

A SARIMAX COUPLED MODELLING APPLIED TO INDIVIDUAL LOAD CURVES INTRADAY FORECASTING

SOPHIE BERCU AND FRÉDÉRIC PROÏA

ABSTRACT. A dynamic coupled modelling is investigated to take temperature into account in the individual energy consumption forecasting. The objective is both to avoid the inherent complexity of exhaustive SARIMAX models and to take advantage of the usual linear relation between energy consumption and temperature for thermosensitive customers. We first recall some issues related to individual load curves forecasting. Then, we propose and study the properties of a dynamic coupled modelling taking temperature into account as an exogenous contribution and its application to the intraday prediction of energy consumption. Finally, these theoretical results are illustrated on a real individual load curve. The authors discuss the relevance of such an approach and anticipate that it could form a substantial alternative to the commonly used methods for energy consumption forecasting of individual customers.

1. INTRODUCTION

The electrical systems have to manage with new challenges: constant increasing demand of electricity, including the arrival of new uses such as electric vehicle, increasing production of renewable energies decentralized, increasing number of energy market participants including aggregators, with a policy of DSM¹ and CO₂ reduction. In particular, we note an increase in peaks load at critical times, intermittent injections of energy at all levels making electrical networks much more difficult to manage, the balance supply/demand more difficult to keep and economic interests much more dispersed. On the other hand, the technological advances are enabling new opportunities: the development of smart-metering and smart grid. Indeed, there is a massive deployment in Europe - 80% of households will be equipped by 2020 - and in North America, mainly. In this context, forecasting consumption of very fine mesh becomes an important issue at the heart of the system of electric power. In this paper, we focus exclusively our attention on the prediction of end-customer on a short-term horizon. These forecasts are useful for the customers who want to optimize their bill and make DSM, for network planning and monitoring, for aggregators who wish to apply real-time load shifting. Forecasting methods on large aggregate of customers exist in abundance but cannot be extrapolated to individual curves because of their extremely irregular behavior. Indeed, aggregation has the

Key words and phrases. SARIMA(X) modelling, Time series analysis, Exogenous covariates, Forecasting, Seasonality, Stationarity, Individual load curve.

¹ For *Demand Side Management*, whose aim is to encourage the consumer to use less energy especially during peak hours.

advantage of reducing the noise and provides little chaotic curves in which trend and seasonality are easily identifiable. For example, in the case of short-term forecasting, the exponentially weighted methods of Taylor [31] give pretty good results. Harvey and Koopman [20] also developed an unobserved components model with time-varying splines to capture the evolution of patterns in hourly electricity loads. Afterwards, bayesian methods that rely on Kalman filter and state space models were suggested by Martin [26], Smith [30] and Cottet and Smith [10]. Multiple time series approaches have also emerged to model and predict intraday aggregated load curves, one can cite in example the heteroskedastic GARCH model of Garcia *et al.* [15], the seasonal ARFIMA-GARCH model of Koopman *et al.* [22], the robust SARIMA estimates of Chakhchoukh *et al.* [7], [8], or the bias reducing iterative algorithm of Cornillon *et al.* [9]. Lately, Dordonnat *et al.* [12] have developed a very comprehensive state space approach for modelling hourly french national electricity load, taking into account different levels of seasonality, calendar events and weather dependence, with one equation for each hour. Semiparametric methods and artificial neural networks to model meteorological effects and seasonal patterns are considered by Liu *et al.* [24]. Poggi [28] proposed a nonparametric approach based on kernel estimators to make forecasts on half-hourly french load curves and more technical studies from Antoniadis *et al.* [1], [2] based on functional kernel-wavelets can also be mentioned. In short, several methods have been implemented of various kinds, from nonparametric to parametric models with exogenous covariates through semi-parametric approaches, heteroskedasticity, state space models and neural networks, each of them providing excellent intraday forecast results on extensively averaged curves, such as national load. Let us conclude by referring the reader to the approach of Devaine *et al.* [11], which is an aggregation of specialized experts combining a set of prediction outputs by independent forecasters.

The issue of individual forecasting is complex and very few literature is available on the subject. The main difficulty of the individual load curves is the deep irregularity resulting from the human behavior. Indeed, we have to deal with phenomena that aggregation usually hides, such as high disturbances, unpredictable local behaviors or thresholds during holiday periods. To the best of our knowledge, very few studies have been conducted in this area. In their work, Espinoza *et al.* [14] are concerned by the short-term load forecasting from a HV-LV substation, and Ghofrani *et al.* [18] propose to model real-time measurement data from customers' smart meters as the sum of a deterministic component and a gaussian noise signal. This paper suggests a statistical parametric approach adapted to an individual load curve which shows a substantial seasonal pattern and a thermosensitive behavior as prerequisites, highly relying on the time series theory. The authors anticipate that it could form an alternative to the commonly used methods for energy consumption forecasting of individual customers. In Section 2, we introduce a dynamic coupled modelling taking temperature into account. We also recall some known results on time series analysis, in particular the concepts of stationarity and causality. We recall some theoretical backgrounds which will be used as a basis for the empirical study on a real curve in the next section. As an example of load curve which we

will investigate, Figure 1.1 displays the energy consumption of a thermosensitive customer and the interpolated temperatures measured every 3 hours by the nearest weather station on the same period of 6 months. Section 3 is devoted to the detailed study on such a load curve based on time series analysis. The main motivations for proposing a coupled modelling are the linear relation between the logarithmic energy consumption and the temperature on the one hand, and the seasonal behavior of the residuals on the other hand. Figure 1.2 represents the scatter plot between temperature and consumption for the same customer, in which one can observe the linear relationship. It also displays the residuals from the linear regression on the right-hand side. One shall investigate seasonality, stationarity and autocorrelations in the residuals from the linear regression, to build a suitable time series modelling and propose a forecasting algorithm, according to some criteria that will be specified. For that purpose, one shall make an extensive use of the well-known Box and Jenkins methodology [3]. A short conclusion is given in Section 4.

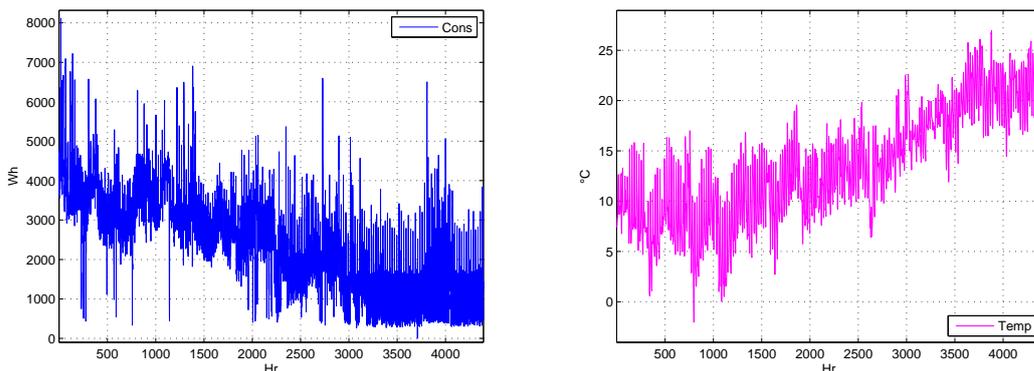


FIGURE 1.1. Energy consumption of a thermosensitive customer (left), temperature during the same period (right).

