space model that we propose and model estimation methods. Section 3 is proposed to show model inference on a simulated dataset. We will talk about Kalman filtering and smoothing, functional principal component analysis. Section 4 will describe the experiments we make on real data with simple application of our procedure. We then explore, in Section 5, some corrections and extension to the simple approach in order to take into account some of the non stationary patters present in the data. The article concludes in Section 6 where some future work lines are discussed.

2. Materials and methods

The starting point of our modeling is a univariate continuous-time stochastic process $Z = \{Z(t), t \in \mathbb{R}\}$. To study this process, an useful device [4] is to consider a second stochastic process $X = \{X_i(t), i \in \mathbb{N}, t \in [0, \delta]\}$ which is now a discrete-time process and at each time step it takes values on some functional space. The process X is derived from Z as follows. For a trajectory of Z observed over the interval $[0, T], T > 0$, we consider the n subintervals of form $I_i = [(i-1)\delta, i\delta], i = 1, \ldots, n$ such that $\delta = T/n$. Then, we can write

$$X_i(t) = Z((i-1)\delta + t), \qquad t \in [0, \delta]\ i = 1, \ldots, n.$$

With this, anticipate the behavior of Z on say $[T, T+\delta]$ is equivalent to predict the next function $X_{n+1}(t)$ of X. The construction is usually called a *functional time series* (FTS). The setting is particularly fruitful when Z presents a seasonal component of size δ. In our practical application, Z will represent the underlying electrical demand, δ will be the size of a day and so X is the sequence of daily electrical loads. Notice that X represents a continuum which is not necessarily completely observed. As mention on the

introduction the records of load demand are only sampled at some discrete grid. We will discuss on this issue below.

2.1. Prediction of functional time series

The prediction task involves making assertions on the future value of the series $X_{n+1}(t)$ having observed the first n elements $X_1(t), \ldots, X_n(t)$. From the statistical point of view one may be interested on the predictor

$$\tilde{X}_{n+1} = E[X_{n+1}|X_1, \ldots, X_n], \tag{1}$$

which minimizes the L_2 prediction error given the available observations at moment n. A useful model is the (order 1) Autoregressive Hilbertian (ARH(1)) process defined by

$$X_{i+1}(t) - \mu(t) = \rho(X_i(t) - \mu(t)) + \epsilon_i(t), \tag{2}$$

where μ is the mean function of X, ρ is a linear operator and $\{\epsilon_i(t)\}$ is a strong white noise sequence of random functions. Under mild conditions, Equation (2) defines a (strictly) stationary random process (see [4]). The predictor (1) for the ARH(1) process is $\tilde{X}_{n+1}(t) = \mu(t) + \rho(X_n(t) - \mu(t))$ which depends generally on two unknown quantities: the function μ and the operator ρ. The former can be predicted by the empirical mean $\hat{\mu}(t) = \bar{X}_n(t)$. The alternative for the latter is to predict ρ by say $\hat{\rho}_n$ and obtain the prediction $\hat{X}_{n+1} = \hat{\mu}_n + \hat{\rho}_n(X_n - \hat{\mu}_n)$, or to estimate directly the application $\rho(X_n - \hat{\mu}_n)$ of ρ over the last observed centered function. Both variants needs an efficient way to approximate the possibly infinite size of either the operator ρ or the function $\rho(X_n - \hat{\mu}_n)$ which are then estimated (see discussion below on this point).

The inherent linearity of (2) makes this model not flexible enough to be used on electricity load forecast. Indeed, having only one (infinite-dimensional) parameter to describe the transition between any two consecutive days is not reasonable. Variants have been studied. We may mention [3] which incorporate weights in (2) making the impact of recent functions more important; the doubly stochastic ARH model that considers the linear operator ρ to be random [10]; or the conditional ARH where an exogenous covariate drives the behavior of ρ [5]. In the sake of more flexibility, we aim to make predictions on a time-varying setting where the mean function $\mu(t)$ and the operator ρ are allowed to evolve.

2.2. Spline smoothing for functional data

In practice, one only disposes a finite sampling $\mathbf{x} = \{x(t_j), j = 1, \ldots, N\}$ observed eventually with noise, from the trajectory $x(t)$ of the random function $X(t)$. Then, one wishes to approximate $x(t)$ from the discrete measurements. A popular choice is to develop $x(t)$ over the elements of a L_2 basis $\phi_1(t), \ldots, \phi_k(t), \ldots$, that is to write

$$x(t) = \sum_k \tilde{y}_k \phi_k(t) \tag{3}$$

where the coefficients $\tilde{y}_k =< x(t), \phi_k(t) >$ are the coordinates resulting of projecting the function x on each of the elements of the basis. Among the several basis usually used, we choose to work with a B-spline basis because they are adapted to cope with the nature of the data we want to model and have nice computational properties.

B-splines is a basis system adapted to represent splines. In our case, we use cubic spline that is 3th-order polynomial piecewise functions. The connections are made at points called knots in order to join-up smoothing, that is warranting the continuity of the second order derivative. An appealing property of B-spline is the compact support of its elements which gives good location properties as well as efficient computation. Figure 1 illustrates this fact from the reconstruction of a daily load curve. The B-spline elements have a support defined over compact subintervals of horizontal axis.

Another important property is that at each point of the domain, the sum of the spline functions is 1. Since the shape of the spline functions on the border knots are clearly different, this fact is clearly observed

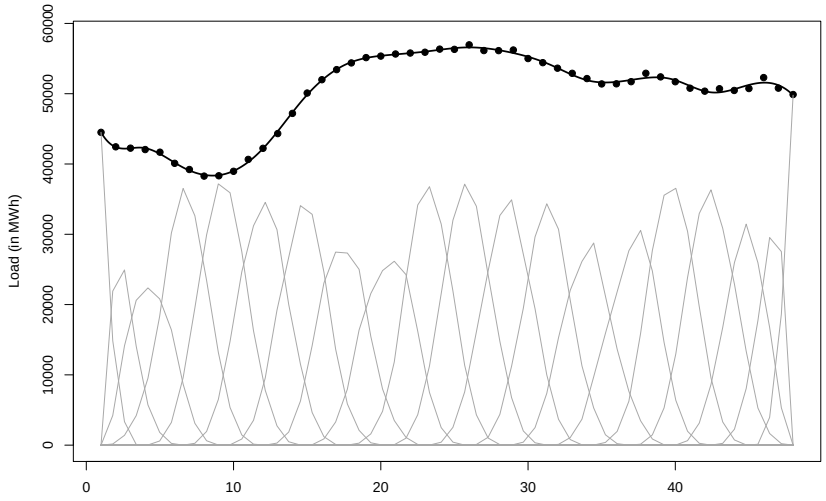


Figure 1: Illustration of the representation of a daily load curve (thick line) by a rank 20 B-spline bases (thin lines).

on the extreme points of the horizontal axis where only one spline has a non null value. Together with the regularity constrains and the additional knots on the extreme the support, these points are subject to a boundary condition. Figure 1 illustrates this important issue concerning the behavior of the boundaries. To avoid this undesirable effect we will use a large number of spline functions on the basis that empirically allows to reduce the boundary condition.

2.3. Functional principal components analysis

Like in multivariate data analysis, Functional Principal Components Analysis (FPCA) provides a mechanism to reduce the dimension of the data by a controlled lost of information. Since data in FDA are of infinite dimension, some care must be given to the sense of dimension reduction. Indeed, what we look for is a representation of the functions like the one in (3) with a relatively low number of basis functions which are now dependent on the data. Moreover, if we demand also that the basis functions form an orthonormal system, then the solution is given by the elements of the eigendecomposition of the associated covariance operator (i.e. the functional equivalent to the covariance matrix) [17].

