TÉCNICO
LISBOA

Prediction of Electrical Energy Consumption at IST

Sofia Fidalgo da Silva

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engeneering

Supervisors: Prof. Dr. João Manuel Lage de Miranda Lemos
Prof. Dr. Mário Serafim dos Santos Nunes

Examinatiom Committee

Chairperson: Prof. João Fernando Cardoso Silva Sequeira
Supervisor: Prof. Dr. João Manuel Lage de Miranda Lemos
Member of the Committee: Prof. Dr. João Pedro Castilho Pereira Santos Gomes

October 2016

Prediction is very difficult, especially if it’s about the future.

Niels Bohr

Acknowledgments

First of all, I would like to thank Professor João Miranda Lemos, whose lectures made me develop an interest in the forecasting world, for the support, and availability shown during the course of this work.

Special thanks to my boyfriend (Tiago) for always being there for me for the last amazing 5 years, being my best friend and my work mate.

I would also like to thank my family for always being supportive in all these years, despite the distance, and also Tiago's family, for being there for me when I needed.

Finally, I would like to thank my friends, for, without them, this journey would have been more difficult, and much less fun.

Abstract

The relevance of electrical energy into the world economy is a motivation for the sustainable energy consumption. Power system management assumes a major role in this field, by contributing to efficient energy consumption. The development and application of forecasting methods to predict the energy consumption contributes to its optimization and to keep a balance between production and demand.

The purpose of this dissertation is to make predictions for the electric energy consumption at the Alameda campus of Instituto Superior Técnico, in Lisbon. The prediction problem consists on computing the estimates of the future terms of a time series, given a sequence of observations. It is assumed that the consumption process is generated by an ARMA model.

To solve the prediction problem, some algorithms were implemented, in order to integrate all the necessary steps, which are data processing, model identification, and prediction. Several architectures are taken into consideration in this study, one based on multiple models, one based on adaptive methods, and other two that combine both multiple models and adaptive approaches.

This dissertation focuses mostly on traditional approaches for model identification, such as the prediction prror method and a recursive and extended version of the least squares method. In addition, several studies on the variance of the prediction error are also shown, including on how it is influenced by the prediction horizon. Finally, some comparisons regarding the performance of the different implemented integration architectures are presented.

Keywords

Prediction, Time Series, ARMA, Multiple Models, Adaptive, Prediction Error

Resumo

O impacto da energia eléctrica na economia mundial é uma motivação para o consumo sustentável de energia. Neste aspecto, a gestão dos sistemas de energia assume um papel fundamental, ao contribuir para o consumo eficiente de energia eléctrica. O desenvolvimento e aplicação de métodos de previsão para prever o consumo de energia contribui para a sua optimização e para manter um balanço entre procura e oferta.

Pretende-se com esta dissertação fazer previsões do consumo de energia eléctrica do campus da Alameda do Instituto Superior Técnico, em Lisboa. O problema da predição consiste em obter as estimativas dos termos futuros de uma série temporal, dada uma série de observações relativas ao consumo de energia. É assumido que esta série temporal é gerada por um modelo ARMA.

De modo a resolver este problema, foram implementados vários algoritmos, por forma a integrar todos os passos necessários, que incluem o processamento de dados, a identificação do modelo e a previsão. Diversas arquitecturas sÃ$o consideradas no estudo, uma baseada em múltiplos modelos, outra baseada em métodos adaptativos, e as últimas duas que resultam na combinação das duas abordagens.

Esta dissertação foca-se sobretudo em métodos tradicionais para a identificação de modelos, como o *prediction error method* e uma versão recursiva e estendida do método dos mínimos quadrados. São também apresentados estudos relativamente à variância do erro de predição, incluindo mostrar de que maneira é influenciada pelo horizonte de predição. Sã também mostradas comparações relativamente ao desempenho das diferentes arquiteturas de integração que foram implementadas.

Palavras Chave

Predição, Série Temporal, ARMA, Mútiplos Modelos, Adaptativa, Erro de Predição

Conteúdo

Lista de Figuras

Lista de Tabelas

Abbreviations

AR Autoregressive

ARIMA Autoregressive Integrated Moving Average

ARMA Autoregressive Moving Average

ARMAX Autoregressive Moving Average with exogenous inputs

HVAC Heating, Ventilation and Air Conditioning

MA Moving Average

R-ELS Recursive Extended Least Squares

R-LS Recursive Least Squares

1

Introduction

Contents

1.1 Motivation

Electrical energy is a crucial need for society development, supplying our houses, industries and the infrastructures our lives rely on [1]. Taking into account the relevance of electrical energy in our lives and on the world economy, the future of society relies on the sustainable energy consumption, in the sense that the amount of energy produced should meet the current needs without compromising future necessities [1]. With the increasing demand of electrical energy, power system management assumes a major role by contributing to the efficient energy consumption.

Power system management is an important issue that intersects this dissertation in several ways. First of all, in a power grid that distributes energy at the national level, there must be, at all times, a balance between production and demand, so that there is enough energy produced to match the consumption needs. Failure to comply with this balance may result in serious drawbacks, such as the degradation of service quality associated to undesired changes in electrical tension and frequency, or even instability phenomena that ultimately may lead to a partial or complete shut-down of the grid.

Since there is a delay on the activation of production, there is a need to perform a demand forecast, a task that may be accomplished with the methods described in this dissertation.

Another motivation for the methods addressed here is the forecast of energy demand in smaller grids, such as a campus grid like the one of Instituto Superior Técnico (IST). In this case, the motivation is mainly economic, in the sense that having a good prediction of what the electrical consumption will be in an hour, day or a week in the future, can improve the management of the power system, by optimizing the energy consumption.

The purpose of this dissertation is to develop MATLAB code for the prediction of the electrical energy consumption at the IST campus applying temporal series methods, using real data, which is provided by the *Campus Sustentável* project.

1.2 State of The Art

Prediction, or forecasting, has been widely used in several areas, such as, with great relevance and representativity, forecasts of electrical power consumption and generation [2] [3] [4] [5] [6], heat loads [7], water demand [8], wind generation [9] [10] [11] [12], and financial time series [13] [14]. Since prediction is still a growing field of study, new applications of this area are being developed, with some examples being, forecasting the time of a disaster [15] in a certain place, disturbance forecasting [4] in order to improve the control of a given variable and diminish the effects of the disturbance, detecting abnormal situations [3], and improvement of instrument precision [16].

When it comes to models to be utilized when predicting, several methods have been addressed [17], starting with traditional models, such as regression models [13], single or multivariate, that express a variable in terms of the variables to which it is related, and exponential smoothing [4], whose purpose is to eliminate the noisy component by weighted averaging the previous samples. Moving to modified traditional methods, that include the Autoregressive Moving Average (ARMA) model family [2] [3] [5] [11] [14] [16] [18], which will be the base model for this work, and Support Vector Machines

(SVM) [12], that are useful for classification purposes. Lastly, there are some more recent methods that fall mostly in the area of Artificial Inteligence (AI), some examples being Neural Networks [3] [8] [14] [16], which is inspired by the neural networks of living things, fuzzy logic [6] [15] [16], that implements human like knowledge to uncertain situation, and Grey Theory [8] [16]. These last two approaches perceive situations as being neither black, nor white, but grey.

Some of these models have been combined in order to generate better forecasts. Ideas like combining linear and non-linear models, such as Autoregressive Integrated Moving Average (ARIMA) and neural networks [14] have been implemented, and even utilizing both off-line and on-line learning [10] [15] , the former for the predictable behavior, and the latter for unexpected behavior.

As for the parameter estimation of the model, for the most classic approaches, the Least Squares Method [14], and Maximum Likelihood [3] [9] [11] [18] are the most used, for the AI approaches, machine learning methods are implemented, such as backpropagation [3] [14] [16] for neural networks.

Lastly, some ideas on data treatment among the revised literature include, for time series analysis, segmentation [2], which means separating weekdays, weekends, holidays and vacations for separate modeling of each, and also the use of the Discrete Wavelet Transform [5] [9], that allows to decompose a non-sationary signal in various components, that when are subtracted from the original signal originate a stationary process, suitable for the identification of models of the ARMA family. Some models require input variables, in which can be included information about the hour, day, or month that is being considered, and in such cases there is no need for data treatment, except for outliers.

1.3 Problem Formulation

The prediction, or forecasting, problem consists of the following:
Given a sequence of observations of a time series (*i.e.*, a sequence of numbers associated to time), compute an estimate of the terms associated to future times.This problem is solved on the basis of a mathematical model that is assumed to generate the time series terms.

More precisely, if $\ldots,\ k-2,\ k-1,\ k,\ k+1,\ \ldots$ are integers that denote discrete time, with k being the present time, the problem consists of computing an estimate of $k+m$, where m is an integer designated *prediction horizon*. This estimate is made from a finite number of previous observations from $k-n,\ k-n+1,\ \ldots,\ k$ and, in some cases, from previous estimates.

In the above context, the problem to be addressed in this thesis consists of the development of algorithms to predict the electrical consumption at IST. Regarding the prediction problem, one must consider the following steps:

1.3.1 Criterion

Establish a criterion to evaluate the predictions. A quadratic criteria is used in this dissertation.

1.3.2 Segmentation

Separate week days, weekends, holidays and vacation. It is also necessary to consider the Heating, Ventilation and Air Conditioning (HVAC) part of the electric energy consumption process separately from the rest, also segmented in the same way mentioned above.

1.3.3 Remove seasonality

After the segmentation what is left are signals with quasi-periodic behavior. That periodicity must be filtered using a filter with zeros placed on the unitary circle, in order to cancel with the poles that cause the periodicity.

1.3.4 Learn the model

After filtering, a stationary signal is obtained. The Box-Jenkins [19] method, which uses ARMA models that will be explained later, shall be used to learn the model that generates that stationary signal, or process. The Box-Jenkins method assumes that the stationary signal is the result of white noise going through a linear filter, which is a quotient of two polynomials, C (numerator) and A (denominator).

Polynomials A and C must be estimated using the prediction-error method as implemented by the MATLAB function, *armax*, and also a recursive least squares method algorithm.

1.3.5 Prediction

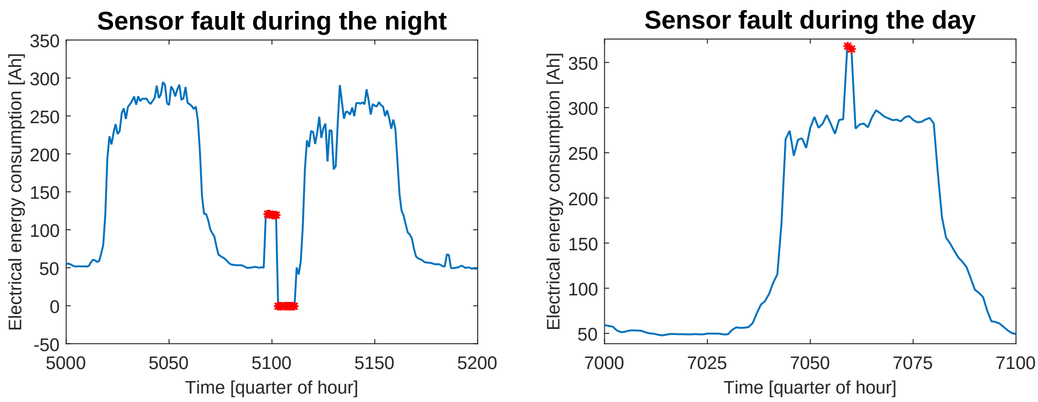
Having the two polynomials, to predict future values, only a division between the two is needed. When predicting values the errors in prediction must be taken into account, particularly, the bigger the prediction interval, the bigger the prediction error, so there needs to be a compromise regarding this matter.

The prediction is done by optimizing a criterium, that in this case will be a quadratic criterium given by the mean-square prediction error.