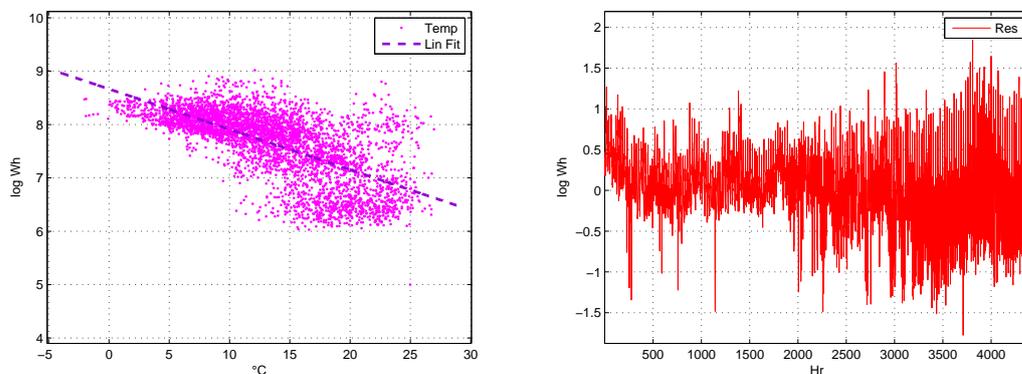


FIGURE 1.2. Scatter plot between energy consumption and temperature (left), residuals from the linear regression (right).

In the case of what we call *thermosensitive* customers, electricity consumption decreases when temperature increases. This phenomenon can be explained mainly

by the use of electric heating². This evolution, relatively to the logarithm of the consumption, is roughly linear - as we see in Figure 1.2. Let us note that at this stage, we do not try to extract the whole information but only part of it. Which is why a linear regression seems sufficient even if non-optimal, since the residuals themselves will be deeply processed subsequently. Of course, the model specification has to be adapted to a different behavior such as the effect of air-conditioning during high temperatures periods, if necessary after studying the underlying scatter plot.

Remark 1.1. In all the sequel, B stands for the backshift operator which operates on an element of a given time series to produce the previous element, $BX_t = X_{t-1}$. The backshift operator is raised to arbitrary integer powers so that $B^s X_t = X_{t-s}$. The difference operator ∇ , defined as $\nabla X_t = (1 - B)X_t$, also generalises to arbitrary integer powers so that $\nabla_s X_t = (1 - B^s)X_t$ and $\nabla^s X_t = (1 - B)^s X_t$.

2. ON A SARIMAX COUPLED MODELLING

The usual time series analysis tools will be repeatedly used throughout the study. The reader will find more details in Chapters 1 and 3 of [5] about the concepts of *stationarity* and *causality*, the *stationary ARMA process*, the *ACF* and *PACF functions*, etc. It is well-known that a stationary solution of an ARMA process defined on \mathbb{Z} does not imply the causality of the underlying process. Besides, for the real curves defined on the positive integers that will be considered in the following, one can observe that the stationarity of the solution and the causality of the process are closely related. Indeed, a zero inside the unit circle results in an explosive behavior of the process that cannot match with stationarity on \mathbb{N}^* , and accordingly we will see a stationary time series as the solution of a causal ARMA(p, q) process that we will try to identify. The hypothesis of stationarity will be tested *via* the commonly used Kwiatkowski-Phillips-Schmidt-Shin *KPSS test* [23] together with the unit root Augmented Dickey-Fuller *ADF test* [29]. In addition, the usual identification methods for the orders of stationary AR(p) and MA(q) processes will also be useful for orders selection. In this connection, let us recall that, for the MA(q) process, the ACF function ρ is such that $\rho(q) \neq 0$ and $\rho(t) = 0$ for all $|t| > q$. Similarly, for the causal AR(p) process, the PACF function α is such that $\alpha(p) \neq 0$ and $\alpha(t) = 0$ for all $t > p$. Chapter 3 of [5] includes all related definitions and results. The hypothesis of white noise will be evaluated through the *portmanteau test* of Ljung-Box [25], [4].

In all the sequel, we denote by (C_t) the individual energy consumption of a given customer, for all $1 \leq t \leq T$. We also denote by (U_t) the temperature associated with (C_t) , supposed to be known for all $-r + 2 \leq t \leq T + H$ where H is the prediction horizon and r is the number of exogenous lags, namely U_{t-r+1}, \dots, U_t explain C_t for all $1 \leq t \leq T$. In addition, we will use a variance-stabilizing Box-Cox logarithmic transformation (Y_t) ensuring homoskedasticity, given, for all $1 \leq t \leq T$, by

$$(2.1) \quad Y_t = \log(C_t + e^\mu)$$

² Electric heating is very common in France, unlike air-conditioning which is still not much widespread.

where μ is a positive parameter, implying that $Y_t = \mu$ in the particular case where $C_t = 0$. This safety precaution is justified by the possible use of relative criteria, such as the *Mean Absolute Percentage Error*, in the evaluations related to Y_t . However, the choice of μ will not be subject to optimization since (2.1) defines a bijection. This translation parameter has to be seen as a positive lower bound for (Y_t) not sensitive as soon as it is not too close to zero, and $\mu = 5$ will be satisfying in all the sequel, according to the dataset.

The dynamic coupled modelling. The first step of the modelling relies in a suitable way to remove the direct influence of the temperature on the consumption. As mentioned above, it exists a strong correlation between (Y_t) and (U_t) . This relationship is modeled through the linear regression given, for all $1 \leq t \leq T$, by

$$(2.2) \quad Y_t = c_0 + \mathcal{C}(B)U_t + \varepsilon_t$$

where $c_0 \in \mathbb{R}$ is an intercept, $\mathcal{C}(B)$ is a polynomial of order r such that, for all $z \in \mathbb{C}$,

$$\mathcal{C}(z) = \sum_{k=1}^r c_k z^{k-1}$$

and the unknown vector parameter $c \in \mathbb{R}^{r+1}$ is estimated by standard least squares. The autocorrelation in the disturbance terms possibly will not produce an optimal and efficient estimator by the method of least squares. Nevertheless, this is not a crucial issue since we only plan to extract some information and to regard (ε_t) as a seasonal time series. In particular, (ε_t) is said to follow a SARIMA(p, d, q) \times (P, D, Q)_s modelling if, for all $1 \leq t \leq T$,

$$(2.3) \quad (1 - B)^d(1 - B^s)^D \mathcal{A}(B)\mathcal{A}_s(B)\varepsilon_t = \mathcal{B}(B)\mathcal{B}_s(B)V_t,$$

according to Definition 9.6.1 of [5], where (V_t) is a white noise of variance $\sigma^2 > 0$, and where the polynomials are defined, for all $z \in \mathbb{C}$, as

$$\begin{aligned} \mathcal{A}(z) &= 1 - \sum_{k=1}^p a_k z^k, & \mathcal{A}_s(z) &= 1 - \sum_{k=1}^P \alpha_k z^{sk}, \\ \mathcal{B}(z) &= 1 - \sum_{k=1}^q b_k z^k, & \mathcal{B}_s(z) &= 1 - \sum_{k=1}^Q \beta_k z^{sk}. \end{aligned}$$

In this modelling, $a \in \mathbb{R}^p$, $b \in \mathbb{R}^q$, $\alpha \in \mathbb{R}^P$ and $\beta \in \mathbb{R}^Q$ are vector parameters estimated by generalized least squares. The differenced process $(\nabla^d \nabla_s^D \varepsilon_t)$ in (2.3) is a stationary solution of the ARMA causal process, *i.e.* $\mathcal{A}(z) \neq 0$ and $\mathcal{A}_s(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.

Definition 2.1 (SARIMAX). In the particular framework of the study, a random process (Y_t) will be said to follow a SARIMAX(p, d, q, r) \times (P, D, Q)_s coupled modelling if, for all $1 \leq t \leq T$, it satisfies

$$(2.4) \quad \begin{cases} Y_t = c_0 + \mathcal{C}(B)U_t + \varepsilon_t, \\ (1 - B)^d(1 - B^s)^D \mathcal{A}(B)\mathcal{A}_s(B)\varepsilon_t = \mathcal{B}(B)\mathcal{B}_s(B)V_t. \end{cases}$$

The orders p, d, q, r, P, D, Q and s shall be evaluated following a well-known Box and Jenkins methodology [3]. Moreover, a straightforward calculation shows that (2.4) can be rewritten in the condensed form given, for all $1 \leq t \leq T$, by

$$(2.5) \quad (1 - B)^d(1 - B^s)^D \mathcal{A}(B)\mathcal{A}_s(B) (Y_t - \mathcal{C}(B)U_t) = \mathcal{B}(B)\mathcal{B}_s(B)V_t,$$

as soon as $d + D > 0$, which will be an assumption always verified as we shall explain in the next section. Indeed, c_0 vanishes by a single differentiation of (ε_t) . In light of foregoing, one can establish the following result, denoting by I the identity matrix of order T , Y and U the observation vector of order T and the design matrix of order $T \times (r + 1)$, respectively given by

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix} \quad \text{and} \quad U = \begin{pmatrix} 1 & U_T & U_{T-1} & \dots & U_{T-r+1} \\ 1 & U_{T-1} & U_{T-2} & \dots & U_{T-r} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & U_1 & U_0 & \dots & U_{-r+2} \end{pmatrix}.$$

Theorem 2.1. *Assume that $U'U$ is invertible. Then, the differenced process $(\nabla^d \nabla_s^D \varepsilon_t)$ where ε_t is given, for all $1 \leq t \leq T$, by the vector form*

$$(2.6) \quad \varepsilon = (I - U(U'U)^{-1}U') Y$$

is a stationary solution of the coupled model (2.4).