However, the problem is that these elements are functions and so of infinite dimension. The solution is to represent themselves into a functional basis system (for instance the one presented on the precedent section). Thus, the initial curve $x(t)$ can be approximated in the eigenfunctions basis system:

$$x_i(t) = \sum_{k}^{p} y_{ik}\xi_k(t) \tag{4}$$

where the number p of eigenfunctions, expected to be relatively small, will be chosen such according to the error of approximation of the curves.

Since the representation system may be non orthogonal then it can be shown that the inner product needed in FPCA is connected to the properties of the representational basis system.

Then, the notion of dimension reduction can be understood when one compares the lower number of eigenfunctions with respect to the number of basis functions needed to represent an observation. FPCA reduction of representation dimension which will yield on dramatical drop of the computational time of the model we describe next.

2.4. State space model

State Space Models (SSM) are a powerful useful tool to describe dynamic behavior of time evolving processes. The shape of the load curve may present long term changes which induce non stationary patterns on the signal. Taking into account these changes is one of the challenge of electricity demand forecast.

The linear SSM [9] includes two terms. An inertial term in the form of an intrinsic state of the whole system being modeled. The observed output is a function of the state, some covariate and a noise structure. The state evolution over time is modeled as a linear equation involving the previous state and other observed variables summarized in a vector η. The general formulation is given by:

$$\left\{ \begin{array}{lcl} \mathbf{y}_i & = & \mathbf{z}_i\alpha_i + \epsilon_i \\ \alpha_{i+1} & = & T_i\alpha_i + R_i\eta_i \end{array} \right. , \tag{5}$$

where $\mathbf{y}_i$ is the target variable observed at time i, $\mathbf{z}_i \in \mathbb{R}^{m+1}$ is a vector of predictors, the state at time i is represented as $\alpha_i \in \mathbb{R}^{m+1}$, T_i and R_i are known matrices, and ϵ_i and η_i are the noise and disturbance processes usually assumed to be independent Gaussian with zero-mean and its respective covariance matrices H_i and Q_i which usually contains unknown parameters.

The evolution of the states are useful to understand the system. Using the celebrated Kalman Filter and smoothing, one is able to extract information about the underlying state vector [9]. The one-step-ahead prediction and prediction error are respectively

$$\begin{array}{lcl} a_{i+1} & = & E[\alpha_{i+1}|\mathbf{y}_1, \ldots \mathbf{y}_i] \\ v_{i+1} & = & \mathbf{y}_i - \mathbf{z}_i a_i. \end{array}$$

Also their associated covariance matrices are of interest so let us define $P_{i+1} = \mathrm{Var}(\alpha_{i+1}|\mathbf{y}_1, \ldots \mathbf{y}_i)$ and $F_i = \mathrm{Var}(v_i) = \mathbf{z}_i P_i \mathbf{z}_i' + H_i$. Since these definitions uses recursion an important step is its initialization. When the observations are unidimensional an exact diffusion method can be used from uninformative diffuse prior. However, the method may fail with multivariate observations because the diffusion phase can yield into a non invertible F_i matrix. Moreover, even when F_i is invertible computations become slow due to its dimensionality. It is however possible to obtain an univariate alternative representation of (5) which theoretically reduces computational cost of the Kalman filter and allows one to use the diffuse initialization.

Inference on SSM can be obtained by maximum likelihood. From the Kalmar filter, the log-likelihood can be written as

$$\log L(Y_n|\psi) = -\frac{np}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(\log(|F_i|) + v_i' F_i^{-1} v_i),$$

where $Y_n = \{(\mathbf{y}_1, \mathbf{z}_1), \ldots, (\mathbf{y}_n, \mathbf{z}_n)\}$ and all the unknown parameters are in the parameter vector ϕ. Notice that $\log L$ quantity depends on data and parameters only through the prediction error v_i term and its associate covariance matrix F_i. Both quantities are obtained from the application of the Kalman filter. The maximum likelihood estimation if obtained by numerical optimization of (slight variants of) $\log L$.

2.5. A functional state space model

Approaches of SSM in continuous-time also exists. For instance, [9] presents the simple mean level model. There, the random walk inherent to the state equation is replace by a Brownian motion that drives the time-varying mean level. Early connections between FDA and SSM yielded on derivations of a SSM with the help of FDA. For example, [20] use spline interpolation to approximate the behavior of a time dependent system which is described by a space model.

Our choice is to keep the time discrete by allowing the observations to be functions or curves. A similar idea is behind the model in [19] where functions are used to represent observation on a SSM model but only dependence between states is considered.

Let us consider the vector $\mathbf{y}_i$ as the p FPCA scores resulting from the projection of $x_i(t)$, the load curve for day i, into the eigenfunctions basis system. Then, we may represent an autoregression system by replacing the covariate $\mathbf{z}_i$ by the past load curve, or more precisely by its spline coefficients $\mathbf{y}_{i-1}$.

We propose the following Functional State Space Model (FSSM),

$$\begin{cases} \mathbf{y}_i = \mathbf{y}_{i-1}\alpha_i + \epsilon_i \\ \alpha_{i+1} = \alpha_i + \eta_i. \end{cases} \tag{6}$$

As before the disturbance terms ϵ_i and η_i follow independent Gaussian distribution with zero mean vector and generally unknown variance matrices H_i and Q_i. The sizes of these matrices are in function of p, the number of FPCA scores retained on the approximation step discussed above.

In order to keep the computation time under control while keeping some flexibility on the modeling, we focus on three structural forms of matrices H_i and Q_i: full, diagonal and null; which yields on 6 possible models. Table 1 summarizes the variants as well as the number of parameters to be estimated on the covariance matrices. The complexity of the variant grows while going from 1 to 6. When Q_i is null, then the state equation establishes that states are simple constant on time. Diagonal structures on Q_i and H_i assumes that all the correlations are null and so only variance terms are to be treated. Conversely, full structures allows for a maximum of flexibility letting all the covariances to be free. However, the important drawback of dimensionality becomes crucial since the number of terms to be estimated if of order p^4.

Variant	H_i	Q_i	nb. of param.
1	Diagonal	Null	p
2	Diagonal	Diagonal	$p + p^2$
3	Diagonal	Full	$p + p^4$
4	Full	Null	p^2
5	Full	Diagonal	$p^2 + p^2$
6	Full	Full	$p^2 + p^4$

Table 1: Variants considered for the model (6) showing different structures of matrices H_i and Q_i and number of unknown parameters as function of p.

The FSSM we propose is a SSM on the FPCA scores. Another choice could have been to apply the SSM directly on the spline basis coefficients $\tilde{\mathbf{y}}_i$, but such choice would be computationaly too expensive. It is illustrative to link these dimensions to the problem of electricity demand forecasting. Recall that the number of daily records on our load curves is 48 (sampled at half-hourly) which is prohibited to treat within our framework. Even if this number can be easily divide by two using spline approximation, the number of coefficients would be still too high. Moreover, since the spline coefficients can not be considered independent, one would need to use full diagonal structures on the covariance matrices H_i and eventually on Q_i. Last, the choice we make to reduce the dimension by using FPCA approach is then justified since with a handy number of eigenfunctions, say less than 10, most of the variants discussed above can be easily computed.

3. Experiments on simulated data

We illustrate in this section our approach to forecast using the proposed functional state space models on a functional time series. There are three steps in our approach. First, we approximate the initial data using a B-spline basis. Then a FPCA is performed using the B-splines approximations of the curves. Finally, a fit of the FSSM is obtained. Prediction can then be done by applying the recursion equations on the last estimated state. The resulting predicted coefficients are then put into the functional expansion equations (see Equation (3) and Equation (4)) to obtain the predicted function. For the experiments we

use the statistical software R [16] to fit our model with the packages `fda` [18] for spline approximation and FPCA computation and `KFAS` [12] for the FSSM estimation.