1.4 Available Data

The acquired data files, in excel, are from the North Tower, one of the buildings of the Alameda Campus of IST, and they correspond to the electric energy consumption measurements in that building. There is one measurement spot in each building that gathers information of the corresponding sensors, which are 26 in total. The data acquired is from August 2015 to March 2016. Also, the data provided is given in Ah.

The consumption is a combination of HVAC and other normal electric energy consumption (illumination, electric equipment,...).The HVAC has a specific working pattern characterized by approximately constant sections depending on the necessity at the moment. Measurement spots take measurements every minute, but they are registered only every 15 minutes, accumulating the previous 15 measurements, which is the interval that is going to be considered when doing predictions.



(a) Night time sensor fault

(b) Day time sensor fault

Figure 1.1: Example of sensor faults

When there is a sensor fault, the missing value is replaced by the measurement value at the same week day and quarter of an hour of the previous week, but only if it is coherent with the current week measurements (if it does not differ by more than 5%). If that is not possible, the corresponding value can either be zero or a high peak, meaning that this value is set to the maximum value that that sensor can capture. There are four possible causes for these faults:

- local electric power fault
- computer network fault
- monitoring equipment anomaly
- human error when operating the measurement equipment.

Figure 1.1 shows probable sensor faults during the night, 1.1a, and during the day, 1.1b. There is a high probability that these measurements do not correspond to the real values of the electric energy used during the corresponding time interval, and that is taken into account during the course of this work.

Besides this, there is a constant consumption (hereafter referred as the ”base line consumption”) in the buildings that corresponds to a fraction of the maximum consumption (from one third to one half).

Figures 1.2 to 1.4 show some graphics that represent the data throughout several months, a week, and a day.

In Figure 1.2 one can visualize the ”base line consumption”, that corresponds to the permanent electric consumption throughout a year and a half, as well as peaks and valleys, associated to week days and weekends, and holidays, respectively, and also a drop in the consumption, corresponding to the summer vacation.

Figure 1.3 shows also peaks and valleys that correspond to day and night time, and a considerable drop in consumption relative to the weekend. High peaks can be visualized at the beginning of every day (around 6 a.m.).

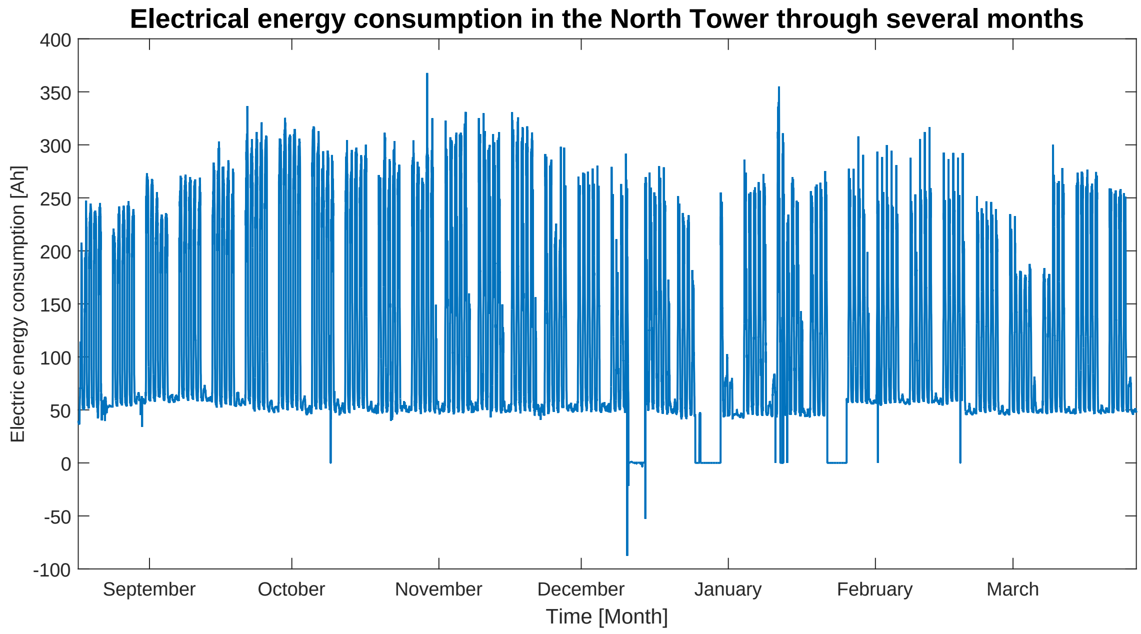


Figure 1.2: Electric energy consumption in the North Tower throughout several months

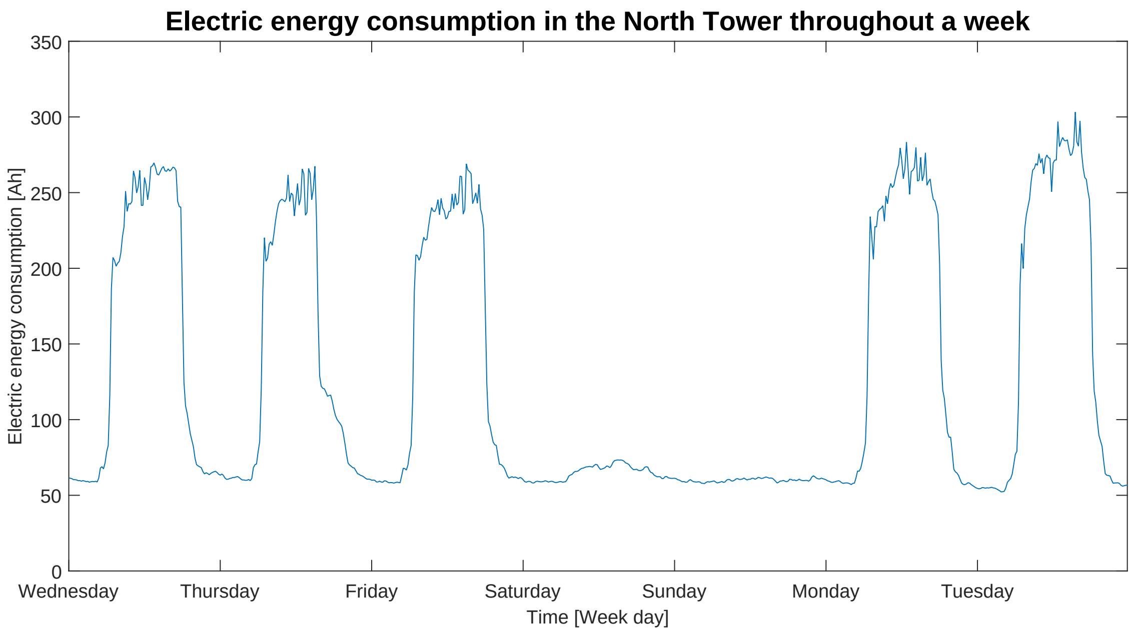


Figure 1.3: Electric energy consumption in the North Tower throughout a week

As for Figure 1.4, the influence of the HVAC is noticeable, given the sudden decreases and increases in the consumption, which correspond to the HVAC being turned off and turned on, respectively.

Figures 1.5 and 1.6 represent the normal and HVAC consumption, respectively, throughout the week depicted in Figure 1.3.

It can be seen that the "base line consumption"is part of the normal consumption, as the HVAC consumption is $0\ Ah$ during the nights and weekends (with a few exceptions). It is also noticeable that the HVAC consumption varies more abruptly than the normal consumption, particularly at the beginning of the school days, where the consumption level goes from nearly $0\ Ah$ to about $100\ Ah$ in

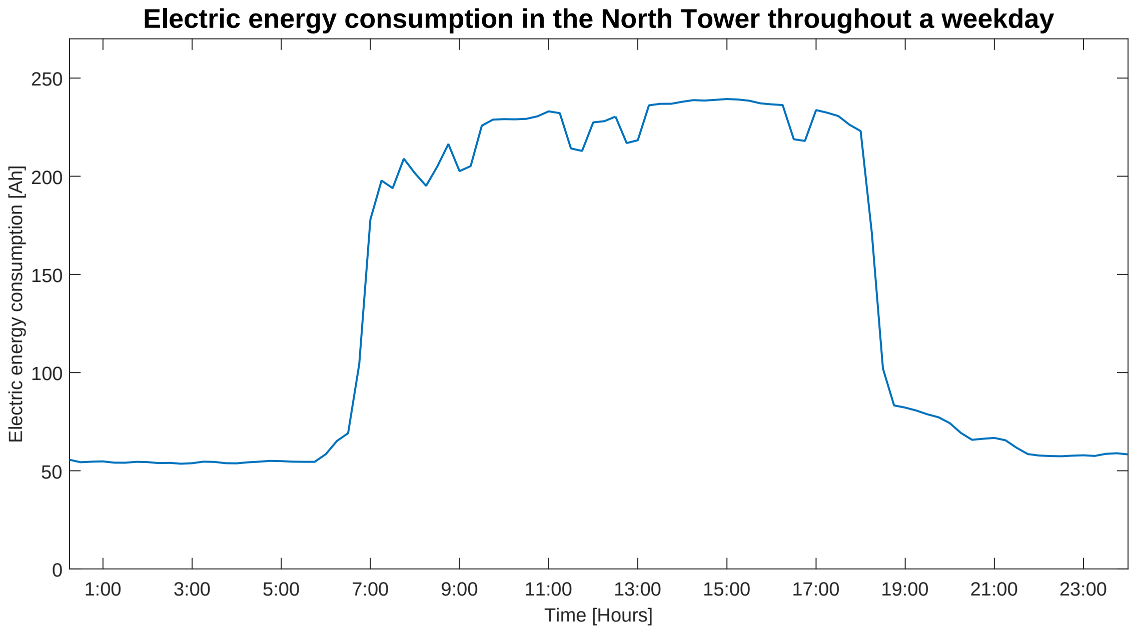


Figure 1.4: Electric energy consumption in the North Tower throughout a day

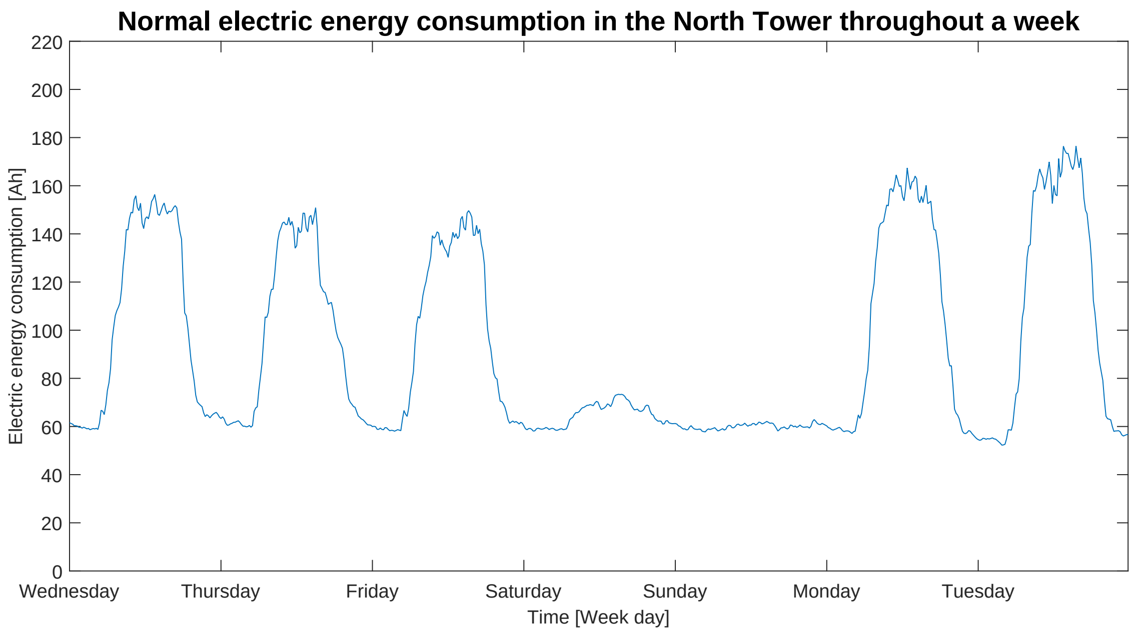


Figure 1.5: Normal electric energy consumption in the North Tower throughout a week

15 minutes, and vice-versa at the end of the school days.