Proof. Theorem 2.1 is a direct consequence of Theorem 3.1.1 of [5] together with a straightforward least squares calculation. \square

Remark 2.1. In the particular case where $r = 0$, we merely obtain $\varepsilon = Y - \bar{Y}$ where

$$\bar{Y} = \frac{1}{T} \sum_{k=1}^T Y_k,$$

and (2.4) reduces to the usual SARIMA(p, d, q) \times (P, D, Q) $_s$ modelling on the recentered load curve. In addition, as soon as $d + D > 0$, the influence of \bar{Y} vanishes.

The t -statistic associated with each parameter, exploiting the asymptotic normality of the estimates, will provide a significance testing procedure, as a confirmation of the criteria minimization strategy. Though, they will not be appropriate in the exogenous regression owing to the strong autocorrelation in the residuals, and will only be applied to the time series coefficients.

Application to forecasting. Whatever prediction method one wishes to apply, see *e.g.* Chapters 5 and 9 of [3] or Chapter 5 of [5], the time series analysis of (2.4) provides the predictor of (ε_t) at stage $T + 1$, denoted by $\tilde{\varepsilon}_{T+1}$. Let \hat{c}_T be the least squares estimator of c in (2.2) and assume that the order r is known. Then, it follows that

$$(2.7) \quad \tilde{Y}_{T+1} = \hat{c}_{0,T} + \sum_{k=1}^r \hat{c}_{k,T} U_{T-k+2} + \tilde{\varepsilon}_{T+1}.$$

Via the same lines, since (U_t) is supposed to be known³ for all $-r + 2 \leq t \leq T + H$, the predictor at horizon H is given by

$$(2.8) \quad \tilde{Y}_{T+H} = \hat{c}_{0,T} + \sum_{k=1}^r \hat{c}_{k,T} U_{T-k+H+1} + \tilde{\varepsilon}_{T+H}.$$

3. APPLICATION TO FORECASTING ON A LOAD CURVE

By virtue of Theorem 2.1, the application of the coupled model (2.4) to real curves merely consists in identifying the seasonality and the orders of differencing ensuring the stationarity of the residual sequence from the regression analysis. Moreover, from a careful analysis of the ACF and PACF, we will get a first approximation of the orders to be considered in the ARMA modelling. We shall first investigate seasonality through a Fourier spectrogram, then stationarity of the deseasonalized series and autocorrelations *via* ACF and PACF, and finally the overall randomness of successive innovations. Different models will be suggested and compared using bayesian criteria on the one hand, and then prediction criteria on the other hand. As mentioned above, the KPSS test [23], the ADF test [29] and the Ljung-Box test [25], [4] will be used as statistical procedures for evaluating the hypothesis of stationarity, of unit root and of white noise up to a certain lag, respectively.

From now on, for all $1 \leq t \leq T$, (C_t) is a load curve and (Y_t) is the associated logarithmic process, given by (2.1) with $\mu = 5$. In addition, (U_t) is the exogenous temperature supposed to be known for all $-r + 2 \leq t \leq T + H$ and H is the prediction horizon. Denote also by $(\hat{\varepsilon}_t)$ the least squares estimated residual set from the regression analysis accordingly given, for all $1 \leq t \leq T$, by

$$(3.1) \quad \hat{\varepsilon}_t = Y_t - \hat{c}_{0,T} - \sum_{k=1}^r \hat{c}_{k,T} U_{t-k+1},$$

directly coming from (2.6). For a sake of simplicity, one shall take $r = 1$ without loss of generality. Besides, one observes on real curves that numerical results are very similar when r increases. Indeed, due to the natural phenomenon it represents, temperature U_t at time t is highly correlated to U_{t-1} and the use of lots of regressors to explain C_t in our modelling would often be redundant and generate statistically nonsignificant coefficients. The load curve is represented on Figure 1.1 together with interpolated temperatures measured by the nearest weather station on the same period of 6 months. The load curve that we consider is this one, but extended to 2 years of consumption. In the Table 3.1 below, we present descriptive statistics for this dataset reference : quantiles, mean and dispersion.

Of course, we will not use the whole dataset for modelling. The amounts of data taken in account for building the estimates will be optimized.

³ Later in Remark 3.2, more precisions are given on U_{T+1}, \dots, U_{T+H} which are necessarily predictions themselves.

	Length	Min	Q _{0.1}	Q _{0.25}	Q _{0.5}	Q _{0.75}	Q _{0.9}	Max	Mean	Range	Std Err
C_t (Wh)	17520	0.0000	461.08	858.87	1789.7	2783.8	3684.9	11052.	1972.8	11052.	1298.1
Y_t (lWh)	17520	5.0000	6.4126	6.9150	7.5695	7.9835	8.2515	9.3237	7.4529	4.3237	.67771
U_t (°C)	17520	-2.0316	7.1192	10.295	15.384	20.872	24.300	30.983	15.597	33.014	6.3992

TABLE 3.1. Descriptive statistics for the load curve and the exogenous variable.

Seasonality. Let us choose for example $T = 17520$, that is 2 years of consumption. We consider the k -th empirical Fourier coefficient of $(\hat{\varepsilon}_t)$, given by

$$\gamma_k = \frac{1}{2\pi T} \left| \sum_{t=1}^T \hat{\varepsilon}_t e^{-if_k t} \right|^2$$

where $f_k = 2\pi k/T$ is the k -th Fourier frequency. Figure 3.1 displays the variation of $\sqrt{\gamma_k}$ on the Fourier frequency spectrum of $(\hat{\varepsilon}_t)$ on the left-hand side and the ones of $(\nabla_{12}\hat{\varepsilon}_t)$ and $(\nabla_{24}\hat{\varepsilon}_t)$ on the right-hand side, with unexploitable low frequencies truncated.

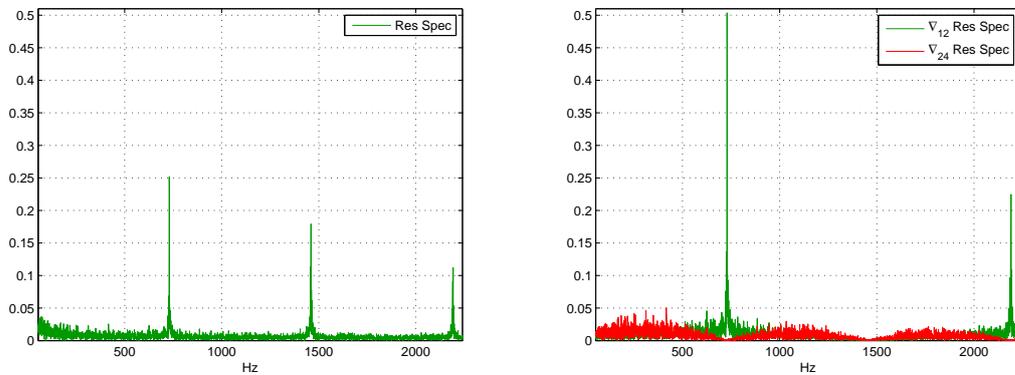


FIGURE 3.1. Fourier spectrograms of residuals (left) and seasonally differenced residuals of period 12 and 24 (right).