3.1. Simulation scheme

Let us consider a process Y generated as follows

$$\begin{aligned} Y(t) &= \beta_0 + \beta_1 m_1(t) + \beta_2 m_2(t) + \epsilon(t) \\ m_1(t) &= \cos(2\pi t/64) + \sin(2\pi t/64) \\ m_2(t) &= \cos(2\pi t/6) + \sin(2\pi t/6) \\ \epsilon(t) &= \nu(t) + \theta\nu(t-1) + \sigma_2 \end{aligned} \quad . \tag{7}$$

where $\nu(t)$ is strong white noise process (i.e. an independent and identical distributed zero-mean normal random variables $\mathcal{N}(0, \sigma_2)$. Following [2] we set $\beta_0 = 8$, $\beta_1 = 0.8$ and $\beta_2 = 0.18$, $\theta = 0.8$ and $\sigma_2 = 0.05$. Expression (7) is evaluated on discrete times ranging from 1 to $\delta \times n$ where n is the number of functions of length $\delta = 64$. Then we consider the segments of length δ as a discrete sampling of some unobserved functional process.

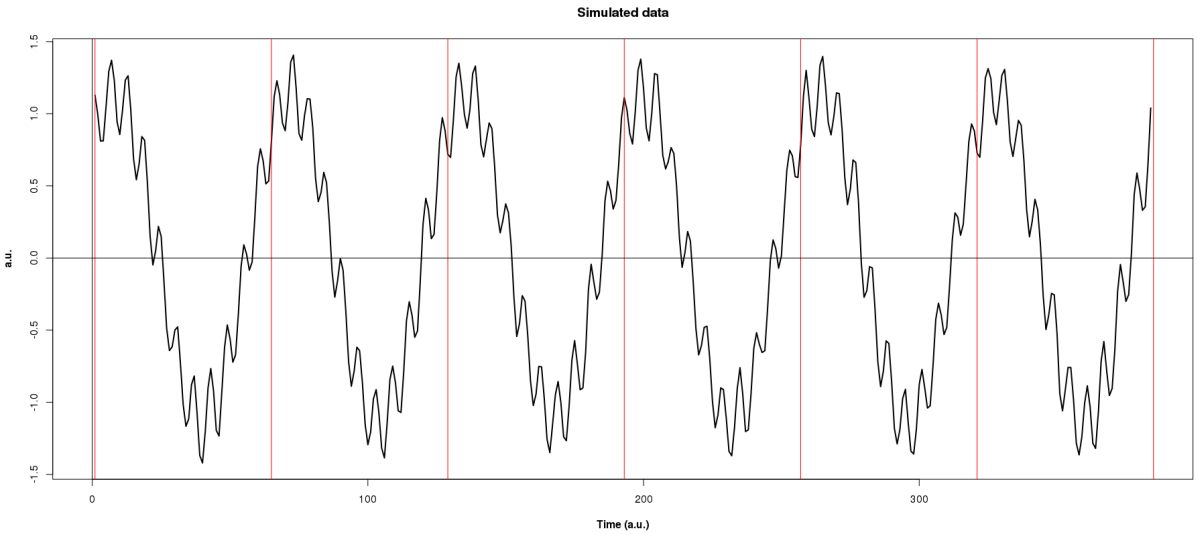


Figure 2: Simulated signal generated via model (7).

Figure 2 represents a time window of the simulated data generated through model (7). Notice that the signal is composed of two additive sinusoidal terms of different frequency plus a moving average structure for the noise term in order to mimic the double seasonal structure of load curves.

3.2. Actual prediction procedure

For each model variant we build and fit the FSSM with the first 26 segments on the simulated signal. That is, each segment is projected on the B-spline basis, and these projections are used into a FPCA. We let the number p of principal components as a tune parameter of the whole procedure. Parameters are estimated and the states are filtered and smoothed as described in Section 2. The last state, together with the last segment coefficients are then used to predict the coefficients of the next segment of the signal which is naturally not used on the estimation of the model. Using the reconstruction expression the actual predicted segment is obtained which closes a prediction cycle.

In order to provide more robust prediction measures, several prediction cycles are used where a sequential increment of the train dataset is done. In what follow we report results on 4 prediction cycles following the describe one-segment-ahead rolling basis.

3.3. How do we measure prediction quality

There are three steps through which the prediction quality must be measured: the splines representation quality of the initial functions $\mathbf{x}$, the functional principal components representation of $\mathbf{x}$, and the forecasting. For all these three steps of quality measurement we use the RMSE (Root Mean-Square-Error) and the MAPE (Mean Absolute Percentage Error). For one-step-ahead forecasting of vector $\mathbf{x}_i$ on time i, if we consider the length of $\mathbf{x}_i$ as h ($h = \delta$ in this case) these metrics are defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{h}(\mathbf{x}_i(t) - \hat{\mathbf{x}}_i(t))^2}{h}}, \qquad \text{MAPE} = \frac{100}{h}\sum_{t=1}^{h}\frac{|\mathbf{x}_i(t) - \hat{\mathbf{x}}_i(t)|}{|\mathbf{x}_i(t)|}.$$

The RMSE is measured in the scale of the data (e.g. kWh for our electricity demand data), and MAPE is expressed in percentage. Notice that MAPE can not be calculated if target variable is zero at some time point. While this is quite unlikely in practice, our simulated signal may present values quite close to zero making MAPE to be unstable. However, this measure is useful to compare prediction performance between signals of different mean magnitude.

3.4. Results

3.4.1. Spline representation and reconstruction

To represent the simulated data we use cubic splines using a regular grid for the knots (with augmented knots on the extremes). To avoid cutting down predictive power of our forecast model we may want to retain here as many spline coefficients as possible (in our case 63). However, we have to make special point here since a boundary condition may yield on artefacts on the spline coefficients near the boundaries. A simple way to reduce this problem was to choose this number of splines (and so the length of the interior knots) to be about 59. This choice produces reasonable quality reconstructions with a MAPE error less than 0.18%.

3.4.2. Functional principal components

The reconstruction quality of the initial functions highly depends on the number of principal components. Of course, the quality of the forecasts will also be impacted by this choice.

Table 2: MAPE and RMSE for the reconstruction step using 59 splines, and 2, 3 or 4 functional principal components.

	Reconstruction error			
	Spline	2 FPC	3 FPC	4 FPC
RMSE	0.0023	0.1282	0.0258	0.0249
MAPE (%)	0.1800	15.0600	2.7900	2.7200

Table 2 reports the reconstruction quality as mean MAPE and RMSE for 2, 3 and 4 principal components.

3.4.3. Forecasting results

In this topic we discuss on the forecasting errors for each choice of the structure of the matrices Q_i and H_i. We take cases of null, diagonal and full matrices Q_i and H_i, as described in Table 1. Table 3 reports RMSE and MAPE values for the forecasting of the simulation data. Both mean and standard deviation are presented. Better prediction performances produce lower MAPE and RMSE. On the one hand, as expected, the number of principal components retained has a large impact on the mean prediction

performance. When only 2 principal components are kept, the prediction error is unreasonably large due to a poor reconstruction. On the other hand, there is no clear advantage for any variant since the standard deviations are large enough to compensate any pairwise difference. This is mainly due to the very small number of prediction segments. Variants with null Q_i matrix are slightly more performant (e.g. smaller errors). This would indicate that a static structure is detected where no time-evolving parameters are needed to predict the signal which is the true nature of the simulated signal.

Table 3: MAPE and RMSE for the forecasting in function of the number of principal components for the simulated signal and for the 6 model variants. Mean values are obtained from 4 one-step-ahead predictions. Standard deviations are reported in parenthesis.