1.5 Original Contributions

In this dissertation, the ARMA family models are used to describe the systems that generate the electric consumption processes. These models, that are to be used to make the predictions of the electrical energy consumption at IST, are identified using traditional approaches, namely, a prediction error method and the least squares method.

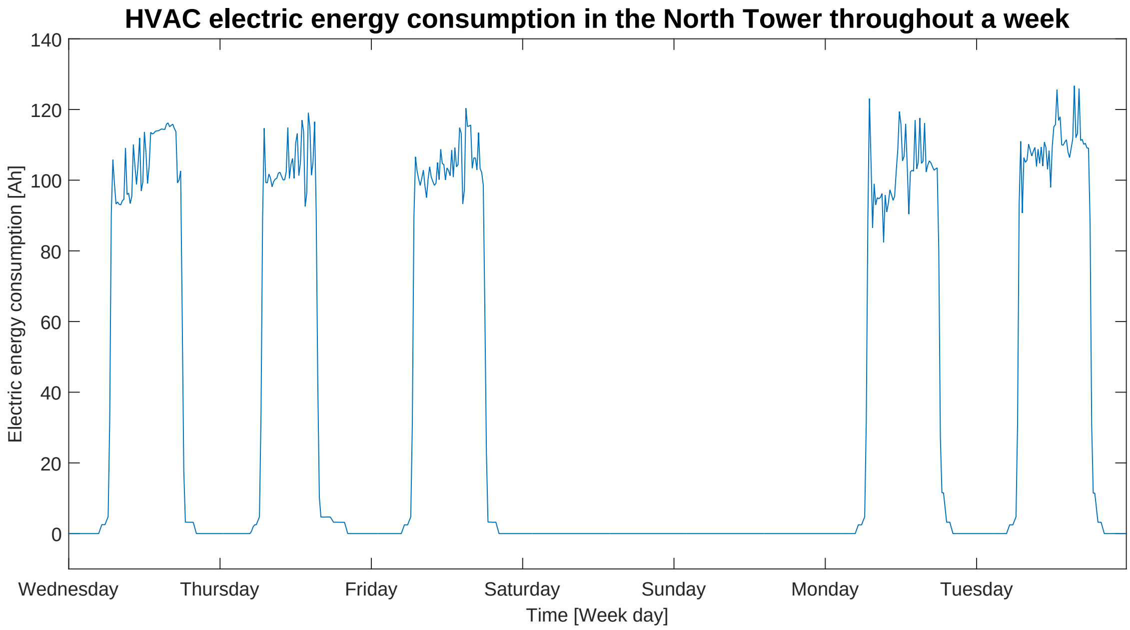


Figure 1.6: HVAC electric energy consumption in the North Tower throughout a week

Two adaptive strategies are also considered to estimate model parameters, one based on the recursive extended least squares method with exponential forgetting, and another that consists of slightly varying a given set of parameters. Several prediction architectures are presented, which include a multiple models approach, an adaptive approach, and some combinations of the previous two.

1.6 Thesis Outline

This document is organized as follows. Chapter 2 provides some background knowledge necessary to understand some concepts and approaches present in this dissertation. In Chapter 3 the methods used to solve the proposed prediction problem are presented, which includes data processing, model identification, and prediction. These methods are then put together in several different architectures and presented in Chapter 4. Chapter 5 is dedicated to results obtained after applying the proposed methods and architectures, along with some comparisons between them. To finalize, some conclusions about the work performed and experimental results obtained, and considerations about possible future work are presented in Chapter 6.

2

Background Knowledge

Contents

The following sections provide some mathematical and nomenclature knowledge necessary to fully understand and implement a solution for the proposed problem.

2.1 Discrete-time white noise

Discrete-time white noise [20] is a stochastic process that consists of a sequence $\{x(k, w),\ k = \ldots, -1, 0, 1, \ldots\}$ of independent and identically distributed ($i.i.d.$) random variables. This means that each random variable has the same probability distribution as the others, and $x(k)$ and $x(l)$ are independent if $k \neq l$. This process can also be characterized by its covariance function,

$$r(\tau) = \begin{cases} \sigma^2 & \tau = 0 \\ 0 & \tau = \pm 1, \pm 2, \ldots \end{cases} \tag{2.1}$$

with σ^2 being the variance of the process, and, consequently, by it's spectral density,

$$\phi(\omega) = \frac{\sigma^2}{2\pi} \tag{2.2}$$

which is constant for all ω, that represents the angular frequency, and $\omega \in \mathbb{R}$.

Many stationary stochastic processes can be generated by making white noise the input of a linear system, making white noise an extremely important process in stochastic control theory and, particularly, for this thesis. Actually, all stochastic processes with a rational spectra can be expressed in this way (Spectral Factorization Theorem [20]). Although it has been shown that some natural phenomena can be better represented with factored models, this other approach is much more complicated to apply and, in this case, is not expected to yield a significant advantage.

2.2 Shift operators

Consider the following discrete linear, time-invariant system,

$$y(k+n) + a_1 y(k+n-1) + \ldots + a_n y(k) = c_0 e(k+m) + c_1 e(k+m-1) + \ldots + c_m e(k), \tag{2.3}$$

where $y(k)$ and $e(k)$ denote the output and input processes of the system, respectively.

In order to write equation (2.3) in a more compact way, let the following expression be considered

$$x(k+1) = qx(k),$$

where q is the forward shift operator, and $x(k)$ is a generic process. Using the above delay description in equation (2.3)

$$q^n y(k) + a_1 q^{n-1} y(k) + \ldots + a_n y(k) = c_0 q^m e(k) + c_1 q^{m-1} e(k) + \ldots + c_m e(k),$$

follows the next expression

$$y(k) = \frac{c_0 q^m + c_1 q^{m-1} + \ldots + c_{m-1} q + c_m}{q^n + a_1 q^{n-1} + \ldots + a_{n-1} q + a_n} e(k) = \frac{C(q)}{A(q)} e(k). \tag{2.4}$$

The same reasoning applies when considering the following expression,

$$x(k-1) = q^{-1} x(k)$$

where q^{-1} is the backward shift operator, and $x(k)$ is a generic process. Rewriting equation (2.3), delaying the processes by n,

$$y(k)+a_1y(k-1)+\ldots+a_ny(k-n)=c_0e(k+m-n)+c_1e(k+m-n-1)+\ldots+c_me(k-n), \quad (2.5)$$

and using the backward shift operator, one gets the following equation

$$y(k)=q^{-d}\frac{c_0+c_1q^{m-1}+\ldots+c_{m-1}q^{-m+1}+c_mq^{-m}}{1+a_1q^{-1}+\ldots+a_{n-1}q^{-n+1}+a_nq^{-n}}e(k)=q^{-d}\frac{C^*(q^{-1})}{A^*(q^{-1})}e(k), \quad (2.6)$$

having $d=n-m$, and $d\geq 0$ assuming that the system is causal. Let the definition of reciprocal polynomial be

$$P^*(q^{-1})=q^{-n}P(q),$$

where n is the order of $P(q)$, and $P(q)$ a generic polynomial. It can be said that $A^*(q^{-1})$ and $C^*(q^{-1})$ are the reciprocal polynomials of $A(q)$ and $C(q)$, respectively, and that the following expression is true

$$\frac{C(q)}{A(q)}=q^{-d}\frac{C^*(q^{-1})}{A^*(q^{-1})}. \quad (2.7)$$

2.3 The ARMA model

Given a stationary signal, $y(k)$, it can be seen as the output of a linear system to which is fed white noise, $e(k)$, as input [20]. Three main models can be identified. The Moving Average (MA) model, , defined by

$$y(k)=e(k)+c_1e(k-1)+\ldots+c_{n_c}e(k-n_c), \quad (2.8)$$

the Autoregressive (AR) model, described by

$$y(k)+a_1y(k-1)+\ldots+a_{n_a}y(k-n_a)=e(k), \quad (2.9)$$

and the ARMA model, defined as

$$y(k)+a_1y(k-1)+\ldots+a_{n_a}y(k-n_a)=e(k)+c_1e(k-1)+\ldots+c_{n_c}e(k-n_c), \quad (2.10)$$

which is a combination of the previous two.

Finding the model that generates the stationary process $y(k)$ requires the estimation of the a_i and c_i coefficients. Representing the ARMA model in terms of the Z-transform, with null initial conditions, results in

$$\mathcal{Z}(y(k)+a_1y(k-1)+\ldots+a_{n_a}y(k-n_a))=\mathcal{Z}(e(k)+c_1e(k-1)+\;+\ldots+c_{n_c}e(k-n_c))\Leftrightarrow$$

$$\Leftrightarrow Y(z)+a_1z^{-1}Y(z)+\ldots+a_{n_a}z^{-n_a}Y(z)=E(z)+c_1z^{-1}E(z)+\;+\ldots+c_{n_c}z^{-n_c}E(z)\Leftrightarrow$$

$$\Leftrightarrow\frac{Y(z)}{E(z)}=\frac{1+c_1z^{-1}+\ldots+c_{n_c}z^{-n_c}}{1+a_1z^{-1}+\ldots+a_{n_a}z^{-n_a}}=H(z),$$

where $H(z)$ is the transfer function of the linear system which, given the previous steps, yields

$$H(z)=\frac{C^*(z^{-1})}{A^*(z^{-1})}=\frac{1+c_1z^{-1}+\ldots+c_{n_c}z^{-n_c}}{1+a_1z^{-1}+\ldots+a_{n_a}z^{-n_a}} \quad (2.11)$$

where n_a and n_c denote the degrees of the polynomials $A^*(z^{-1})$ and $C^*(z^{-1})$, respectively This representation will be useful when presenting the estimates of the a_i and c_i coefficients.

Consider now a more general approach on models. The previous models can only generate stationary processes, whereas one may want to generate processes that are not stationary. Introducing the *autoregressive integrated moving average* - ARIMA - models which allow to have a vastest range of signal at the output, defined as

$$H(z) = \frac{1 + c_1 z^{-1} + \ldots + c_n z^{-n}}{1 + a_1 z^{-1} + \ldots + a_n z^{-n}} \times \frac{1}{1 - z^{-d}}. \tag{2.12}$$

Keep in mind that a multiplication in the Z-transform domain corresponds to a convolution in the time domain, so, equation (2.12), represents the transfer function of a cascade of two systems, the first corresponding to the ARMA model, which generates a stationary process, and the second corresponding to the integration of the stationary output process that is obtained. Analyzing the integration part, when $d = 1$, one has a pole in $z = 1$, which corresponds to a simple integrator, and when $d > 1$, there are d poles on the unit circle, and that corresponds to adding periodicity to the output process of the ARMA model.

2.4 Prediction-error method

The prediction-error method [21] consists of computing the estimates of the parameters that minimize the difference between the actual observations and their predictions.

Consider an arbitrary model

$$y(k) = f(\boldsymbol{\theta}) + e(k),$$

where f is a function that depends on the set of parameters $\boldsymbol{\theta}$, which are the quantities to be estimated.

To compute the estimates of the model parameters, the first step is to describe the model using a suitable predictor for the observations one step ahead, $\hat{y}(k|k-1)$,

$$\hat{y}(k|k-1) = p(O^{k-1}),$$

with p being a function that depends on past observations, O^{k-1}, then, the predictor needs to be parametrized so that it depends on the set of parameters that are to be estimated,

$$\hat{y}(k|k-1) = p(O^{k-1}, \boldsymbol{\theta}).$$

Finally, the estimates can be computed by minimizing the difference between the observed data and their predictions,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{M} l(y(k) - \hat{y}(k|k-1)) = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{M} l(y(k) - p(O^{k-1}, \boldsymbol{\theta})),$$

where M is the total number of observations and l is suitable norm function.