Figure 3.1 shows that the estimated residual set $(\hat{\varepsilon}_t)$ has a seasonality and the abscissa of the main peak indicates that the pattern repeats itself 730 times on 2 years, that is daily. The second peak also suggests a seasonality of 12 hours. On the right side, one can see that $(\nabla_{12}\hat{\varepsilon}_t)$ still has a periodicity whereas $(\nabla_{24}\hat{\varepsilon}_t)$ is quasi-a-periodic. This is the reason why one shall choose $s = 24$ in the SARIMA modelling, and that also leads us to choose $D = 1$ in (2.4). By studying lower frequencies, one can also notice that a weekly cycle is likely to appear, but in a far less marked way.

Stationarity. The KPSS and ADF statistical procedures both suggest that, on 6 months of consumption, $(\hat{\varepsilon}_t)$ is not stationary whereas $(\nabla\hat{\varepsilon}_t)$, $(\nabla_{24}\hat{\varepsilon}_t)$ and $(\nabla\nabla_{24}\hat{\varepsilon}_t)$ are stationary. As a consequence, $(\hat{\varepsilon}_t)$ is difference-stationary and the differenced series can all be solutions of a causal ARMA modelling, which leads to ARIMA models with $d = 1$ and SARIMA models with $d = 0$ and $D = 1$ or $d = 1$ and $D = 1$.

Autocorrelations. On the ACF and PACF of $(\hat{\varepsilon}_t)$, one can clearly observe the daily periodicity of the series, as it appears on Figure 3.2.

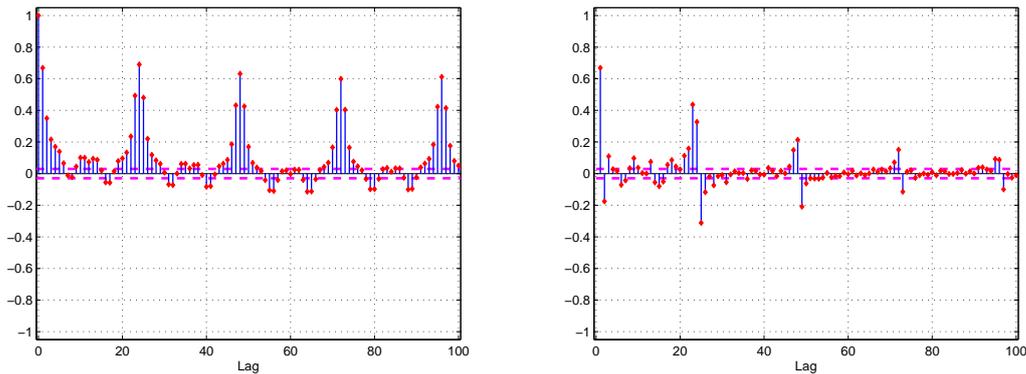


FIGURE 3.2. ACF (left) and PACF (right) of the residuals $(\hat{\varepsilon}_t)$.

In addition, the sample ACF of $(\nabla_{24}\hat{\varepsilon}_t)$ on Figure 3.3 below shows either an exponential decay or a mixture with damped sine wave, while the sample PACF has a relatively large spike at lag 1 and can reasonably be considered as nonsignificant afterwards, with uncertainty up to lag 5. One can also detect a pattern around lag 24 on the ACF. This behavior suggests an $AR(p)$ modelling with a seasonal moving average autoregression on the seasonally differenced series, that is a $SARIMA(p, 0, 0) \times (0, 1, Q)_{24}$ model with $p \leq 5$, and $Q = 1$ on $(\hat{\varepsilon}_t)$.

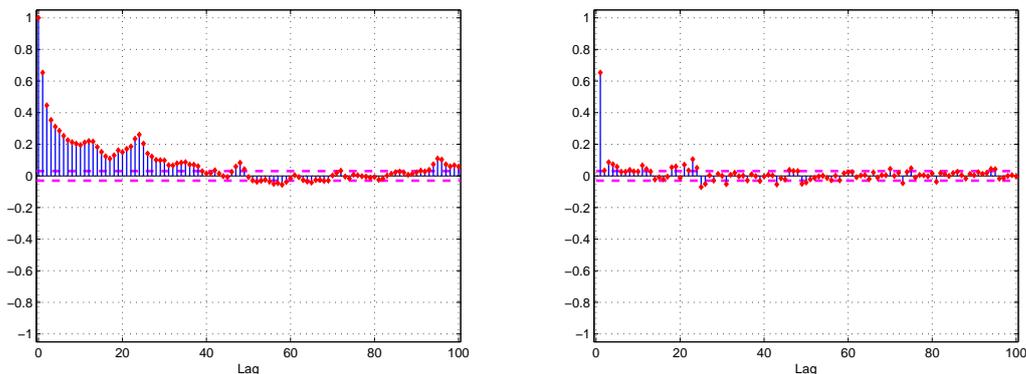


FIGURE 3.3. ACF (left) and PACF (right) of the seasonally differenced residuals of period 24 $(\nabla_{24}\hat{\varepsilon}_t)$.

Finally, Figure 3.4 displays the sample ACF and PACF of $(\nabla\nabla_{24}\hat{\varepsilon}_t)$. One can observe that the PACF tails off exponentially from lag 1 and that the ACF cuts off after lag 2, with small seasonal contributions. As a consequence, the series is likely to be generated by a $SARIMA(p, 1, q) \times (0, 1, Q)_{24}$ process with $p = 1$, $q = 2$ and $Q = 1$.

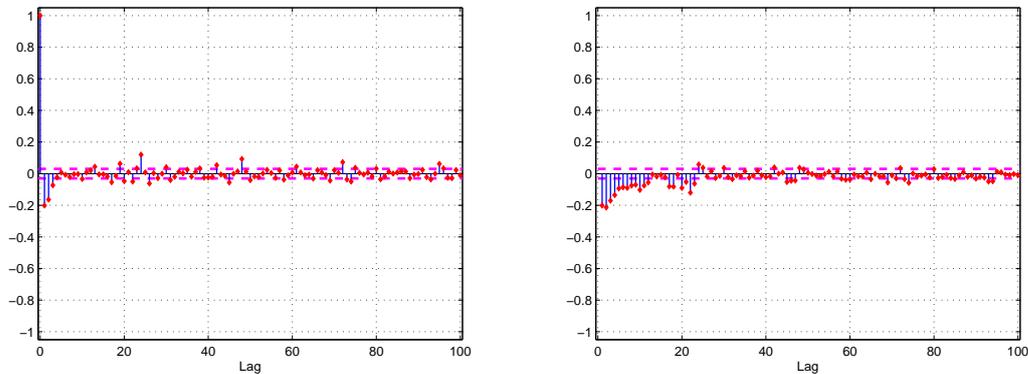


FIGURE 3.4. ACF (left) and PACF (right) of the doubly differenced residuals of period 24 ($\nabla\nabla_{24}\hat{\varepsilon}_t$).