	MAPE			RMSE		
Variant (H_i/Q_i)	2	3	4	2	3	4
1. Diag / Null	18.14 (6.37)	3.32 (0.47)	3.34(0.46)	0.1753 (0.0291)	0.0279 (0.0012)	0.0279 (0.0017)
2. Diag / Diag	31.00 (14.89)	3.63 (0.4)	3.6 (0.55)	0.1832 (0.0330)	0.0279 (0.0018)	0.0280 (0.0024)
3. Diag / Full	23.38(8.74)	3.68(0.48)	3.92(0.59)	0.1832 (0.0330)	0.0279 (0.0018)	0.0280 (0.0024)
4. Full / Null	18.15(6.34)	3.32(0.47)	3.34(0.46)	0.1753 (0.0291)	0.0279 (0.0013)	0.0279 (0.0017)
5. Full / Diag	18.15(6.39)	3.39(0.5)	3.95(0.99)	0.1832 (0.0330)	0.0279 (0.0018)	0.0280 (0.0024)
6. Full / Full	18.96(7.59)	3.53(0.53)	3.96(0.73)	0.1832 (0.0330)	0.0279 (0.0018)	0.0280 (0.0024)

Finally, we compare now the variants from the computational time needed to obtain the prediction. We can see in Table 4 that differences in computing times are significant since standard deviations are quite small. For a fixed number of principal components, there is a clear ranking that can be obtained where the more parsimonious structures produce smaller computing times. Conversely, when the number of principal components increases the computation time increases. However, the increment is more important for the variants of covariances matrices having more parameters.

Table 4: Computing time (in seconds) for the whole procedure by number of principal components and for the 6 model variants. Mean values are obtained from 4 one-step-ahead predictions of the simulated signal. Standard deviations are reported in parenthesis.

Variant (H_i/Q_i)	2	3	4
1. Diag / Null	0.24 (0.03)	0.29 (0.05)	0.38 (0.01)
2. Diag / Diag	0.66 (0.26)	0.47 (0.05)	13.5 (4.96)
3. Diag / Full	4.13 (0.32)	26.24 (10.88)	319.23 (31.98)
4. Full / Null	0.38 (0.10)	1.73 (1.97)	19.5 (14.84)
5. Full / Diag	1.20 (0.22)	6.63 (1.6)	34.82 (7.76)
6. Full / Full	6.91 (0.50)	52.78 (14.33)	421.22 (12.11)

4. Experiments on real electrical demand data

We treat now the case of the electricity load data from the French electricity supplier ENERCOOP[1]. The utility has several kind of customers such as householders, industries as well as small and medium-sized enterprise (SME) with different profiles of electricity consumption. Householders for example, use electricity mainly for lightning, heating and, sometimes cooling and cooking. We work with the aggregated electricity data that is the simple sum of all the individual demands.

We first introduce the data paying special attention to those salient features that are important for the prediction task. Then, we introduce a simple benchmark prediction technique based on persistence

[1] `enercoop.fr`

that we compare to the naive utilization of our prediction procedure. Next, in Section 5, we study some simple variants constructed to cope with the existence of non stationary patterns.

4.1. Description of the dataset

Let us use some graphical summaries to comment on the features of these data. Naturally we adopt the perspective of time series analysis to describe the demand series. Figure 3 represents the dataset which consists in half-hourly sampled records over 6 years going from January 1, 2012 to December 31, 2017. Vertical bars delimits years which are shown on top of the plot. Each record represent the load demand measured in kWh.

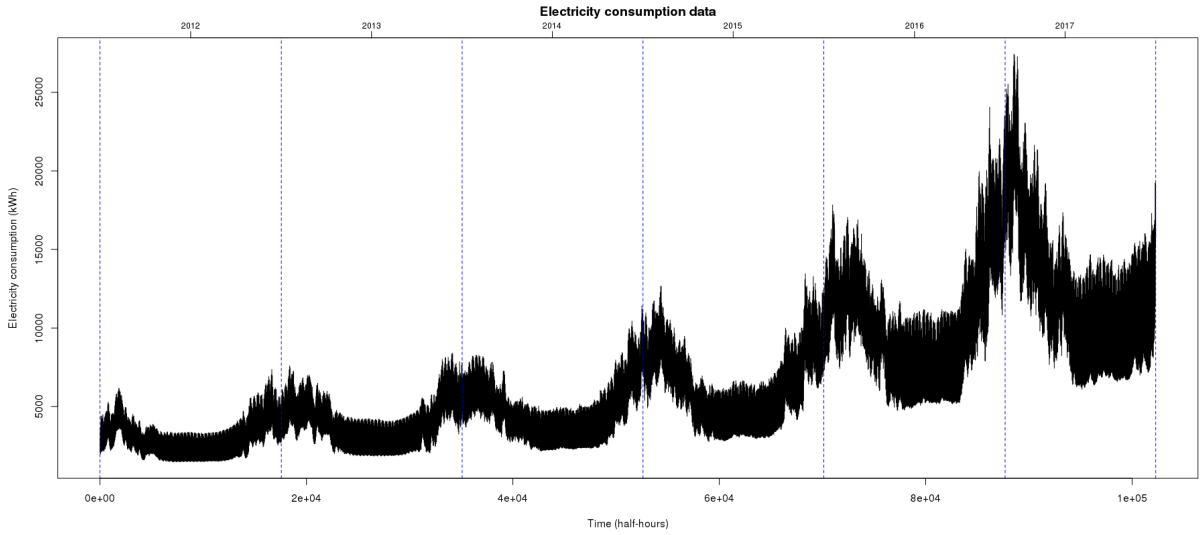


Figure 3: Electricity demand for the supplier ENERCOOP during six years.

First, we observe a clear up growing long-term trend that is connected to a higher variability of the signal at the end of the period. Second, an annual cycle is present with lower demand levels during summer and higher during winter. Both the industrial production calendar and the weather impacts on the demand explain this cyclical pattern.

Figure 4 shows the profile of daily electricity consumption for a period of three weeks (from October 31, 2016 to November 20, 2016). This unveils new cycles presented in data that can be seen as two new patterns: the weekly one and the daily one. The former is the consequence of how human activity is structure around the calendar. Demand during workable days is larger than during weekend days. The latter is also mainly guided by human activity with lower demand levels during night, usual plateau during afternoon and peaks during evening. However, a certain similarity can be detected among days. Indeed even if the profile of Fridays is slightly different, the ensemble of workable days share a similar load shape. Similarly, the ensemble of weekends form a homogeneous group. Holiday banks and extra days off may impact also on the demand shape. For instance in Figure 4 the second segment, which corresponds to the 1st November, is the electrical demand on a bank Holiday. Even if this occurs on a Tuesday, the shape of load of this days and the preceding one (usually an extra day off) are closer to weekend days.