2.5 Least squares method

Consider the following model

$$y(k) = \boldsymbol{\theta}^T \boldsymbol{\varphi}(k) + e(k), \tag{2.13}$$

where M is the number of observations $y(k)$, $\boldsymbol{\theta}$ is a column vector of all the parameters to be estimated, the upper script T denotes that the matrix is transposed, and $\varphi(k)$ is a column vector of known values to with dimension equal to the number of parameters in $\boldsymbol{\theta}$.

Considering that there are M observations of the system output $y(k)$, one may define a quadratic cost function to minimize [20] as

$$J(\boldsymbol{\theta}) = \frac{1}{2M}\sum_{k=1}^{M}[e(k)]^2 = \frac{1}{2M}\sum_{k=1}^{M}[y(k) - \boldsymbol{\theta}^T \boldsymbol{\varphi}(k)]^2, \tag{2.14}$$

which means to find $\boldsymbol{\theta}$ such that the variance of the error is minimized.

Since there are M observations the model defined in (2.13) may be represented in matrix form as

$$\begin{bmatrix} y(1) \\ y(2) \\ \cdots \\ y(M) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varphi}^T(1) \\ \boldsymbol{\varphi}^T(2) \\ \cdots \\ \boldsymbol{\varphi}^T(M) \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} e(1) \\ e(2) \\ \cdots \\ e(M) \end{bmatrix},$$

which can be represented in a more compact way

$$\overline{\boldsymbol{y}} = \boldsymbol{\Phi}\boldsymbol{\theta} + \boldsymbol{\epsilon}. \tag{2.15}$$

The cost function can be rewritten using matrix notation

$$J(\boldsymbol{\theta}) = \frac{1}{2M}||\boldsymbol{\epsilon}||^2 = \frac{1}{2M}\boldsymbol{\epsilon}^T\boldsymbol{\epsilon},$$

and from equation (2.15), follows

$$J(\boldsymbol{\theta}) = \frac{1}{2M}(\overline{\boldsymbol{y}} - \boldsymbol{\Phi}\boldsymbol{\theta})^T(\overline{\boldsymbol{y}} - \boldsymbol{\Phi}\boldsymbol{\theta}).$$

The estimates $\boldsymbol{\theta}$ can be obtained by solving the system of linear equations given by

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0.$$

Solving with respect to $\hat{\boldsymbol{\theta}}$, the equation that allows the estimation of the parameters is given by

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T[\overline{\boldsymbol{y}}]. \tag{2.16}$$

3

Methods

Contents

This chapter presents the proposed methods to address all the parts required to solve the problem as discussed in section 1.3.

3.1 Data Processing

In terms of data processing there is the need to deal with outliers, and seasonality. To do so, the following methods are proposed.

3.1.1 Outlier removal

The proposed approach to remove the outliers is to compare the difference between the values at a current time instant, k, and the previous, $k-1$, with a certain threshold, $T(k)$. The value of $T(k)$ is computed using equation (3.1)

$$T(k) = \frac{a}{M}\sum_{j=1}^{M}|x(k-j) - x(k-1-j)| + b, \tag{3.1}$$

where a and b, are two parameters to adjust according to the given data, $x(k)$, the process from which the outliers are to be removed, and M is the number of previous differences to take into account. The reason for the term b to exist is that when the average absolute value for the previous differences is small, a considerable jump is necessary for a given value to be considered as an outlier, whereas, when that average is bigger, the threshold, $T(k)$, can be approximated by said average times a constant, a. Both a and b are to be determined after some experiments are performed, to see which pair of values provides the best results.

If the difference between the current time instant, k, and the previous, $k-1$, is bigger than $T(k)$, $x(k)$ is replaced by another value, according to equation (3.2)

$$x(k) = \begin{cases} \frac{1}{M}\sum_{j=1}^{M}|x(k-j) - x(k-1-j)| & x(k) - x(k-1) \geq 0 \\ -\frac{1}{M}\sum_{j=1}^{M}|x(k-j) - x(k-1-j)| & x(k) - x(k-1) < 0 \end{cases}. \tag{3.2}$$

3.1.2 Seasonal differencing

A seasonal differencing filter [19] allows to remove seasonal aspects of the process by applying a filter with poles on the unit circle. If in the unfiltered process, $x(k)$, exhibits seasonality of period T, it can be removed by applying the following filter, represented in the delay operator as

$$y(k) = (1 - q^{-T})x(k). \tag{3.3}$$

If $x(k)$ has no more seasonal parts, the process at the output of the filter, $y(k)$, is non-seasonal.

3.2 Model Identification

For model identification, two main methods are presented, prediction-error method and a variation on the classic least squares method. Besides that, in this section, the method to estimate the ideal number of parameters, p, is described.

3.2.1 Prediction-error method

The method to be used is the one implemented in the already existing MATLAB function *armax* that is based on a prediction error method [21]. This function returns the identified model, sys, given a data set, x, and the degrees of the polynomials $A^*(q^{-1})$ and $C^*(q^{-1})$, which are n_a and n_c, respectively. The syntax is the following,

$$sys = armax(x, [n_a \; 0 \; n_c \; 0]).$$

3.2.2 Least squares method

Another method to implement is the least squares method. Though equation (2.16) provides a solution for the estimation of the parameters of the model, it assumes that the parameters do not change as time, k, advances, so a recursive algorithm [22] is proposed below.

Considering that n_c is different from zero, meaning that the noise is colored, there are no guarantees that the estimates given by the regular least squares method are not biased, and the $\boldsymbol{c}$ parameters need to be identified for prediction purposes. Below, an extended least squares method [21] to workaround this issue is proposed.

3.2.2.A Recursive least squares method

For the recursive version of the least squares method, let $P(k)$, denote the estimation error covariance matrix, and $K(k)$ the Kalman gain, for a given time instant, k. The Recursive Least Squares (R-LS) method requires initial values for the estimates, $\hat{\boldsymbol{\theta}}(\boldsymbol{k})$, and the covariance matrix, $P(k)$. For the following time instants, these quantities are updated taking into account their value in the previous time instant and new information. Equations (3.4) to (3.6) allow to compute parameter estimates for every k. The equations that propagate R-LS estimates are

$$P(k) = P(k-1) - \frac{P(k-1)\boldsymbol{\varphi}(k)\boldsymbol{\varphi}^T(k)P(k-1)}{1 + \boldsymbol{\varphi}^T(k)P(k-1)\boldsymbol{\varphi}(k)} \tag{3.4}$$

$$K(k) = P(k)\boldsymbol{\varphi}(k) \tag{3.5}$$

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + K(k)[y(k) - \hat{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)] \tag{3.6}$$

The initial estimates carry great uncertainty, so the corresponding covariance matrix's values should be high. This leads to a Kalman gain composed of high values as well, to speed up convergence of the estimates to the real values of the parameters. As the estimates converge, the uncertainty decreases lowering the values of the covariance matrix and Kalman gain, which lowers convergence time but leads to smaller fluctuations in the values of the estimates.

3.2.2.B Exponential forgetting

When a sufficient amount of time has passed, and, in that time, the data represents well a system with non varying parameters, the Kalman gains decrease, and the parameter estimates converge to certain values, becoming almost constant. If the system parameters change, since the Kalman gains

are very small, it takes a lot of time for the estimates to converge to the new values. Let the new cost function be

$$J(\boldsymbol{\theta}) = \frac{1}{2M}\sum_{k=1}^{M}\lambda^{M-k}[y(k) - \boldsymbol{\theta}^T\boldsymbol{\varphi}(k)]^2, \tag{3.7}$$

where λ is the forgetting factor, $\lambda \in]0,1[$, and its purpose is to give less weight to past data. The smaller the forgetting factor, the less data samples influence the estimates, and the faster the algorithm adapts to changes in the parameters of the system that is being identified.

A constant forgetting factor can lead to the rapid growth of the covariance matrix and Kalman gain. A proposed solution for this problem is to have the forgetting factor that varies in time. The equations of the algorithm with a time varying forgetting factor are presented below.

$$\boldsymbol{\epsilon}(k) = y(k) - \hat{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k), \tag{3.8}$$

$$K(k) = \frac{P(k-1)\boldsymbol{\varphi}(k)}{\lambda(k-1) + \boldsymbol{\varphi}^T(k)P(k-1)\boldsymbol{\varphi}(k)}, \tag{3.9}$$

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + K(k)\boldsymbol{\varepsilon}(k), \tag{3.10}$$

$$\lambda(k) = 1 - [1 - \boldsymbol{\varphi}^T(k)K(k)]\boldsymbol{\varepsilon}^2(k)/\varepsilon_0, \tag{3.11}$$

$$P(k) = [I - K(k)\boldsymbol{\varphi}^T(k)]P(k-1)/\lambda(k). \tag{3.12}$$

The constant ε_0 represents the mean value of the prediction error, and that value is determined after performing some experiments.

In order to a have a sufficient amount of information to perform the identification it is convenient to impose a minimum value for the forgetting factor. Because of this, it is required to include the condition

$$\text{If } \lambda(k) < \lambda_{min} \implies \lambda(k) = \lambda_{min}. \tag{3.13}$$

3.2.2.C Extended least squares method

For an ARMA model, equation (2.13) can be written as

$$y(k) = \boldsymbol{\theta}^T\boldsymbol{\varphi}(k) + v(k), \tag{3.14}$$

where $v(k)$ is a linear combination of white noise samples $e(k)$, and is given by

$$v(k) = e(k) + c_1 e(k-1) + \ldots + c_{n_c},$$

with

$$\boldsymbol{\theta} = [a_1 \; \ldots \; a_{n_a}]^T,$$

and

$$\boldsymbol{\varphi}(k) = [-y(k-1) \; \ldots \; -y(k-n_a)]^T.$$

The regular least squares method does not estimate the $\boldsymbol{c}$ parameters of the system, nor guarantees a non biased estimation of the $\boldsymbol{a}$ parameters.

Let equation (3.14) be written in a different way,

$$y(k) = \boldsymbol{\theta}_{\boldsymbol{es}}^T\boldsymbol{\varphi}_{\boldsymbol{es}}(k) + e(k),$$

where $\boldsymbol{\theta_{es}}$ and $\varphi_{\boldsymbol{es}}$ are given by

$$\boldsymbol{\theta_{es}} = [a_1 \ \dots \ a_{n_a} \ c_1 \ \dots \ c_{n_c}]^T,$$

and

$$\boldsymbol{\varphi_{es}}(k) = [-y(k-1) \ \dots \ -y(k-n_a) \ \varepsilon(k-1) \ \dots \ \varepsilon(k-n_c)]^T,$$

where $\varepsilon(k-1) \ \dots \ \varepsilon(k-n_c)$ are the estimates of $e(k-1) \ \dots \ e(k-n_c)$, which are not available for measurement, so they must be estimated. Suppose these estimates are known at a certain time instant, k.

The equations for the Recursive Extended Least Squares (R-ELS) method with exponential forgetting are

$$\boldsymbol{\varepsilon}(k) = y(k) - \boldsymbol{\hat{\theta}}^T_{\boldsymbol{es}}(k-1)\boldsymbol{\varphi_{es}}(k), \tag{3.15}$$

$$K(k) = \frac{P(k-1)\boldsymbol{\varphi_{es}}(k)}{\lambda(k-1) + \boldsymbol{\varphi}^T_{\boldsymbol{es}}(k)P(k-1)\boldsymbol{\varphi_{es}}(k)}, \tag{3.16}$$

$$\boldsymbol{\hat{\theta}_{es}}(k) = \boldsymbol{\hat{\theta}_{es}}(k-1) + K(k)\boldsymbol{\varepsilon}(k), \tag{3.17}$$

$$\lambda(k) = 1 - [1 - \boldsymbol{\varphi}^T_{\boldsymbol{es}}(k)K(k)]\boldsymbol{\epsilon}^2(k)/\varepsilon_0, \tag{3.18}$$

$$\text{If } \lambda(k) < \lambda_{min} \implies \lambda(k) = \lambda_{min}. \tag{3.19}$$

$$P(k) = [I - K(k)\boldsymbol{\varphi}^T_{\boldsymbol{es}}(k)]P(k-1)/\lambda(k). \tag{3.20}$$

With equations (3.15) to (3.20) all the conditions are gathered to perform the identification of the system characterized by an ARMA model using a modified least squares method.