Modelling. This identification methodology seems quite rough, and one shall make the parameters vary in their neighborhood to determine the optimal modelling. Table 3.2 gives the bayesian criteria associated with a set of SARIMAX models fitted on 6 months of consumption. AIC, SBC, LL and WN respectively stand for *Akaike Information Criterion*, *Schwarz Bayesian Criterion*, *Log-Likelihood* and *White Noise*. Let us recall that

$$\text{AIC} = -2 \log \mathcal{L} + 2k \quad \text{and} \quad \text{SBC} = -2 \log \mathcal{L} + k \log T$$

where \mathcal{L} is the model likelihood and k is the number of parameters. In addition, VAR is the estimated variance of (V_t) . The Ljung-Box *portmanteau* test is used to evaluate the hypothesis of white noise on the fitted innovations, considering arbitrarily that (V_t) is a white noise if it has nonsignificant autocorrelations up to lag 3.

	p	d	q	r	P	D	Q	s	AIC	SBC	LL	VAR	WN
SARIMAX	1	0	0	1	0	1	1	24	-362.4	-330.4	186.2	0.053	
SARIMAX	3	0	0	1	0	1	1	24	-393.6	-348.8	203.8	0.053	✓
SARIMAX	5	0	0	1	0	1	1	24	-424.7	-367.2	221.4	0.053	✓
SARIMAX	3	0	2	1	0	1	1	24	-446.9	-389.4	232.5	0.052	✓
SARIMAX	3	0	2	2	0	1	1	24	-457.2	-393.3	238.6	0.052	✓
SARIMAX	0	1	2	1	0	1	1	24	-238.7	-200.4	125.4	0.055	
SARIMAX	1	1	1	1	0	1	1	24	-389.5	-351.2	200.8	0.053	
SARIMAX	2	1	2	1	0	1	1	24	-435.1	-384.0	225.6	0.052	✓

TABLE 3.2. Bayesian criteria associated with a set of SARIMAX models on 6 months of consumption.

Estimations of $a \in \mathbb{R}^p$, $b \in \mathbb{R}^q$ and $\beta \in \mathbb{R}^Q$ related to (2.3) come from an optimized mixing of conditional sum-of-squares and maximum likelihood [6], [13], [16], [19] provided by the software environment, and $c \in \mathbb{R}^r$ is estimated by standard least squares. In order to keep this section brief, only the most representative results are summarized in the table above even if more models have been evaluated. In conclusion, on the basis of the bayesian criteria, the most adequate modelling for

the given load curve is a $\text{SARIMAX}(3, 0, 2, 2) \times (0, 1, 1)_{24}$ whose explicit expression is as follows, for all $28 \leq t \leq T = 4380$,

$$(3.2) \quad \begin{cases} Y_t = c_0 + c_1 U_t + c_2 U_{t-1} + \varepsilon_t, \\ \varepsilon_t = \varepsilon_{t-24} + a_1(\varepsilon_{t-1} - \varepsilon_{t-25}) + a_2(\varepsilon_{t-2} - \varepsilon_{t-26}) + a_3(\varepsilon_{t-3} - \varepsilon_{t-27}) \\ \quad + (V_t - b_1 V_{t-1} - b_2 V_{t-2}) - \beta_1(V_{t-24} - b_1 V_{t-25} - b_2 V_{t-26}), \end{cases}$$

in which the estimates at stage $T = 4380$ are approximately given by

$$\hat{c}_0 = 7.9871, \quad \hat{c}_1 = 0.0166, \quad \hat{c}_2 = -0.0420, \quad \hat{a}_1 = 0.4776, \quad \hat{a}_2 = 0.9030,$$

$$\hat{a}_3 = -0.4305, \quad \hat{b}_1 = 0.0801, \quad \hat{b}_2 = -0.8524, \quad \hat{\beta}_1 = -0.8125, \quad \hat{\sigma}^2 = 0.0522,$$

and the t -statistics of the time series coefficients, justifying their significance, by

$$t_{\hat{a}_1} = 12.58, \quad t_{\hat{a}_2} = 26.13, \quad t_{\hat{a}_3} = 15.53, \quad t_{\hat{b}_1} = 2.50, \quad t_{\hat{b}_2} = 28.17, \quad t_{\hat{\beta}_1} = 75.42.$$

Moreover, one can easily check that the estimation of the autoregressive polynomial $\hat{\mathcal{A}}(z) = 1 - \hat{a}_1 z - \hat{a}_2 z^2 - \hat{a}_3 z^3$ is causal, for all $z \in \mathbb{C}$. The fitted values (\hat{C}_t) are obtained *via* (2.1), that is, for all $28 \leq t \leq T$,

$$(3.3) \quad \hat{C}_t = e^{\hat{Y}_t} - e^\mu$$

with $\mu = 5$. On Figures 3.5 and 3.6, the fitted values (\hat{Y}_t) and (\hat{C}_t) from model (3.2) and (3.3) are represented over the logarithmic load curve (Y_t) and the real load curve (C_t) , respectively, with a zoom. The temperature (U_t) during the same period is represented on Figure 3.7. One can see that, except for some unpredictable local behaviors related to the individual nature of the curve, there is a pretty good adequation between modeled and real values.

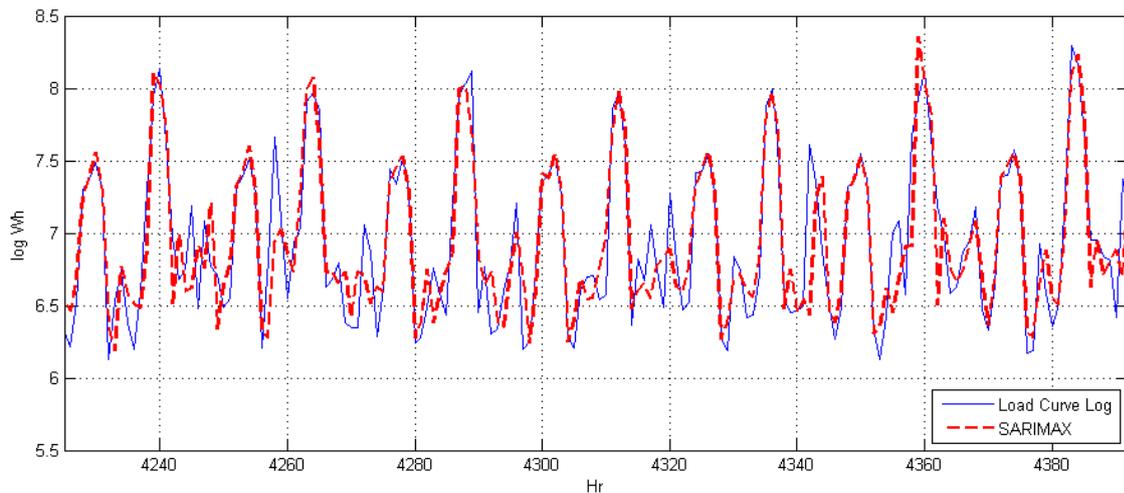


FIGURE 3.5. $\text{SARIMAX}(3, 0, 2, 2) \times (0, 1, 1)_{24}$ modelling on the logarithmic curve in red over the observed values in blue.

Forecasting. Our goal is now to propose an intraday forecasting methodology for the electrical consumption of individual customers. Let us start by introducing two

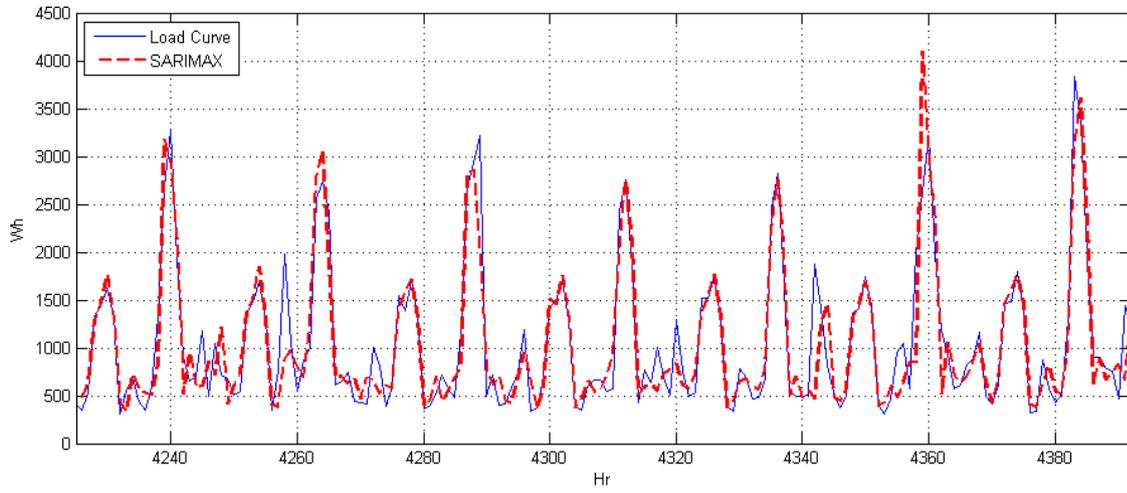


FIGURE 3.6. SARIMAX(3, 0, 2, 2) \times (0, 1, 1) $_{24}$ modelling on the real curve in red over the observed values in blue.