We may also inspect the shape of the load curve across the annual cycle. A simple way to do this is to inspect the mean form between months. Figure 5 represents the mean load on four months, one per season of the year. Some differences are easy to notice, for instance the peak demand is during afternoon in Autumn and Winter but at midday in Spring and Summer. Other are more subtle, like the effect of daylight saving time clock change which horizontally shifts the whole demand. Transitions between the

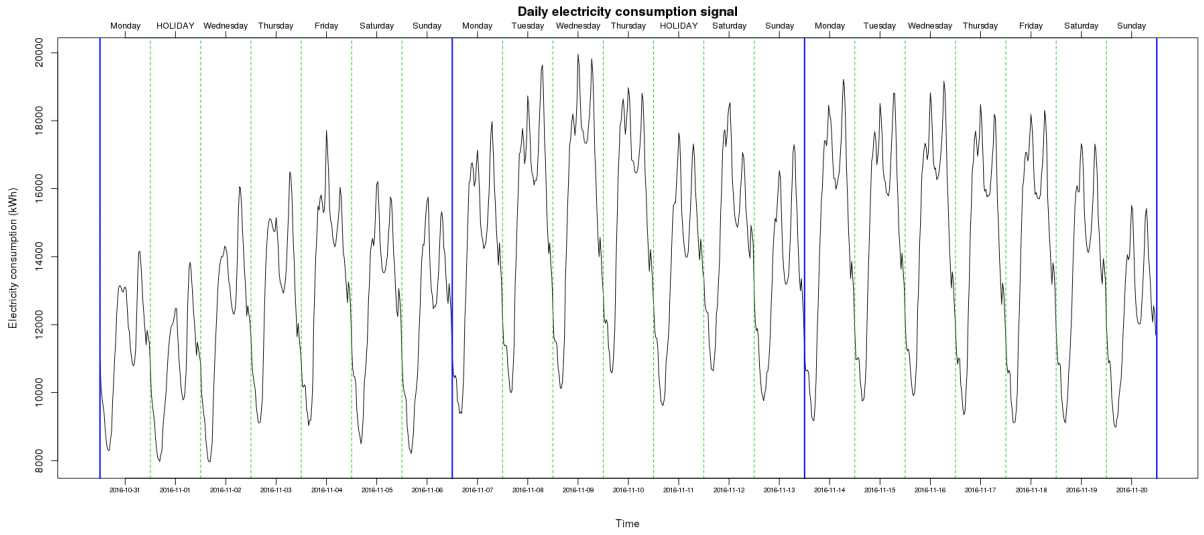


Figure 4: Electricity consumption from October 31, 2016 to November 20, 2016.

cold and warm seasons are particularly sensitive for the forecasting task, specially when the change is abrupt.

4.2. Spline and FPCA representation

As before, we report on the reconstruction error resulting from the spline and FPCA representations.

Table 5: RMSE and MAPE between the splines approximation and the electrical load data in function of the number of splines.

	Number of splines				
	12	24	40	45	47
MAPE (%)	1.310	0.480	0.160	0.060	0.010
RMSE (kWh)	130.030	48.800	19.480	8.860	3.850

For comparison purposes, we compute the error criteria for 5 choices on the number of splines (12, 24, 40, 45 and 47 splines) on the reconstruction of the coded functions. Table 5 shows the MAPE and RMSE between the reconstruction and the real load data. As expected, the lower the number of splines the higher the reconstruction error. This shows that using only spline interpolation our approach is not pertinent because a relatively large number of spline coefficients is needed. The extremest case of 12 splines, which would make the computing times reasonable, produces a too large MAPE of 1.310% which hampers the performance of any forecasting method based on this reconstruction. On the other extreme, using 47 cubic splines to represent the 48-length discrete signal of the daily demand produces boundary effects that will dominate the variability of the curves.

Since spline approximation is connected to the FPCA in our approach, we may check the reconstruction quality for all the choices issued from the crossing of the selected number of splines and a number of principal component between 2 and 10. Table 6 shows the MAPE and RMSE of the reconstructions obtained by each of the possible crossings. We may target a maximum accepted MAPE value of 1% in reconstruction. Then, the options are just a few, most of them with very close MAPE values. In what follows, we choose to work with 45 cubic splines and 10 principal components.

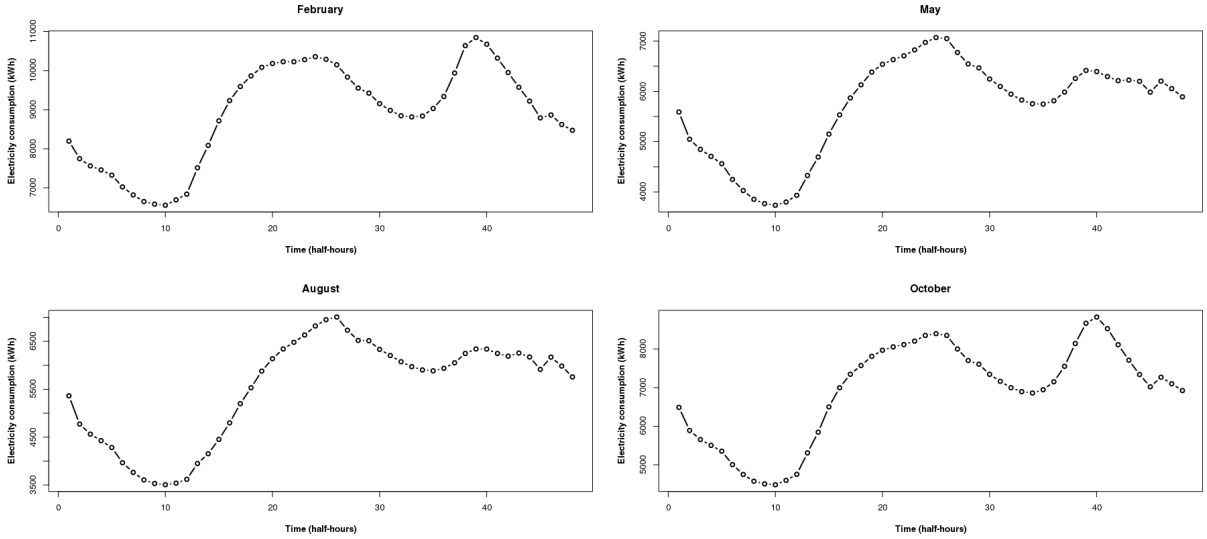


Figure 5: Mean of electricity daily consumption for four months: February, May, August and October.

Table 6: RMSE and MAPE errors for splines smoothing load data reconstitution via FPCA in function of the number of splines and the number of principal components.

	MAPE (in %) nb. of splines					RMSE (in kWh) nb. of splines				
nb PC	12	24	40	45	47	12	24	40	45	47
2	3.400	3.250	3.320	3.770	4.190	343	331	332	351	385
3	2.560	2.340	2.420	2.940	3.690	250	232	233	260	333
4	2.120	1.860	1.960	1.820	2.780	205	180	182	177	262
5	1.770	1.440	1.550	1.400	2.280	176	145	149	141	217
6	1.650	1.250	1.200	1.210	1.770	160	121	116	116	168
7	1.540	1.120	1.040	1.060	1.370	148	103	97	98	135
8	1.440	0.900	0.820	0.830	1.160	139	84	76	76	109
9	1.390	0.830	0.740	0.770	0.960	135	75	65	68	92
10	1.350	0.760	0.680	0.750	0.790	132	69	58	64	71

4.3. Forecasting results

Forecasting is done in a rolling basis over the whole last year in the data. The first five years are used as initial training dataset. Predictions are obtained at horizon one-segment ahead. This means that actually we are making predictions for the next 48 time steps (1 day) if we adopt a traditional time series point of view. Once the prediction is obtained, we compare it with the actual data and incorporate the observation into the training dataset. Thanks to the recursion in the SSM only an update step is necessary here.

To give a comparison point to our methodology we propose to use a simple but powerful benchmark based on persistence forecasting.

4.3.1. Persistence-based forecasting

A persistence-based forecasting method for a functional time series equals to anticipate $X_{n+1}(t)$ with the simple predictor $\hat{X}_{n+1}(t) = X_n(t)$. Thus, the predictor can be connected to the ARH model (Equation (2)) where the linear operator is the identity operator $\rho = Id$. However, this approach is not convenient

since there are two groups of load profiles in the electricity demand given by workable days and the other days (e.g. weekends or holidays). We use then a smarter version of the persistence model which takes into account the existence of these two groups. The predictor is then written

$$\hat{X}_{n+1}(t) = \begin{cases} X_n(t) & \text{if day } n \text{ is Monday, Tuesday, Wednesday or Thursday} \\ X_{n-7}(t) & \text{otherwise} \end{cases} . \quad (8)$$

Table 7 summaries the MAPE on prediction by type of day for the persistence-based forecasting method. We can observe the global MAPE errors is several times the reconstruction error we observed above. There is a clear distinction between those days predicted by the previous day and the other ones (`i.e.` Saturdays, Sundays and Mondays). The lack of recent information for the latter group is severe drawback and impacts on its prediction performance. The increased difficult of predicting bank holidays is translated into the highest error levels.