3.2.3 Estimation of the number of parameters

In order to identify any model, the number of parameters, p, necessary to characterize the ARMA model that best fits the data, needs to be estimated. To do so, the data to be used in the identification is divided in two parts, a training set, that contains about 65% of the total data, and a test set that contains the remaining 35%. The training set is used for identification purposes, and then the test set is used to make predictions one step ahead, $\hat{y}(k+1|k)$. The number of parameters, p, which in this particular case is the set of n_a and n_c parameters of the ARMA model, is chosen according to

$$n_a*, n_c* = \arg\min_{n_a, n_c} \frac{1}{M} \sum_{k \in test \ set} (y(k-1) - \hat{y}(k|k-1))^2, \tag{3.21}$$

where M denotes the size of the test set. In other words, the goal is to minimize the variance of the prediction error when making predictions using the test set.

The goal is to have a model that would fit all data corresponding to a given segment, but the given data is limited, so a training set and a test set are used to avoid overfitting.

For the sake of simplicity, the number of parameters of each of the polynomials of the model, n_a and n_c shall not be larger than 7.

3.3 Prediction

Let $\hat{y}(k+m|k),\ m \geq 1$, be the predicted value, and

$$J = E[(y(k+m) - \hat{y}(k+m|k))^2|O^k] \tag{3.22}$$

be the variance of the prediction error in steady state, which is the value to minimize, making it the cost function, where $E[\ ]$ denotes the expected value, O^k are the observations until time instant k, given that $\hat{y}(k+m|k)$ does not depend on future values [20].

The process $y(k+m)$ is given by

$$y(k+m)+a_1 y(k+m-1)+\ldots+a_n y(k+m-n) = e(k+m)+c_1 e(k+m-1)+\ldots+c_n e(k+m-n), \tag{3.23}$$

which, in the backward shift operator, is given by

$$y(k+m) = \frac{1+c_1 q^{-1}+\ldots+c_n q^{-n}}{1+a_1 q^{-1}+\ldots+a_n q^{-n}} e(k+m) = \frac{C^*(q^{-1})}{A^*(q^{-1})} e(k+m). \tag{3.24}$$

The quotient $\frac{C^*(q^{-1})}{A^*(q^{-1})}$ can be expanded, using long division

$$\frac{C^*(q^{-1})}{A^*(q^{-1})} = F_m^*(q^{-1}) + q^{-m} \frac{G_m^*(q^{-1})}{A^*(q{-1})}, \tag{3.25}$$

the polynomial $F_m^*(q^{-1})$ being of order $m-1$, can be described as

$$F_m^*(q^{-1}) = 1 + f_1 q^{-1} + \ldots + f_{m-1} q^{-m+1}, \tag{3.26}$$

and the quotient $\frac{G_m^*(q^{-1})}{A^*(q{-1})}$ can be described as a polynomial of infinite order

$$\frac{G_m^*(q^{-1})}{A^*(q{-1})} = f_m + f_{m+1} q^{-1} + f_{m+2} q^{-2} \ldots, \tag{3.27}$$

thus yielding the following equation

$$y(k+m) = e(k+m) + f_1 e(k+m-1) + \ldots + f_{m-1} e(k+1) + f_m e(k) + f_{m+1} e(k-1) + \ldots. \tag{3.28}$$

The process $y(k+m)$ can then be written as the sum of two terms

$$y(k+m) = F_m^*(q^{-1}) e(k+m) + \frac{G_m^*(q^{-1})}{A^*(q{-1})} e(k), \tag{3.29}$$

where the first term corresponds only to future values of $e(k)$, given that polynomial $F_m^*(q^{-1})$ is of order $m-1$, and the second term corresponds to past and present values of $e(k)$. This representation will be advantageous later.

Substituting now, in equation (3.22), $y(k+m)$ according to equation (3.29), follows the next expression

$$J = E[(F_m^*(q^{-1}) e(k+m) + \frac{G_m^*(q^{-1})}{A^*(q{-1})} e(k) - \hat{y}(k+m|k))^2|O^k]. \tag{3.30}$$

Some terms of equation (3.30) only depend on future observations and other terms only depend on past and present values. Denoting

$$\alpha = \frac{G_m^*(q^{-1})}{A^*(q{-1})} e(k) - \hat{y}(k+m|k),$$

which has all the terms that depend on past and present observations,and

$$\beta = F_m^*(q^{-1})e(k+m),$$

which has all the terms that depend on future observations. The cost function can, thus, be expressed in terms of α and β

$$J = E[(\alpha+\beta)^2|O^k],$$

the square can be expanded

$$J = E[\alpha^2+\beta^2+2\alpha\beta|O^k],$$

and being the expected value a linear operator, the cost function can be written as

$$J = E[\alpha^2|O^t] + E[\beta^2|O^t] + 2E[\alpha\beta|O^k].$$

Since $e(k)$ is white noise, its mean is zero and its samples are uncorrelated, and the system causal, past and present observations do not depend on future observations of $e(k)$, and neither do those future values of $e(k)$ depend on past and present observations making α and β uncorrelated. This means that the cost function can be written as

$$J = E[\alpha^2|O^k] + E[\beta^2|O^k] + 2E[\alpha|O^k]E[\beta|O^k],$$

because the expected value of the product of two uncorrelated terms is the product of the expected value of each of them. Further more, besides being uncorrelated, α and β have zero mean, so the next equality is true

$$J = E[\alpha^2|O^k] + E[\beta^2|O^k].$$

Considering that the expected value of a term that depends on certain values, given those values, is the term itself, and the expected value of a term that does not depend on the given values, is the expected value of that term unconditioned. Given that the term α^2 only depends on past and present observations, the term β^2 only depends on future observations and O^k represents past and present observations, the following equation holds

$$J = \alpha^2 + E[\beta^2].$$

Substituting α and β by the corresponding expressions

$$J = (\frac{G_m^*(q^{-1})}{A^*(q{-}1)}e(k) - \hat{y}(k+m|k))^2 + E[(F_m^*(q^{-1})e(k+m))^2]. \tag{3.31}$$

The minimum value of the cost function is achieved when

$$\frac{G_m^*(q^{-1})}{A^*(q{-}1)}e(k) - \hat{y}(k+m|k) = 0,$$

because the other term does not depend on $\hat{y}(k+m|k)$. So, the optimal predictor is given by

$$\hat{y}(k+m|k) = \frac{G_m^*(q^{-1})}{A^*(q{-}1)}e(k). \tag{3.32}$$

The previous expression is not very useful in the means that the signal $e(k)$ cannot be measured. Considering the next equality

$$e(t) = \frac{A^*(q{-}1))}{C^*(q{-}1)} y(k), \tag{3.33}$$

which is valid because $A(q)$ and $C(q)$ are polynomials the same order, $e(k)$ in equation (3.32) can be substituted using according to equation (3.33), thus yielding the expression for the optimal predictor in steady state

$$\hat{y}(k+m|k) = \frac{G_m^*(q^{-1})}{C^*(q{-}1)} y(k). \tag{3.34}$$

To have a measure of how good is the prediction, the variance of the prediction error can be analyzed, which is the cost function used to obtain the optimal predictor. Given this, the variance of the prediction error, considering the optimal predictor, is given by

$$E[(y(k{+}m){-}\hat{y}(k{+}m|k))^2|O^k] = E[(F_m^*(q^{-1})e(k{+}m))^2] = E[((1{+}f_1q^{-1}{+}\ldots{+}f_{m-1}q^{-m+1})e(k{+}m)^2].$$

Considering the fact that the signal $e(k)$ is white noise, meaning that its samples are uncorrelated, the variance of the prediction error can be written as

$$E[(y(k+m) - \hat{y}(k+m|k))^2|O^k] = (1 + f_1^2 + \ldots + f_{m-1}^2)\sigma_e^2, \tag{3.35}$$

where σ_e^2 is the variance of the signal $e(k)$.

4

Architectures

Contents

This chapter presents the proposed architectures to be developed in order to solve the prediction problem.

4.1 Multiple Models

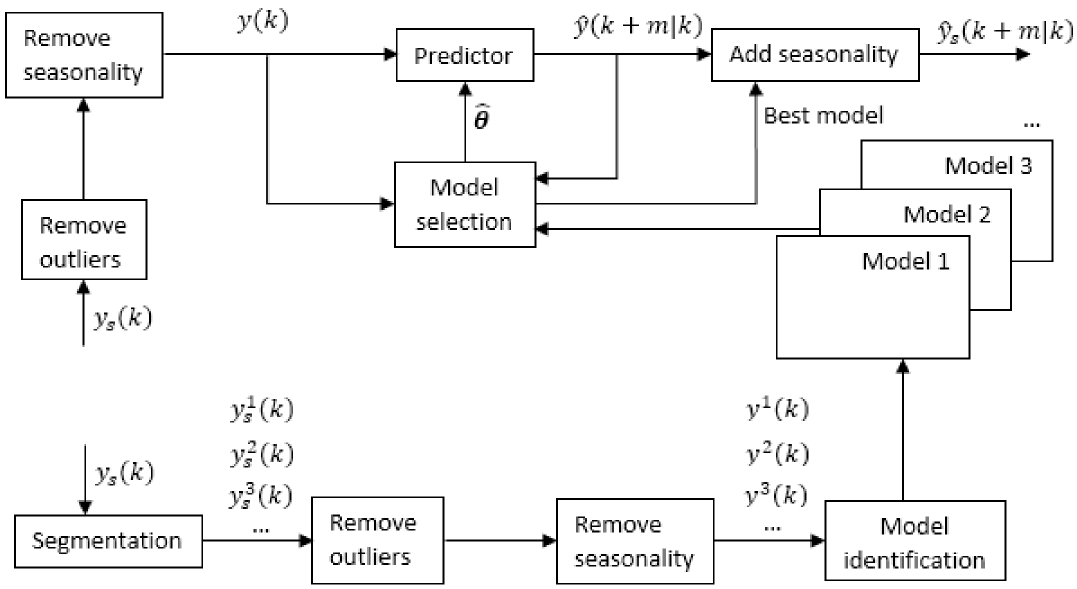


Figure 4.1: Architecture for prediction using multiple models

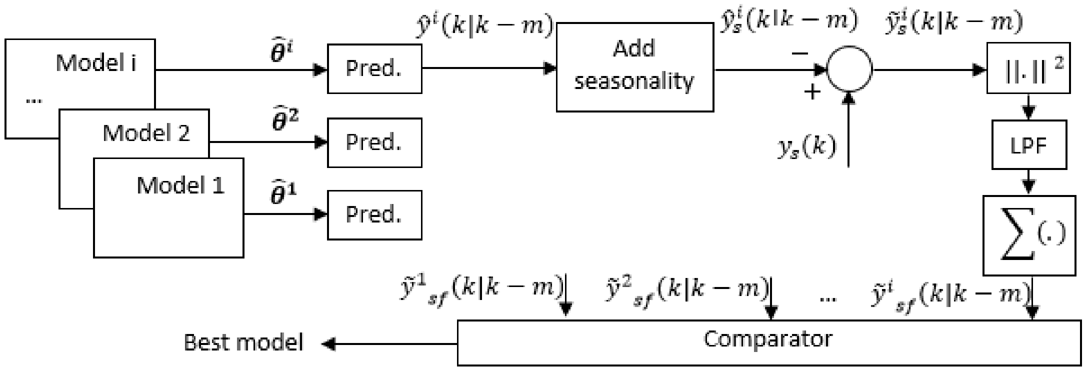


Figure 4.2: Architecture for the model selection box from Figure 4.1

The first proposed architecture is based on multiple models, as illustrated on Figure 4.1.