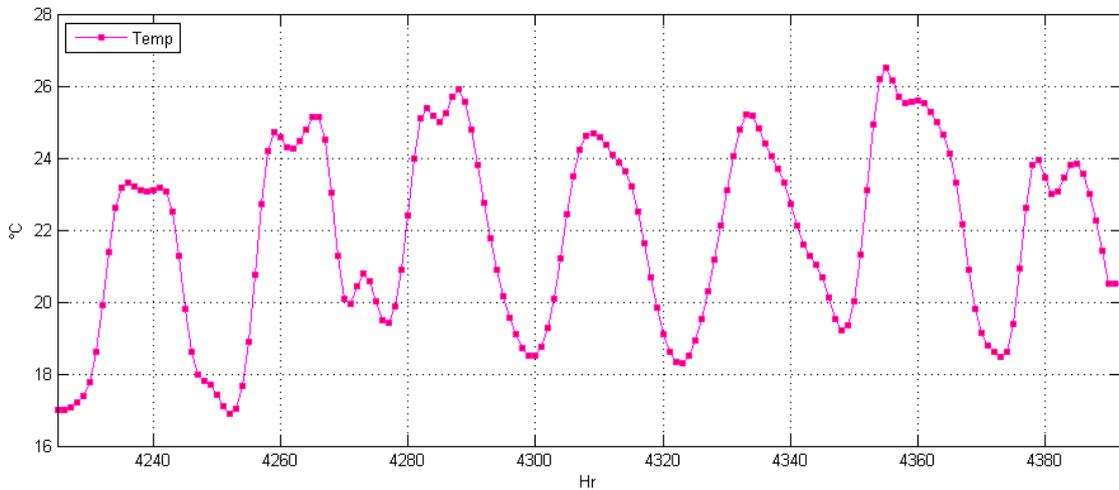


FIGURE 3.7. Temperature measured on the same period by the nearest weather station.

criteria that will help us to select the most suitable forecasting model. Denote by $\tilde{C}_{T+1}, \dots, \tilde{C}_{T+NH}$ the values of N consecutive predictions at horizon H from time T . Then, the absolute criterion C_A and the relative criterion C_R are defined as follows,

$$C_A = \frac{1}{NH} \sum_{k=1}^{NH} \left| \tilde{C}_{T+k} - C_{T+k} \right| \quad \text{and} \quad C_R = \left(\sum_{k=1}^{NH} C_{T+k} \right)^{-1} \sum_{k=1}^{NH} \left| \tilde{C}_{T+k} - C_{T+k} \right|.$$

Following the same lines as in the identification step, one has to make the parameters vary in their neighborhood to determine the most powerful forecasting model, considering the SARIMAX(3, 0, 2, 2) \times (0, 1, 1) $_{24}$ modelling as a basis. A Kalman filtering finite-history prediction method [13], [19], [21] is used to produce (\tilde{Y}_{T+k}) from the modelling for all $1 \leq k \leq NH$, also provided by the software environment,

and the forecasts (\tilde{C}_{T+k}) are obtained by

$$(3.4) \quad \tilde{C}_{T+k} = e^{\tilde{Y}_{T+k}} - e^\mu$$

with $\mu = 5$.

Remark 3.1. It is important to note that $\tilde{C}_{T+1}, \dots, \tilde{C}_{T+H}$ is not a sequence of predictions at horizon H but a sequence in which only the last component is a prediction at horizon H . By misuse of language, one shall consider in the sequel that a sequence of predictions at horizon H corresponds to H successive predictions without additional meantime information. By extension, a sequence of N predictions at horizon H needs N estimations of the parameters.

Remark 3.2. We only focused our attention on an individual load curve forecasting process and measured temperatures were considered as true values. However, the exogenous contributions cannot be known during the area $T+1 \leq t \leq T+H$ unless we also predict them, and temperatures forecasts usually come from well-known prediction models that we did not plan to investigate in this paper⁴. By virtue of the usual accuracy of these algorithms, the authors are pretty convinced that very similar results can be obtained using unobserved values of the exogenous variable in the forecasting experiments.

Our experiments are based on $N = 14$ days of daily forecasting, *i.e.* $H = 24$, the coefficients are evaluated on 3 months of data, that is $T = 2190$, and the numerical results are summarized in the Table 3.3 below. We have added a nonparametric approach NW, built on the Nadaraya-Watson kernel estimator [27], [32], that we have optimized for C_R on this curve, choosing the gaussian kernel and a window $h_T = T^{-0.45}$. We have also added a seasonal additive exponential smoothing approach SHW based on the Holt-Winters algorithm [17], where the parameters $\lambda_1 = 0.17$, $\lambda_2 = 0.01$, $\lambda_3 = 0.07$ and $s = 24$ have been optimized on a fine mesh for C_R .

	p	d	q	r	P	D	Q	s	C_A	C_R
SARIMAX	1	0	0	1	0	1	1	24	241.0	0.2279
SARIMAX	1	0	1	2	0	1	1	24	242.2	0.2290
SARIMAX	3	0	0	1	0	1	1	24	245.1	0.2318
SARIMAX	3	0	2	1	0	1	1	24	251.6	0.2380
SARIMAX	3	0	2	2	0	1	1	24	250.3	0.2368
SARIMAX	1	1	1	1	0	1	1	24	253.8	0.2400
SARIMAX	2	1	2	1	0	1	1	24	254.1	0.2403
NW	/	/	/	/	/	/	/	/	344.0	0.3253
SHW	/	/	/	/	/	/	/	/	243.9	0.2307

TABLE 3.3. Prediction criteria associated with a set of daily SARIMAX, NW and SHW forecasts on 3 months of consumption.

First, all experiments that we made lead to the conclusion that our model performs a lot better than every classic nonparametric approach. The seasonal Holt-Winters algorithm also gives worse results than the best SARIMAX modellings, even if they

⁴ Such models are usually provided by specialized laboratories in weather science.

are closer to the optimal solution. In addition, the parsimony in the time series analysis is a central issue in forecasting applications, and it is not surprising that the models minimizing C_A and C_R are not the same as those minimizing the bayesian criteria, and tend to reject uncertainty coming from overparametrization. Moreover, by selecting an optimal sliding window in the modelling, one is able to slightly improve our results. For example, the SARIMAX(1, 0, 0, 2) \times (0, 1, 1)₂₄ model provides $C_A = 231.0$ and $C_R = 0.2185$ in the particular case where $M = 1$ month, that is $T = 730$. On Figure 3.8, we investigate the influence of the size of the sliding window M together with the one of the exogenous regression dimension r on the relative criterion C_R for the latter modelling and the same experiment. This enables us to select the most powerful forecasting model for this particular curve.