Table 7: MAPE (in %) for the persistence-based forecasting method.

	Minimum	1st quartile	Median	3rd quartile	Maximum	Mean (sd)
Monday	0.86	3.10	4.99	9.82	24.41	7.48 (5.72)
Tuesday	0.66	1.52	2.03	5.00	13.97	3.61 (2.86)
Wednesday	0.47	1.39	2.23	3.11	11.34	2.85 (2.25)
Thursday	0.39	1.26	2.27	3.87	10.99	2.78 (2.12)
Friday	0.27	1.45	2.33	3.73	11.55	2.74 (1.78)
Saturday	0.27	3.04	6.85	11.05	24.21	7.60 (5.61)
Sunday	0.42	2.80	6.11	9.15	21.41	6.86 (5.11)
Bank Holiday	1.13	6.27	10.80	12.65	25.73	10.90 (5.82)
Global	0.27	2.60	4.70	7.30	25.73	5.60 (1.78)

4.3.2. FSSM forecasting

We now present the results for the proposed FSSM. Only the variant 1 in Table 1 is used, namely we consider the covariance matrix of the observations H_i as diagonal and the covariance matrix of the states Q_i as null. The reason is twofold. First, we show on simulations that basic models give as satisfactory results as the more involved ones. Second, computing time must be kept into reasonable standards for the practical application.

Table 8: Daily MAPE (in %) on prediction for the FSSM forecasting.

	Minimum	1st quartile	Median	3rd quartile	Maximum	Mean (sd)
Monday	1.27	2.49	3.19	4.21	10.97	3.66 (1.61)
Tuesday	0.79	1.64	2.64	3.63	7.96	2.93 (1.57)
Wednesday	0.83	1.52	2.20	3.67	11.20	2.65 (1.4)
Thursday	0.77	1.70	2.47	3.93	8.98	2.90 (1.71)
Friday	0.77	1.66	2.41	3.12	8.70	2.59 (1.21)
Saturday	0.83	2.59	4.32	6.19	19.98	4.66 (2.89)
Sunday	1.26	3.91	5.94	7.80	19.98	6.07 (2.55)
Bank Holiday	2.20	5.35	6.03	6.86	11.20	6.18 (2.26)
Global	0.77	2.61	3.65	4.93	19.98	3.96 (1.21)

Tables 8 and 9 show the MAPE on prediction for days and months respectively for the FSSM approach. In comparison with the persistence-based forecasting, the global error is sensibly reduced with improvement on almost all day types. Also improvement are observed in all the months of the year but August (results

Table 9: Monthly MAPE (in %) on prediction for the FSSM forecasting.

	Minimum	1st quartile	Median	3rd quartile	Maximum	Mean (sd)
January	0.79	2.15	2.86	4.21	8.54	3.39 (1.81)
February	1.11	2	2.535	3.57	4.36	2.72 (0.94)
March	1.27	2	3.82	5.42	10.71	4.05 (2.29)
April	1.01	2.54	3.67	5.81	10.44	4.31 (2.37)
May	0.83	1.99	3.19	6.03	8.70	3.85 (2.26)
June	0.93	1.69	2.43	4.11	10.54	3.25 (2.11)
July	1.26	2.8	3.94	5.05	7.37	3.97 (1.46)
August	0.77	1.45	2.43	5	11.20	3.69 (3.1)
September	0.83	2.14	2.87	5.41	19.98	4.29 (3.67)
October	1.04	1.66	2.7	3.69	19.98	3.22 (2.01)
November	1.45	3.26	4.415	5.68	9.48	4.65 (1.95)
December	1.28	1.58	2.59	4.26	8.98	3.09 (1.81)

for persistence-based forecasting are not shown). If we look at the distribution of MAPE we see that the range is compressed with a lower maximum error but also a higher minimum error. This last effect is the price to pay for having an approximate representation of the function. We may think the MAPE on approximation as a lower bound for the MAPE on forecasting. The higher this bound the more limited the approach is. Despite this negative result, the gain on the largest errors observed before more than compensates the increment on the minimum MAPE and yields on a globally better alternative. Among workable days, Mondays presents the higher MAPE. FSSM being an autoregressive approach, it builds on the previous days which present a different demand structure. Moreover, the mean load level of these two consecutive days is sharply different. Undoubtedly incorporating the calendar information would help the model to better anticipate this kind of transition. Both mean load level corrections and calendar structure are modification or extensions that can be naturally incorporated in our FSSM. We discuss some clues for doing this in the next section.

5. Corrections to cope with non stationary patterns

We explore two kind of extensions to add exogenous effects. In the first one the days are grouped into two groups, `workable` and `non workable` days, and the prediction is done separately in each group. In the second one, the day and the month are introduced as exogenous fixed effect in the model.

5.1. Adding effects as grouping variables

We aim to tackle some of the difficulties non stationary patterns imposes on the forecasting of load data by explicitly consider two groups of days: `workable` (i.e. Monday, Tuesday, Wednesday, Thursday, Friday), and `non workable` (i.e. Saturday, Sunday and Holidays). We adapt our model FSSM described in Equation (6) by adapting it on each group separately, that is we consider the model

$$\begin{cases} \mathbf{y}_i^g = \mathbf{y}_{i-1}^g \alpha_i^g + \epsilon_i^g \\ \alpha_{i+1}^g = T_i^g \alpha_i^g + \eta_i^g \end{cases} \quad , \qquad g \in \texttt{workable, non-workable} \tag{9}$$

The only difference between models (6) and (9) is the data used in estimation of parameters. In the case of (9), we have two groups of data, workable days and, non workable days. The structure of the matrices are the same as in the model 6, but now they are specific to each group of days. In term of forecasting procedure, if we want to predict a workable day, we choose the data for the group `workable`. Similarly for non workable days. In Tables 10 and 11) we present MAPE obtained with this procedure

reported by day type and month. The results for this approach globally improve the forecast with respect the initial model since the global MAPE decreases. Also, reduced MAPE are obtained for most of the day types. However, we can see also that for Saturday and Monday the error are significantly increased. These days are those where the transition between the groups occur. They share then the additional difficult of not having the most recent information (that one from the precedent day) in the model. Some cold months prediction is not improved. Improvement are observed during summer months or months around summer. The high level load demand and variability of winter months impacts on the rise of prices and thus makes these months of particular interest. The accuracy in forecasting for these months is important because bad prediction can have a large economical impact for electricity suppliers.

Table 10: Daily MAPE (%) for model (9).

	Minimum	1st quartile	Median	3rd quartile	Maximum	Mean (sd)
Monday	0.92	2.61	3.68	7.42	15.85	5.10 (3.38)
Tuesday	0.64	1.20	2.02	3.24	15.85	2.51 (1.83)
Wednesday	0.60	1.39	2.03	2.71	8.62	2.18 (1.12)
Thursday	0.66	1.43	2.01	3.10	6.76	2.42 (1.3)
Friday	0.75	1.47	2.15	3.23	9.51	2.53 (1.32)
Saturday	1.16	2.39	3.24	7.46	27.72	5.69 (5.34)
Sunday	1.15	2.01	2.33	3.67	8.06	3.00 (1.6)
Bank Holiday	1.39	3.77	4.51	5.88	9.51	4.89 (1.92)
Global	0.60	2.03	2.75	4.59	27.72	3.54 (1.32)

Table 11: Monthly MAPE (%) for model (9).