A model is identified, using the prediction-error method, for each segment, after the data has been properly treated (segmented, with outliers removed, and without seasonality).

After obtaining the parameter estimates, $\hat{\theta}$, for each segment, the prediction values m steps ahead,

$\hat{y}(k+m|k)$ can be calculated, using the parameters that correspond to the appropriate model, selecting one of the available ones through a method that is described later.

After obtaining the predicted values of the stationary process, seasonality, which is also dependent on the data segment, must be added by using the inverse of the seasonal differencing filter that was used to remove it.

Figure 4.2 shows how the model selection block works. To select the best model to fit the data, several predicted processes are calculated, $\hat{y}^i(k|k-m)$, one for each model i, followed by adding the corresponding seasonality to obtain the prediction of seasonal processes, $\hat{y}_s^i(k|k-m)$. Then, again for each one of the models, the prediction error, $\tilde{y_s}^i(k|k-m) = y_s(k) - \hat{y}_s^i(k|k-m)$, is computed, squared, passed through a low pass filter (LPF) for smoothing and through an integrator. The segment in which the prediction starts is known. The corresponding model is used for prediction until a time instant k_1 is reached where $\tilde{y}_s^i f(k|k-m)$ goes above a certain threshold T. When that occurs, another model is selected for prediction after the time instant k_1, namely the one with the lower filtered prediction error, $\tilde{y}_{sf}^i(k|k-m)$. This process is repeated for all the time series, so that, for each time instant, there is a corresponding model to be used for prediction.

Recall that the consumption processes correspond to the sum of the normal and HVAC parts. This means that the process described in this section is applied to both normal and HVAC consumption processes separately. The two parts are then summed in order to have a predicted process of the total consumption data. This procedure is repeated for all the architectures presented in the following sections.

4.2 Multiple models with adaptation

This next approach is still based on multiple models, but with a slight variation. There are as many main models as in the previous case, only it is assumed now that the parameters corresponding to each segment may vary in time. So, considering again the diagram from Figure 4.2, after the best main model is selected there is and adaptation step that aims to make variations on the parameters estimates of the chosen segment. Two approaches are presented below.

4.2.1 Parameter tunning

In this first case, the goal is to take the parameter estimates of the chosen model, a_i^* and c_i^*, and slightly vary them. For this, the uncertainty corresponding to each parameter, which is provided by the *armax* function in the form of a covariance matrix, gives an idea of the amplitude of its variation. Figure 4.3 shows a grid, for a case where there are only two parameters, a_1^* and a_2^*, in which, each intersection represents a slightly different model from the main one.

The best combination of parameters is then chosen similarly to way the main models are chosen, checking which of them provides the lowest prediction error.

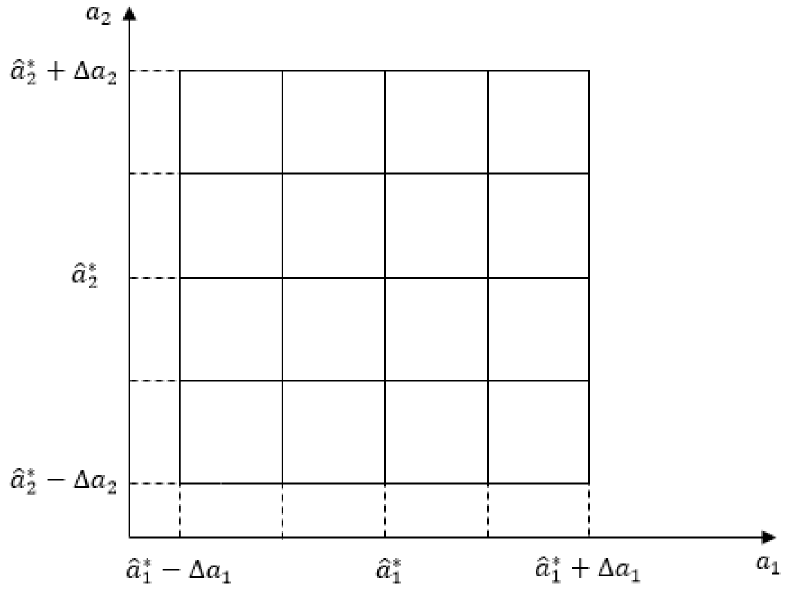


Figure 4.3: Grid of possible variations of the main model, with estimated parameters a_1^* and a_2^*

4.2.2 Recursive Extended Least Squares

In this case the variations on the main models parameter estimates are computed using the recursive extended least squares method with exponential forgetting. The steps to take when applying the method are presented in Algorithm 4.1 [22].

Algorithm 4.1 Recursive extended least squares (R-ELS) with exponential forgetting

Result:set of parameter estimates for each time instant, $\hat{\boldsymbol{\theta}}(k)$
Initialization: Define ε_0; Compute previous estimates $\hat{\boldsymbol{\theta}}(k-1)$, and $\varepsilon(k-1)\ \ldots\ \varepsilon(k-n_c)$; Define previous forgetting factor $\lambda(k-1)$; Define previous covariance matrix $P(k-1)$
for $k = 1 : T$
    Read current system output $y_s(k)$
    Compute the prediction error $\varepsilon(k) = y(k) - \varphi^T(k-1)\hat{\theta}(k-1)$
    Compute the Kalman gain $K(k) = \frac{P(k-1)\varphi(k-1)}{\lambda(k-1)+\varphi^T(k-1)P(k-1)\varphi(k-1)}$
    Compute the parameter estimates $\hat{\theta}(k) = \hat{\Theta}(k-1) + K(k)\varepsilon(k)$
    Compute forgetting factor $\lambda(k) = 1 - [1 - \varphi^T(k-1)K(k)]\varepsilon^2(k)/\varepsilon_0$
    If $\lambda(k) < \lambda_{min} \implies \lambda(k) = \lambda_{min}$
    Compute the covariance matrix $P(k) = [I - K(k)\varphi^T(k-1)]\frac{P(k-1)}{\lambda(k)}$
end

Before the algorithm starts running, some initial conditions must be computed such as the values of $\boldsymbol{\theta}(k)$ for each of the main models, and $\varepsilon(k)$. The former are the estimates computed in the simple multiple models approach,and the latter, are computed using the appropriate inverse filter . As for the initial conditions of the covariance matrix $P(k)$ and the value of ε_0, these quantities are initialized with appropriate values after some experiments are performed.

As the algorithm progresses, whenever there is a change of model, still following the diagram presented in Figure 4.2, the R-ELS algorithm needs to be re-initialized, namely the initial conditions,

$\boldsymbol{\theta}(k)$ and $P(k-1)$. In the mean time the R-ELS takes charge and performs the adaptation so that the parameter estimates can better fit the data.

4.3 Adaptive prediction

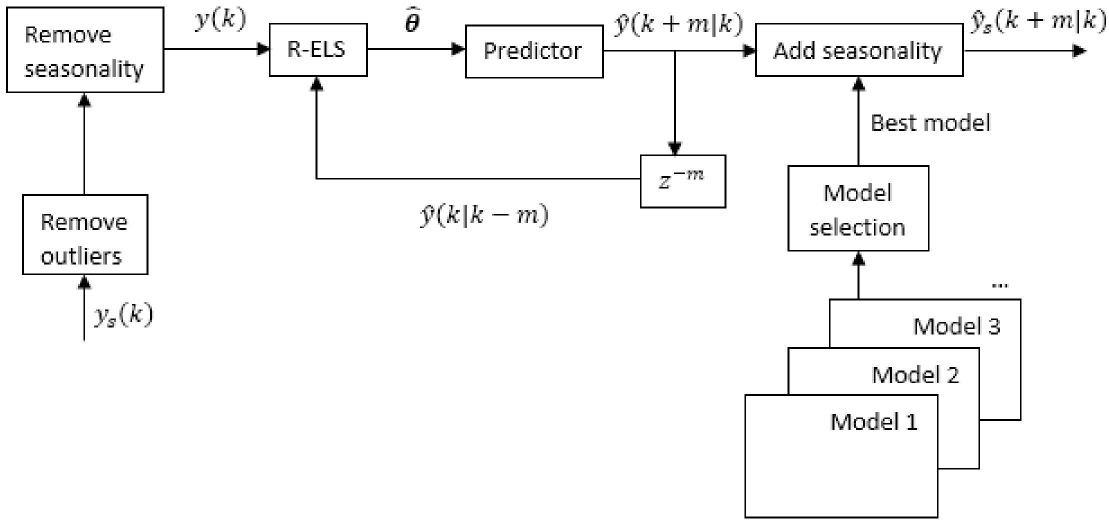


Figure 4.4: Architecture using adaptive prediction - recursive extended least squares method

The last proposed architecture does not require the identification of multiple models when predicting the values for the stationary process $y(k)$. After removing the outliers and the seasonality, the parameters of the system are estimated using the recursive extended least squares (R-ELS) method with exponential forgetting, similarly to the algorithm that was presented in Algorithm 4.1. Similarly to the previous case initial conditions must be calculated, only this time there is no re-initialization, as there are no multiple models describing the stationary process $y(k)$ in this particular case. For each time instant a prediction is then made using the estimates calculated by the R-ELS algorithm. Lastly, seasonality must be added to the stationary predicted process. Although there is no need for multiple models when computing the predicted values of the process $y(k)$, they are still required when performing this last task. The model selection box works similarly to the one presented in Figure 4.2, except there are only multiple processes after adding seasonality, whereas for the other presented cases, there were multiple ARMA model parameters, $\hat{\boldsymbol{\theta}}^i$, one for each segment.

5

Experimental Results

Contents

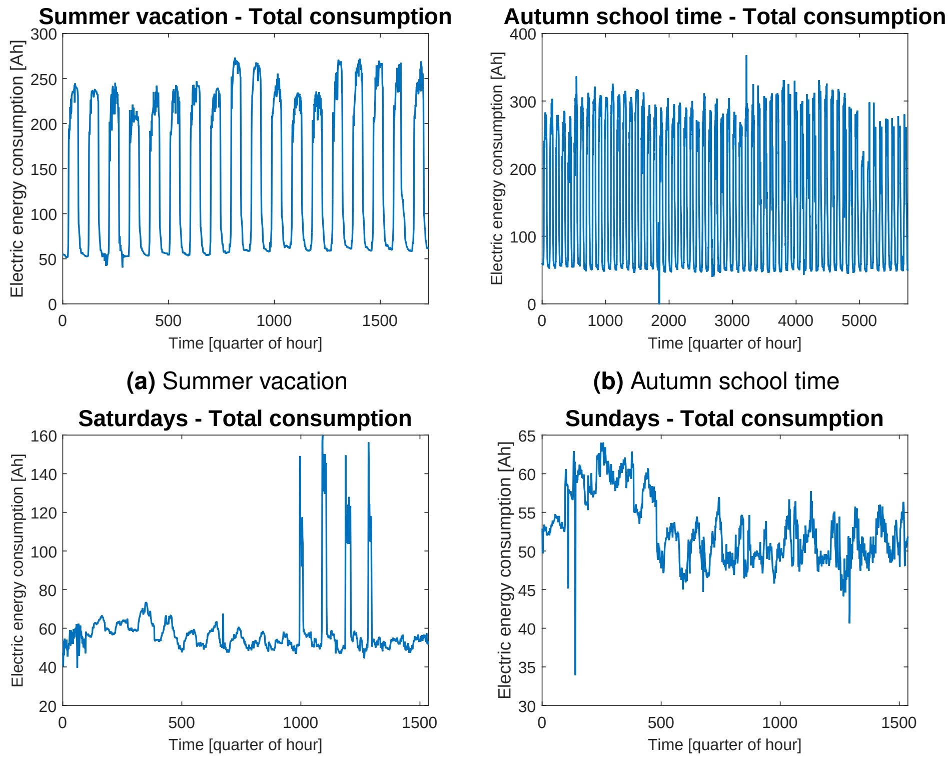
These chapter contains the experimental results of applying the methods from Chapter 3, and the architectures from Chapter 4, to the data set provided.