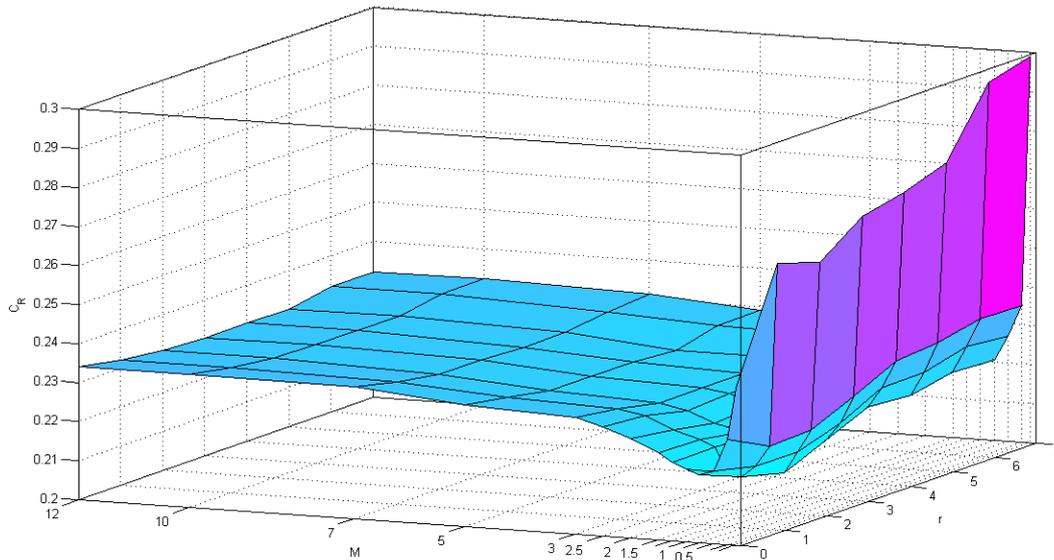


FIGURE 3.8. Influence of r and M on C_R calculated from 14 daily forecasts from the SARIMAX(1, 0, 0, r) \times (0, 1, 1)₂₄ modelling.

Accordingly, one shall consider the SARIMAX(1, 0, 0, 2) \times (0, 1, 1)₂₄ modelling with $M = 0.75$ month, even if one can see that r is not playing a substantial role as soon as it is greater than 2 for a reason of strong correlation of the exogenous phenomenon, already mentioned above. One can also observe that prediction results can be improved when the parameters are evaluated on rather small amounts of data. This can once again be explained by the nature of the curve and the underlying human behavior whose consumption is highly influenced by local circumstances such as weather, holiday period, *etc.* Whereas too few data are not sufficient to take into account the seasonality of period 24 and properly estimate the very significant β_1 parameter, conversely, results tend to stabilize when M increases disproportionately. The explicit expression of the predictive model is as follows, for all $26 \leq t \leq T = 548$,

$$(3.5) \quad \begin{cases} Y_t = c_0 + c_1 U_t + c_2 U_{t-1} + \varepsilon_t, \\ \varepsilon_t = \varepsilon_{t-24} + a_1(\varepsilon_{t-1} - \varepsilon_{t-25}) + V_t - \beta_1 V_{t-24}, \end{cases}$$

in which the estimates at stage $T = 548$ are approximately given by

$\hat{c}_0 = 7.2494$, $\hat{c}_1 = 0.0497$, $\hat{c}_2 = -0.0629$, $\hat{a}_1 = 0.3540$, $\hat{\beta}_1 = -0.7086$, $\hat{\sigma}^2 = 0.0708$, and the t -statistics of the time series coefficients, justifying their significance, by

$$t_{\hat{a}_1} = 8.66, \quad t_{\hat{\beta}_1} = 21.02.$$

The estimation of the autoregressive polynomial $\hat{A}(z) = 1 - \hat{a}_1 z$ is actually causal, for all $z \in \mathbb{C}$. On Figures 3.9 and 3.10, we display an example of 7 daily predictions from the latter model and a sliding window of 0.75 month of consumption, for the logarithmic curve (\tilde{Y}_t) as well as for the load curve (\tilde{C}_t). It also contains the 95% and 90% prediction confidence intervals, rather large owing to the horizon of prediction. The width of these intervals is $\pm q_{1-\alpha/2} \hat{\sigma}_h$ where $\hat{\sigma}_h$ is the estimated standard error at horizon h and $q_{1-\alpha/2}$ stands for the $(1 - \alpha/2)$ -th quantile of the Gaussian distribution. The hypothesis of residual normality is entirely reasonable in this context, in view of sample sizes and statistical procedures.

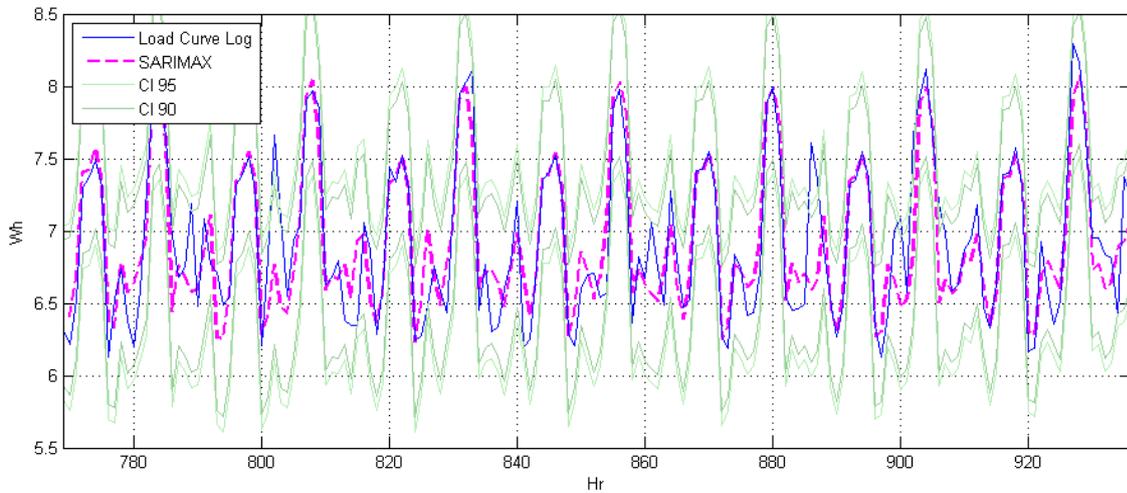


FIGURE 3.9. SARIMAX(1, 0, 0, 2) \times (0, 1, 1)₂₄ daily predictions on the logarithmic curve in magenta over the observed values in blue.

To conclude the empirical study, let us add that $\tilde{Y}_{T+1}, \dots, \tilde{Y}_{T+NH}$ only differs of 3% from Y_{T+1}, \dots, Y_{T+NH} when we consider the whole 14 daily predictions, that is $NH = 336$. Once again, one can notice that the coupled dynamic model provides very interesting results of prediction as soon as it is correctly specified, in spite of the noise on the load curve due to its individual nature.

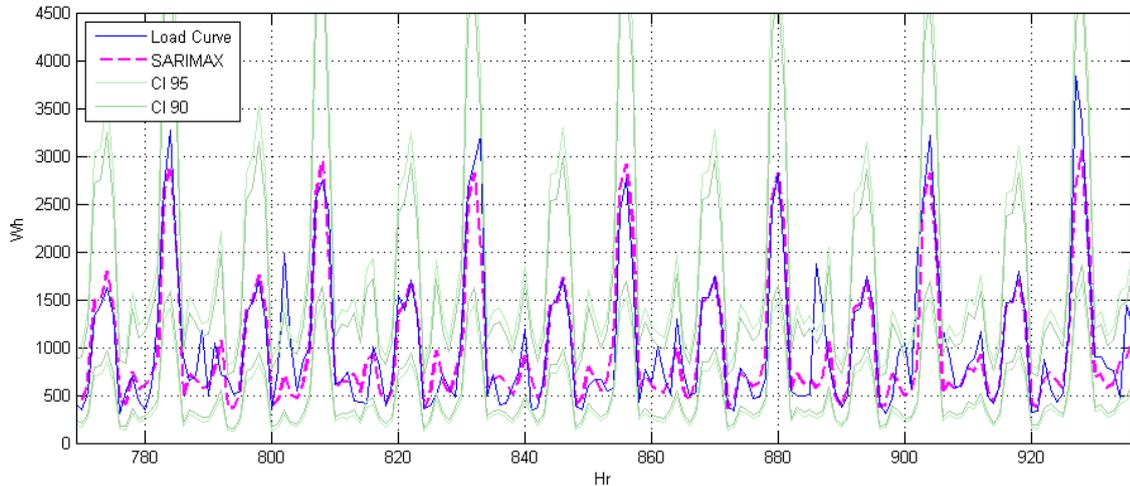


FIGURE 3.10. SARIMAX(1,0,0,2) \times (0,1,1)₂₄ daily predictions on the real curve in magenta over the observed values in blue.