	Minimum	1st quartile	Median	3rd quartile	Maximum	Mean (sd)
January	0.75	1.75	3.16	4.37	13.88	4.26 (3.8)
February	0.60	1.63	2.515	3.98	14.82	3.60 (3.24)
March	0.82	1.81	3.23	5.36	22.10	4.40 (4.48)
April	1.14	2.16	2.885	4.11	8.62	3.71 (2.25)
May	0.91	2.01	2.39	3.63	9.51	2.91 (1.77)
June	0.90	1.54	2.2	3.15	4.78	2.35 (1)
July	0.96	1.47	2.01	2.78	4.51	2.12 (0.85)
August	1.16	2.04	2.63	3.24	9.24	2.84 (1.41)
September	1.08	1.93	2.32	3.08	4.39	2.49 (0.86)
October	0.66	1.2	1.91	3.57	10.40	2.84 (2.53)
November	1.40	2.26	4.36	6.35	13.04	4.93 (2.87)
December	0.64	1.56	2.8	5.29	27.72	4.27 (5.05)

One thing we can also do to improve forecasting of load demand, is to integrate some exogenous variables such as days types and weather variables. Day types are fixed variables and weather variables are random variables. In this paragraph, we have just implicitly introduced days types in our modeling but not as exogenous variable. In the next paragraph we introduce in our model the variable day types.

5.2. Adding effects as covariates

In this paragraph, we introduce in model (6) the variables day type and months. We must have an appropriate presentation of this exogenous deterministic variables before predicting. We choose to create for each day of the week, a binary dummy variable. In total we have eight days (including Holidays) type which we use as eight dummies variables. Each variable takes values in $\{0, 1\}$. The value 1 corresponds to the case where the response vector $\mathbf{y}_i$ is observed on the same day as the day which is being used. For

example, if the response variable $\mathbf{y}_i$ is observed on Sunday, the dummy variable for Sunday take value 1. The dummy variable for Sunday will take 0 if the response vector $\mathbf{y}_i$ is not observed on Sunday, but $\mathbf{y}_i$ takes 1 for other day dummy variable on which it was observed. In addition we choose a numeric variable which represents the months of the year. We would like to control the seasonal effect of the series with this variable. The months variable has 12 values representing the month of the year. Let's choose **Day** as days dummy variables and **Month** as the numeric months variable. The model (6) becomes:

$$\begin{cases} \mathbf{y}_i = \mathbf{Day}_i\beta_i^D + \mathbf{Month}_i\beta_i^M + \mathbf{y}_{i-1}\alpha_i + \epsilon_i \\ \alpha_{i+1} = T_i\alpha_i + \eta_i. \end{cases} \tag{10}$$

$\mathbf{Day}_i \in \{0,1\}^{1\times 8}$, $\beta_i^D \in \mathbb{R}^{8\times p}$, $\mathbf{Month}_i \in \{1,\ldots,12\}$ and $\beta_i^M \in \mathbb{R}^{1\times p}$. $\mathbf{Day}_i$ and $\mathbf{Month}_i$ are fixed in the time but β_i^D and β_i^M are not fixed in time because the profile of the series changes with time. These regression coefficients can be interpreted as estimation of each day and month mean profiles of the series. Lets note $\beta_i = \{\beta_i^D, \beta_i^M\}$. In term of mixed linear models modeling, β_i controls the fixed effects of the load data. Parameters estimation of (10) is bit different when using the package `KFAS` which allows to estimate states space models parameters. In the case of the model (6), we choose random approach to estimate α_i, which consequently explains the randoms part of the load curves. That means we assume Q_i exists.

Table 12: Daily MAPE (in%) for model (10).

	Minimum	1st quartile	Median	3rd quartile	Maximum	Mean (sd)
Monday	1.48	2.35	2.81	3.96	8.90	3.30 (1.39)
Tuesday	0.65	1.51	2.16	2.93	8.06	2.56 (1.46)
Wednesday	0.65	1.44	2.04	2.73	8.06	2.21 (1.12)
Thursday	0.78	1.41	1.87	2.97	7.50	2.42 (1.41)
Friday	0.77	1.42	1.93	3.39	8.39	2.39 (1.43)
Saturday	0.77	2.14	3.04	4.44	13.11	3.55 (1.96)
Sunday	0.96	3.02	4.49	6.02	13.11	4.67 (1.92)
Bank Holiday	1.83	3.10	4.28	7.12	8.14	4.90 (2.03)
Global	0.65	2.05	2.83	4.20	13.11	3.25 (1.43)

Table 13: Monthly MAPE (in%) for model (10).

	Minimum	1st quartile	Median	3rd quartile	Maximum	Mean (sd)
January	0.90	1.87	2.59	3.86	8.14	3.10 (1.68)
February	1.24	1.925	2.33	2.9	4.44	2.52 (0.88)
March	0.94	1.81	3.08	5.09	8.90	3.50 (1.95)
April	1.34	2.17	3.38	4.78	8.06	3.75 (2)
May	0.82	2.03	3.07	5.34	7.12	3.40 (1.76)
June	0.77	1.34	1.985	2.93	8.03	2.46 (1.52)
July	0.87	1.83	2.65	3.37	5.19	2.61 (1.08)
August	0.78	1.41	2.1	3.3	6.71	2.61 (1.63)
September	0.65	1.4	2.2	3.16	13.11	2.74 (2.24)
October	0.89	1.61	2.15	3.56	13.11	2.68 (1.5)
November	1.62	3.35	4.205	5.48	9.70	4.57 (1.88)
December	1.13	1.61	2.26	3.58	7.86	2.85 (1.65)

It can be observed that, the global MAPE decreases. On Monday, the MAPE decreases significantly. It is still difficult to have accurate prediction on winter and spring months, except December and February.

For the month of November, the great value of MAPE is due to unprecedented increase of the number of customers on November 2016. Figure 6 shows the errors of observations equation and their variance H_i. It can be observed at the end of year 2016 (from November 2016), that there was a big change in the errors variance. This change comes from the change in the rhythm of electricity consumers subscription as clients of ENERCOOP. In the end of year 2016, there were COP21 (21^{st} Sustainable Innovation Forum) in France. As ENERCOP's politic is to support renewable electricity production and consumption, at this moment, the communication of the image of ENERCOOP increased in the population. ENERCOOP was well known in French public than ever. This communication brought many customers to subscribe to electricity supplied by ENERCOOP. For this moment, it is difficult for the model to be accurate because at this moment outliers values were being observed.