5.1 Data processing

For the practical purpose of this work, not all of the available data was used. Sets of observation where the data acquisition seems to have failed were left out.

5.1.1 Segmentation

Four different segments have been considered, summer vacation, autumn school time, Saturdays, and Sundays (with no differentiation between summer and autumn in the case of the latter two segments). Figure 5.1 shows the segmented data, $y_t^i(k)$.



(a) Summer vacation

(b) Autumn school time

(c) Saturdays

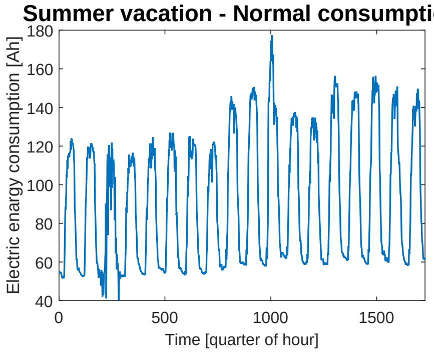
(d) Sundays

Figure 5.1: Total consumption data for each segment

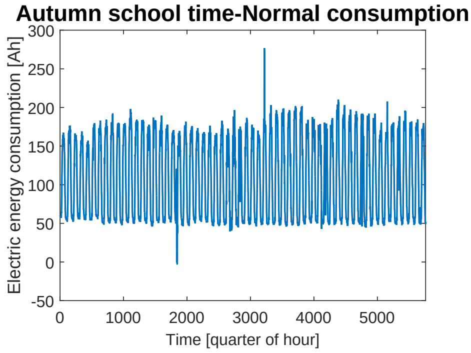
The data corresponding to each segment has been concatenated, for example, for week days, the end of a Friday is followed by the beginning of the next Monday.

Besides the segmentation in various sets of days, the normal and HVAC consumption data were separated, because it is assumed that systems that generate each part are different. Figures 5.2 and 5.3 depict, for each of the four segments, the normal and HVAC consumption data, denoted by $y_s^i(k)$, respectively.

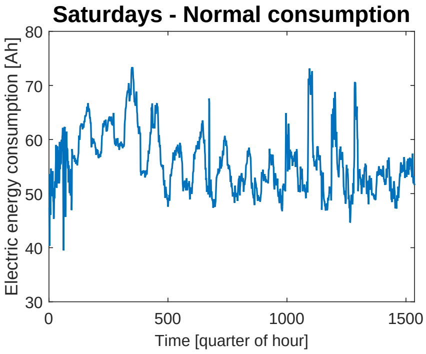
It can be seen that the normal consumption is lower in weekends than in week days, which makes sense given that a lot less people come to campus on weekends, and the HVAC is almost always



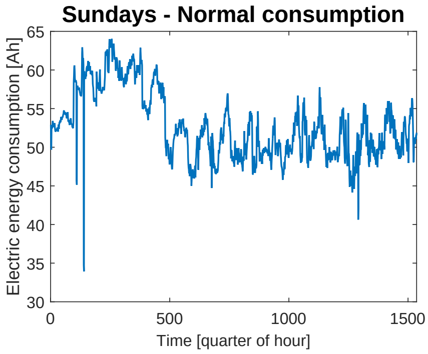
(a) Summer vacation



(b) Autumn school time

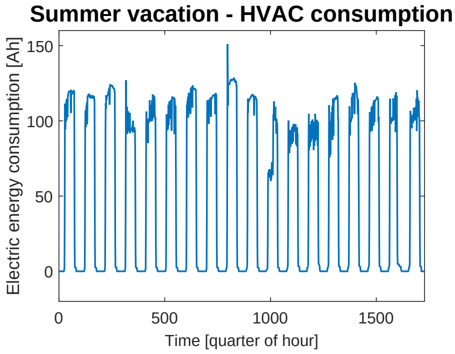


(c) Saturdays

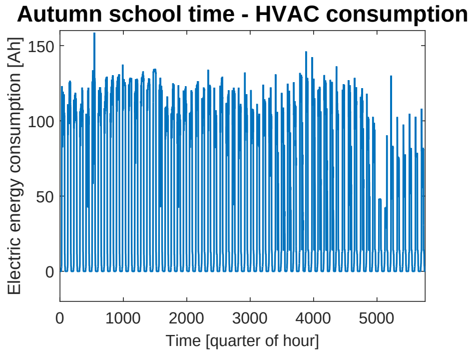


(d) Sundays

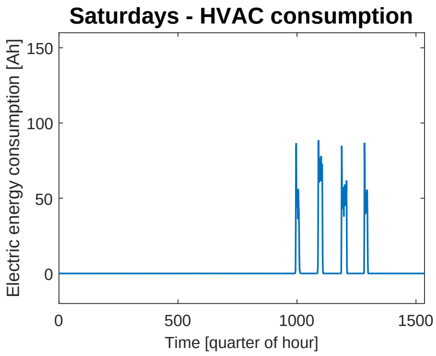
Figure 5.2: Normal consumption data for each segment



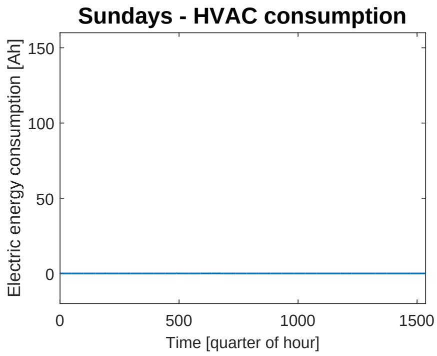
(a) Summer vacation



(b) Autumn school time



(c) Saturdays



(d) Sundays

Figure 5.3: HVAC consumption data for each segment

turned off during the weekends and nights.

5.1.2 Outlier removal

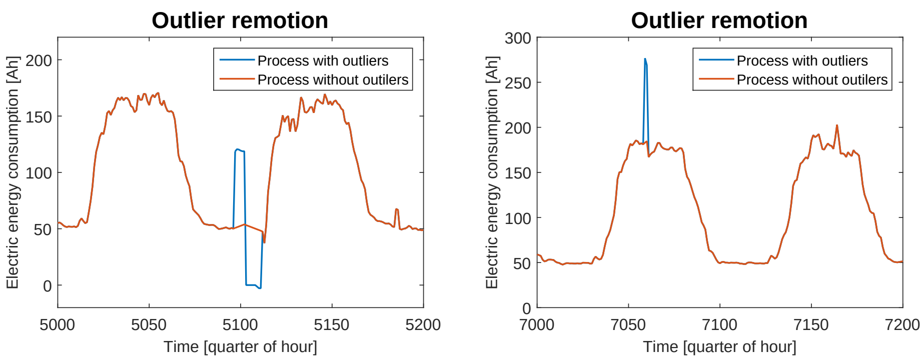
Following the segmentation and separation of the normal and HVAC components, it is necessary to remove the outliers from the processes.

For the normal consumption process, the threshold (recall equation (3.1)) above which a sample is considered an outlier, determined after conducting some experiments with several pairs of values a and b, is

$$T(k) = \frac{10}{M}\sum_{j=1}^{M} |x(k-j) - x(k-1-j)| + 10.$$

It was decided to not apply the outlier removal algorithm to the HVAC consumption process, because it is very common to have high jumps, and that is normal behavior. There would have to be abnormal jumps for it to be considered an outlier, and by visualization of the data, there are no such occurrences.

Figure 5.4 shows some data segments in which outliers were detected, before and after they were removed, using the method described in Subsection 3.1.1.



(a) Night time sensor fault

(b) Day time sensor fault

Figure 5.4: Comparison between segments of the processes $y_s^i(k)$ before and after outlier removal

5.1.3 Seasonal differencing

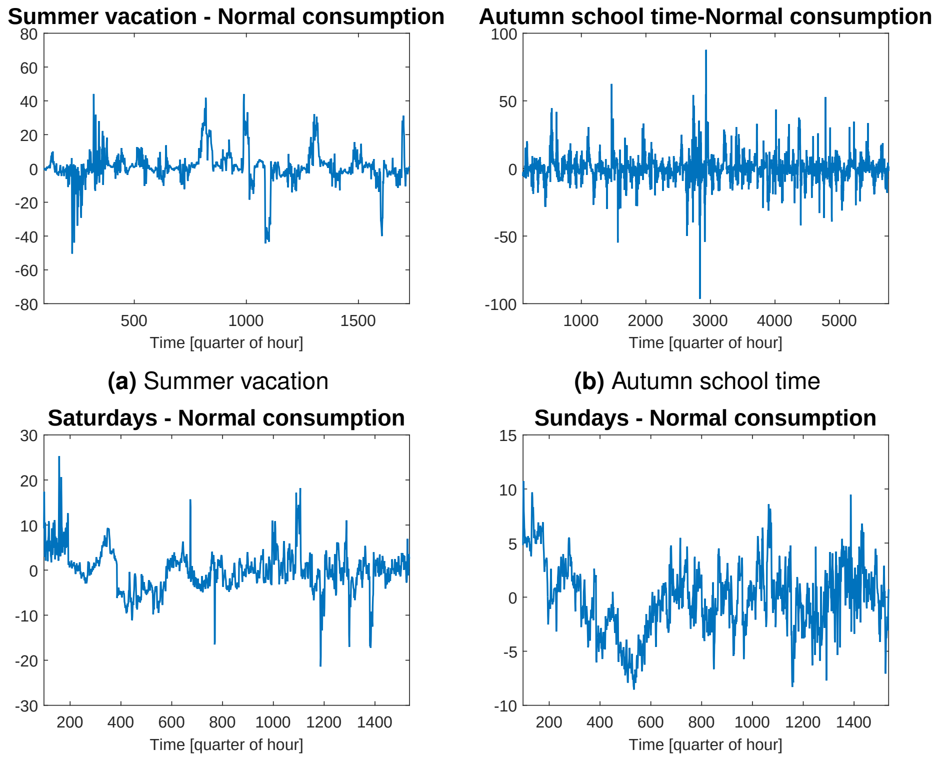
After the segmentation and outlier removal, the seasonal part of the data needs to be removed for later identification. For that, the data corresponding to each segment is passed through the filter presented in equation (3.3). In this particular case, the data has only one seasonal component, and its period corresponds to a day. Given that the data is acquired every fifteen minutes, a day corresponds to 96 samples, so $T = 96$. The filter is, then,

$$y(k) = (1 - q^{-96})y_s(k) \tag{5.1}$$

There are two particular cases to adress, which concern the data relative to the HVAC consumption during the weekends. Regarding the case of Sundays, as said before, the HVAC is always turned off, and as for Saturdays it is only turned on certain days. These two data sets do not present a daily seasonality, so for this work it is assumed that the Sunday segment is only composed of the normal

consumption, and the HVAC part of the Saturday segment is already a stationary process, so that it is not required to go through the seasonal differencing filter.

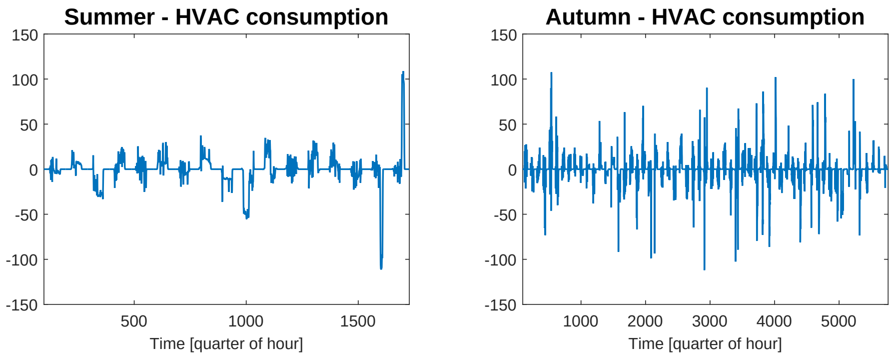
Figures 5.5 and 5.6 show the processes presented in Figures 5.2 and 5.3 without the seasonal part, here denoted as $y^i(k)$, result of the applying the filter from (5.1), with the exception for the Saturday and Sunday processes that correspond to the HVAC consumption because of the reasons mentioned above.