4. CONCLUSION

To conclude, we would like to draw the significance of the exogenous covariates to the reader's attention. Indeed, let us notice that in some cases, the empirical study suggests to select $r = 0$, meaning no temperature influence despite the manifest linear relation between the latter and the consumption. The authors interpret this observation by the fact that seasonality and local circumstances totally prevail on the effect of temperature and that all information has already been recovered by the deep study of the signal, resulting in very few significant coefficients and equivalent forecasts for $r = 0, 1, \dots$. In addition, one should not overlook the possible irrelevance of the temperature measured by the weather station related to a customer without other criterion than geolocation, especially when altitude is concerned, coastal residence, cloud covering, or more generally when substantial differences may be observed at the same time between the weather station and the customer's home. Also, we should not forget that the exogenous inputs assumed to be known during the prediction period of the time series are nothing else than predictions themselves, with all attendant uncertainty. In addition to the required seasonality, the relevance of the exogenous measures is a central issue for this approach to be applied to a load curve with profit. Nevertheless, and despite many irregularities due to the individual nature of the curves, this study shows that some very interesting results of daily forecasts may be obtained under certain conditions already described, and above all a careful study of each curve. Finally, this intraday forecasting approach has been conducted on a whole set of individual customers from EDF, leading to the same satisfying conclusions. For large datasets, there is a major concern that it is impractical to automatically select the orders of the model for each customer in an

optimal manner, and the strategy used was to define a standard parameterization on a subset of heterogeneous curves, and then to make parameters vary in their neighborhood to optimize C_R .

Acknowledgments. *The authors thank the associate editor and the anonymous reviewer for their constructive comments which helped to improve the paper.*

REFERENCES

- [1] ANTONIADIS, A., PAPANODITIS, E., AND SAPATINAS, T. A functional wavelet-kernel approach for time series prediction. *J. Roy. Statistical Society B.* 68 (2006), 837–857.
- [2] ANTONIADIS, A., PAPANODITIS, E., AND SAPATINAS, T. Bandwidth selection for functional time series prediction. *Stat. Probabil. Lett.* 79-6 (2009), 733–740.
- [3] BOX, G. E. P., JENKINS, G. M., AND REINSEL, G. C. *Time Series Analysis, Forecasting and Control*. Third Edition. Holden-Day, Series G, 1976.
- [4] BOX, G. E. P., AND PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Am. Stat. Assn. Jour.* 65 (1970), 1509–1526.
- [5] BROCKWELL, P. J., AND DAVIS, R. A. *Time Series: Theory and Methods*. Springer-Verlag, New-York, 1991.
- [6] BROCKWELL, P. J., AND DAVIS, R. A. *Introduction to Time Series and Forecasting*. Springer, New-York, 1996.
- [7] CHAKHCHOUKH, Y., PANCIATICI, P., AND BONDON, P. Robust estimation of SARIMA models: Application to short-term load forecasting. *In IEEE workshop on Statistical Signal Processing. Cardiff, UK* (2009).
- [8] CHAKHCHOUKH, Y., PANCIATICI, P., AND MILI, L. New robust method applied to short-term load forecasting. *In IEEE Power Tech Conference, PowerTech. Bucharest, Romania* (2009).
- [9] CORNILLON, P. A., HENGARTNER, N., LEFIEUX, V., AND MATZNER-LØBER, E. Prévission de la consommation d’électricité par correction itérative du biais. *SFdS, Bordeaux* (2009).
- [10] COTTET, R., AND SMITH, M. Bayesian modeling and forecasting of intraday electricity load. *J. Amer. Statistical Assoc.* 98 (2003), 839–849.
- [11] DEVAINE, M., GOUDE, Y., AND STOLTZ, G. *Technical report, Ecole Normale Supérieure Paris and EDF R&D, Clamart* (2009).
- [12] DORDONNAT, V., KOOPMAN, S. J., OOMS, M., DESSERTAINE, A., AND COLLET, J. An hourly periodic state space model for modelling french national electricity load. *Int. J. Forecasting.* 24 (2008), 566–587.
- [13] DURBIN, J., AND KOOPMAN, S. *Time Series Analysis by State Space Methods*. Oxford University Press, 2001.
- [14] ESPINOZA, M., JOYE, C., BELMANS, R., AND DE MOOR, B. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. *IEEE Trans. Power Syst.* 20-3 (2005), 1622–1630.
- [15] GARCIA, R., CONTRERAS, J., VAN AKKEREN, M., AND GARCIA, J. B. C. A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Trans. Power Syst.* 20 (2005), 867–874.
- [16] GARDNER, G., HARVEY, A. C., AND PHILLIPS, G. D. A. Algorithm AS154. An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Appl. Stat.* 29 (1980), 311–322.
- [17] GELPER, S., FRIED, R., AND CROUX, C. Robust forecasting with exponential and Holt-Winters smoothing. *J. Forecast.* 29-3 (2010), 285–300.
- [18] GHOFRANI, M., HASSANZADEH, M., ETEZADI-AMOLI, M., AND FADALI, M. S. Smart meter based short-term load forecasting for residential customers. *North American Power Symposium (NAPS)* (2011), 1–5.

- [19] HARVEY, A. C. *Time Series Models*. Second Edition. Harvester Wheatsheaf, 1993.
- [20] HARVEY, A. C., AND KOOPMAN, S. J. Forecasting hourly electricity demand using time-varying splines. *J. Amer. Statistical Assoc.* 88 (1993), 1228–1236.
- [21] HARVEY, A. C., AND MCKENZIE, C. R. Algorithm AS182. An algorithm for finite sample prediction from ARIMA processes. *Appl. Stat.* 31 (1982), 180–187.
- [22] KOOPMAN, S. J., OOMS, M., AND CARNERO, M. A. Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices. *J. Amer. Statistical Assoc.* 102 (2007), 16–27.
- [23] KWIATKOWSKI, D., PHILLIPS, P. C. B., SCHMIDT, P., AND SHIN, Y. Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? *J Econometrics.* 54 (1992), 159–178.
- [24] LIU, J. M., CHEN, R., LIU, L. M., AND HARRIS, J. L. A semi-parametric time series approach in modeling hourly electricity loads. *J. Forecasting.* 25 (2006), 537–559.
- [25] LJUNG, G. M., AND BOX, G. E. P. On a measure of a lack of fit in time series models. *Biometrika.* 65-2 (1978), 297–303.
- [26] MARTIN, M. M. Filtrage de Kalman d’une série temporelle saisonnière. Application à la prévision de consommation d’électricité. *Rev. Statist. Appl.* 4 (1999), 69–86.
- [27] NADARAYA, E. A. On a regression estimate. *Teor. Veroyatnost. i Primenen* 9 (1964), 157–159.
- [28] POGGI, J. M. Prévision non paramétrique de la consommation d’électricité. *Rev. Statist. Appl.* 42 (1994), 83–98.
- [29] SAID, E. S., AND DICKEY, D. A. Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika.* 71-3 (1984), 599–607.
- [30] SMITH, M. Modeling and short-term forecasting of New South Wales electricity system load. *J. Bus. Econ. Statist.* 18 (2000), 465–478.
- [31] TAYLOR, J. W. Triple seasonal methods for short-term electricity demand forecasting. *Eur. J. Oper. Res.* 204 (2010), 139–152.
- [32] WATSON, G. Smooth regression analysis. *Sankhya Ser. A* 26 (1964), 359–372.

E-mail address: Sophie.Bercu@edf.fr

E-mail address: Frederic.Proia@inria.fr

EDF RECHERCHE & DÉVELOPPEMENT, DÉPARTEMENT ICAME. 1, AVENUE DU GÉNÉRAL DE GAULLE, 92141 CLAMART CEDEX.

UNIVERSITÉ BORDEAUX 1, INSTITUT DE MATHÉMATIQUES DE BORDEAUX, UMR 5251, AND INRIA BORDEAUX, TEAM ALEA, 200 AVENUE DE LA VIEILLE TOUR, 33405 TALENCE CEDEX, FRANCE.