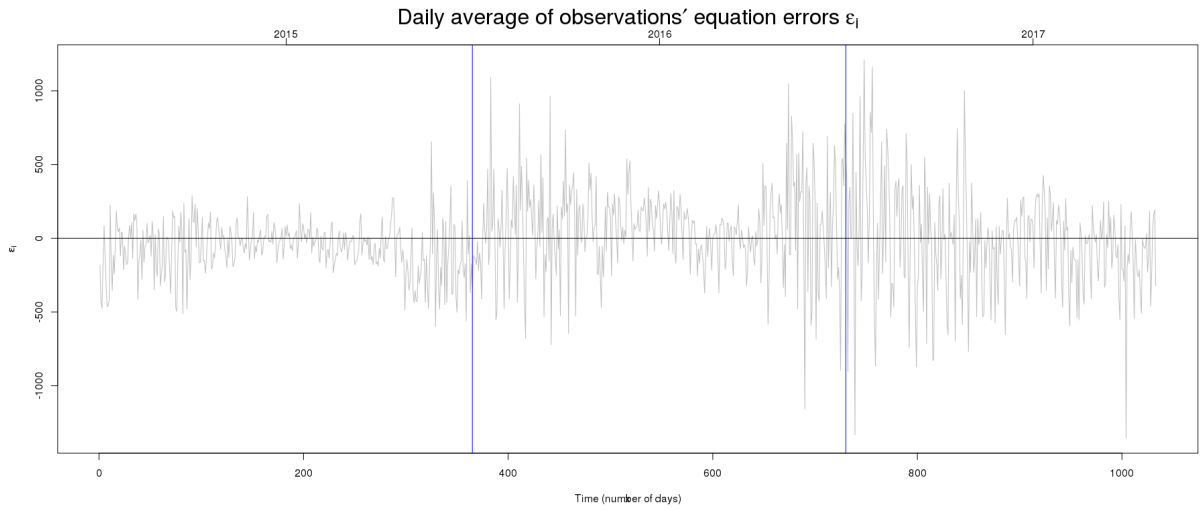
6. Conclusion and discussion

In a concurrent environment, electrical companies need to anticipate load demand from data that presents non stationary patterns induced by the arrives and departs of customers. Forecasting in this context is a challenge since one desires to use as much past data as possible but needs to reduce the usable date to the records that describe the current situation. In this trade-off, adaptive methods have their role to play.

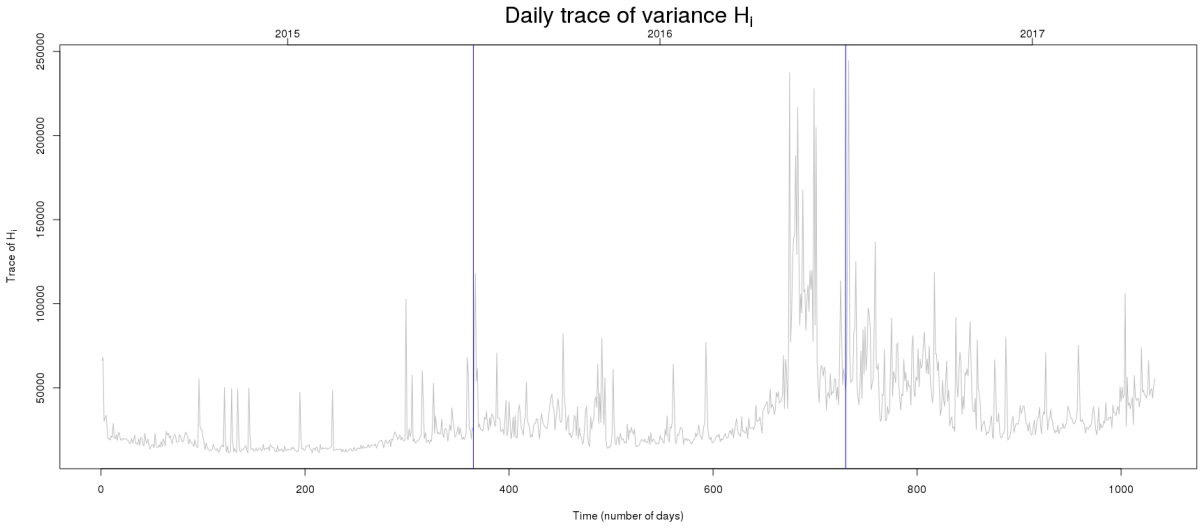
Figure 6 witnesses of the ability that FSSM has to adapt into certain extent to the changing environment. The impact of an external event to the electrical demand is translated into larger variability on the error and so an inflation of the trace of the errors variance matrix (cf. at the end of 2016).

Forecasting process in this paper is mainly endogenous. Only some calendar information is used as exogenous variable. However, in electricity forecasting it is well known that weather have a great influence on the load curve. For instance temperature impacts through the use of cooling systems in hot season and electrical heating during cold seasons. In France this dependence is known to be asymmetrical with higher influence of temperature on cold days. The nature of this covariable on forecasting is different to the ones we used. Indeed, while calendar can be deterministically predicted it is not the case for the temperature. Using forecasted weather on an electrical demand forecast inserts the eventual bias and the uncertainty of the weather forecasting system to the electrical demand prediction. Integration of weather information into our model, eventually changing the structure of the matrix H_i, and obtaining prediction interval for the predictions are perspectives of future work.

Also, only point predictions are obtained. In a probabilistic framework one would like to have not only an idea of the mean level anticipation of the load, but also some elements about the predictive distribution of the forecast. Whether the whole distribution, some quantile levels or a predictive interval, this information is not trivially obtained from our approach. While SSM do provides intervals through the Gaussian assumptions coupled with the estimation of the variance matrices, FSSM have these information on the coefficients of the functional representation. Transporting these information to the whole curve need to be studied.



(a)



(b)

Figure 6: Diagnostics plot from model FSST: (a) Daily mean of errors $\hat{\epsilon}_i$; (b) Daily trace of variance matrix H_i.

References

[1] Antoniadis A, Brossat X, Cugliari J, Poggi JM. Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. application à la consommation d'électricité. Journal de la Société Française de Statistique 2012;153(2):52–78.

[2] Antoniadis A, Paparoditis E, Sapatinas T. A functional wavelet–kernel approach for time series prediction. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2006;68(5):837–57.

[3] Besse P, Cardot H. Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. Canadian Journal of Statistics 1996;24(4):467–87.

[4] Bosq D. Linear Processes in Function Spaces. Springer New York, 2000.

[5] Cugliari J. Non parametric forecasting of functional-valued processes : application to the electricity load. Theses; Université Paris Sud - Paris XI; 2011.

[6] Cugliari J, Poggi JM. Electricity demand forecasting. In: An Hourly Periodic State Space Model for Modelling French. Wiley; 2018. .

[7] Dordonnat V, Koopman S, Ooms M, Dessertaine A, Collet J. An hourly periodic state space model for modelling french national electricity load. International Journal of Forecasting 2008;24(4):566 –87. Energy Forecasting.

[8] Dordonnat V, Koopman SJ, Ooms M. Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling. Computational Statistics & Data Analysis 2012;56(11):3134 –52.

[9] Durbin J, Koopman SJ. Time series analysis by state space methods. volume 38. OUP Oxford, 2012.

[10] Guillas S. Doubly stochastic Hilbertian processes. Journal of Applied Probability 2002;39(3):566–80.

[11] Hays S, Shen H, Huang JZ. Functional dynamic factor models with application to yield curve forecasting. The Annals of Applied Statistics ;6(3):870–94. doi:10.1214/12-AOAS551.

[12] Helske J. Kfas: Exponential family state space models in r. Journal of Statistical Software, Articles 2017;78(10):1–39.

[13] Holmes EE, Ward EJ, Wills K. Marss: Multivariate autoregressive state-space models for analyzing time-series data. The R Journal 2012;4(1).

[14] Liebl D. Modeling and forecasting electricity spot prices: A functional data perspective. The Annals of Applied Statistics ;7(3):1562–92.

[15] Álvarez Liébana J. A review and comparative study on functional time series techniques. arXiv 2017;.

[16] R Core Team . R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria; 2017. URL: https://www.R-project.org/.

[17] Ramsay JO, Hooker G, Graves S. Functional data analysis with R and MATLAB. Springer New York, 2009.

[18] Ramsay JO, Wickham H, Graves S, Hooker G. fda: Functional data analysis, 2011. R package http://CRAN R-project org/package= fda Cited on ;:24.

[19] Samé A, El-Assaad H. A state-space approach to modeling functional time series application to rail supervision - IEEE conference publication. In: Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European. IEEE; 2014. .

[20] Valderrama MJ, Ortega-Moreno M, González P, Aguilera AM. Derivation of a state-space model by functional data analysis. Computational Statistics 2003;18(3):533–46.