(a) Summer vacation **(b)** Autumn school time

(c) Saturdays **(d)** Sundays

Figure 5.5: Normal consumption data without seasonality for each segment



(a) Summer vacation **(b)** Autumn school time

Figure 5.6: HVAC consumption data without seasonality for the appropriate segments

5.2 Model identification

The result of finding out the appropriate number of parameters of the systems of the corresponding segments, which means finding the degrees of the polynomials of the corresponding ARMA models, n_a and n_c, using the method introduced in Subsection 3.2.3, is presented in Table 5.1.

	Normal		HVAC	
	n_a	n_c	n_a	n_c
Summer vacation	4	1	3	1
Autumn school time	4	0	2	2
Saturdays	4	1	3	2
Sundays	7	0	–	–

Table 5.1: Number of parameters necessary to describe the system of each segment

Most of the systems are to be identified assuming they are characterized by a purely AR model, and none of them by a ARMA model with a degree for the $C^*(q^{-1})$ polynomial larger than 1.

After obtaining the number of parameters for each segment, their values were estimated using the *armax* function. The transfer function representation of each of the identified models is presented below.

- Summer vacation
  - Normal consumption:
$$\frac{C(z)}{A(z)} = \frac{1 + z^{-1}}{1 - 0.2549z^{-1} - 0.32z^{-2} - 0.05866z^{-3} - 0.2086z^{-4}}$$
  - HVAC consumption:
$$\frac{C(z)}{A(z)} = \frac{1 + z^{-1}}{1 - 0.04883z^{-1} - 0.566z^{-2} - 0.2085z^{-3}}$$
- Autumn school time
  - Normal consumption:
$$\frac{C(z)}{A(z)} = \frac{1}{1 - 0.093z^{-1} + 0.5794z^{-2} - 0.4174z^{-3} + 0.105z^{-4}}$$
  - HVAC consumption:
$$\frac{C(z)}{A(z)} = \frac{1 - 0.5611z^{-1} - 0.1871z^{-2}}{1 + 0.5603z^{-1} - 0.2508z^{-2}}$$
- Saturdays
  - Normal consumption:
$$\frac{C(z)}{A(z)} = \frac{1 + 0.8589z^{-1}}{1 - 0.3823z^{-1} - 0.279z^{-2} - 0.02412z^{-3} - 0.01608z^{-4}}$$

– HVAC consumption:

$$\frac{C(z)}{A(z)} = \frac{1 + 1.22z^{-1} + 0.2214z^{-2}}{1 - 0.6587z^{-1} + 0.05777z^{-2} - 0.2895z^{-5}}$$

- Sundays

  – Normal consumption:

$$\frac{C(z)}{A(z)} = \frac{1}{1 - 1.106z^{-1} + 0.7389z^{-2} - 0.6836z^{-3} + 0.453z^{-4} - 0.4039z^{-5} + 0.2449z^{-6} - 0.144z^{-7}}$$

For latter analysis, the estimates of the variance of the process $e(k)$, here denominated by $\hat{\sigma}_e^2$, were computed, using the inverse filters of the ones presented above, to obtain the estimates, $\varepsilon(k)$, of the noise process, and then using the following variance estimator,

$$\hat{\sigma}_e^2 = \frac{1}{M}\sum_{k=1}^{M} \varepsilon^2(k),$$

where M is the number of samples of each process.

The resulting estimates are presented in Table 5.2.

	$\hat{\sigma}_e^2$	
	Normal	HVAC
Summer vacation	14.9797	40.7040
Autumn school time	24.7887	55.3629
Saturdays	3.1720	9.1049
Sundays	1.4212	–

Table 5.2: Estimates of the variance of the noise process, $e(k)$, for each segment

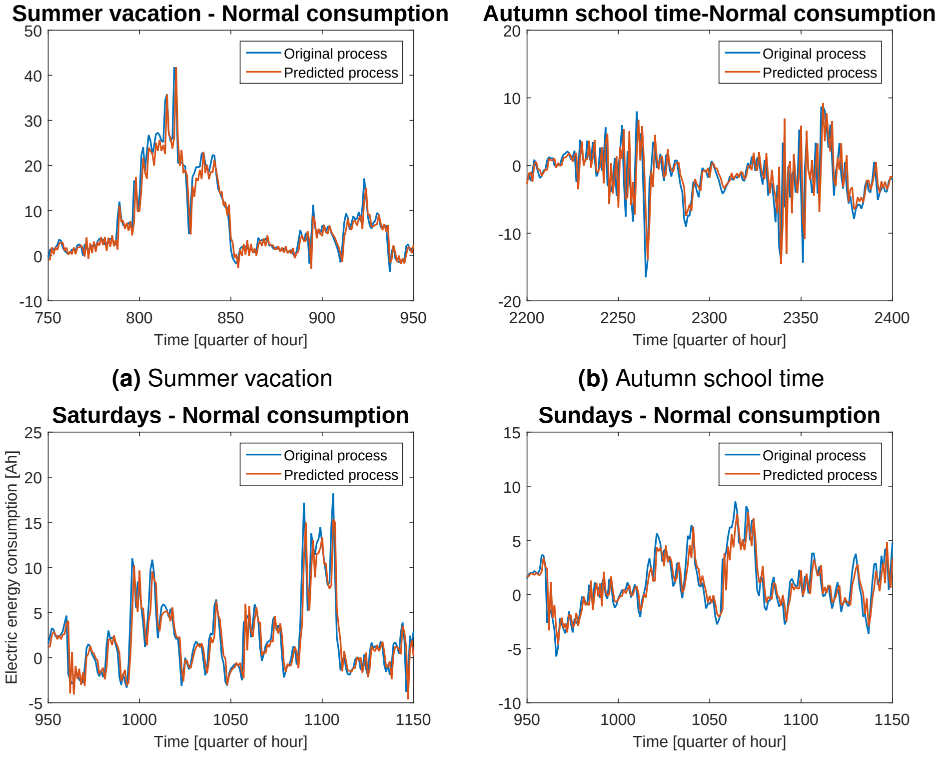
5.3 Prediction

After the system for each segment has been identified, they were used for predicting future values of the process. Figures 5.7 and 5.8 show a comparison between the stationary processes, $y^i(k)$, from Figures 5.5 and 5.6, and the one step ahead ($m = 1$) predictions, $\hat{y}^i(k+1|k)$.

The following step was to re-add the seasonality to the predicted processes. As mentioned above, it was not necessary to remove the seasonality from the process that represents the HVAC consumption on Saturdays as it is assumed to be already stationary, so the prediction is calculated using the process from Figure 5.3c. The result is shown in Figures 5.9 and 5.10, for the normal and HVAC consumption, respectively.

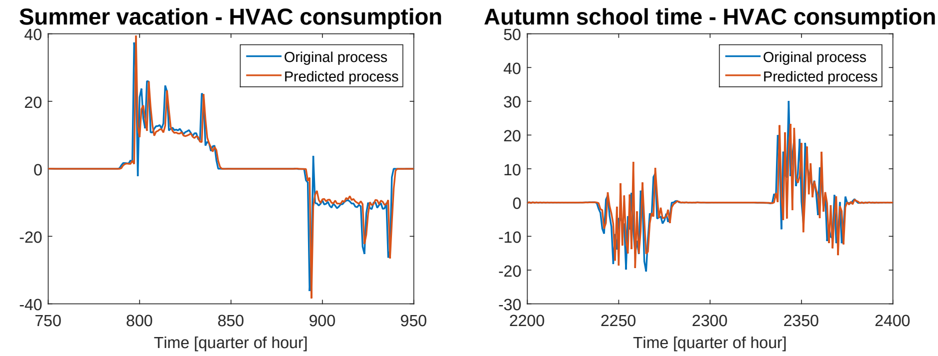
As seen in section 3.3, the variance of the prediction error is given by

$$E[(\tilde{y}_s^i(k+m|k))^2|O^k] = (1 + f_1^2 + \ldots + f_{m-1}^2)\sigma_e^2,$$



(a) Summer vacation **(b)** Autumn school time **(c)** Saturdays **(d)** Sundays

Figure 5.7: Comparison between the original and predicted ($m = 1$) stationary normal consumption processes for each segment.



(a) Summer vacation **(b)** Autumn school time

Figure 5.8: Comparison between the original and predicted ($m = 1$) stationary HVAC consumption processes for each segment.

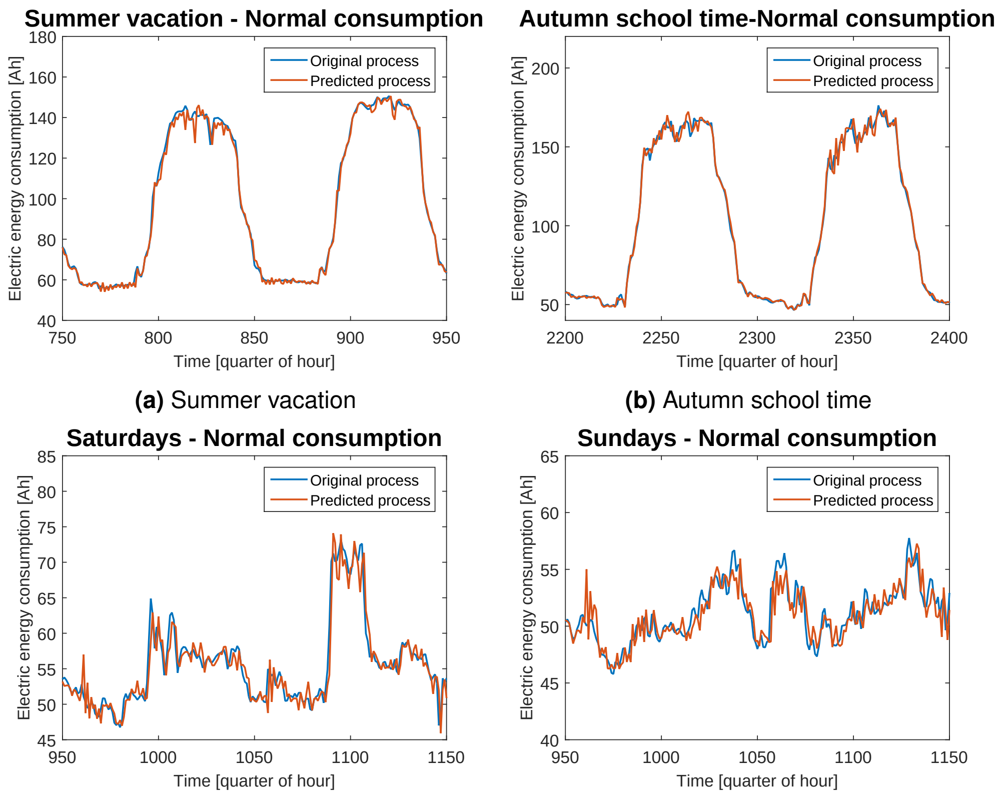
where σ_e^2 is the variance of the process $e(k)$. In particular case when $m = 1$, follows that

$$E[(\tilde{y}_s^i(k+1|k))^2|O^k] = \sigma_e^2.$$

Supposedly, the variance of the prediction error is the same as the variance of the noise process. Table 5.3 presents the estimates of the variance of the prediction error for each segment, as calculated using the following estimator

$$\hat{E}[(\tilde{y}_s^i(k+1|k))^2|O^k] = \frac{1}{M}\sum_{k=1}^{M}(\tilde{y}_s^i(k+1|k))^2, \tag{5.2}$$

where M is the number of samples of each process.



(a) Summer vacation **(b)** Autumn school time

(c) Saturdays **(d)** Sundays

Figure 5.9: Comparison between the original and predicted ($m = 1$) normal consumption processes for each segment.

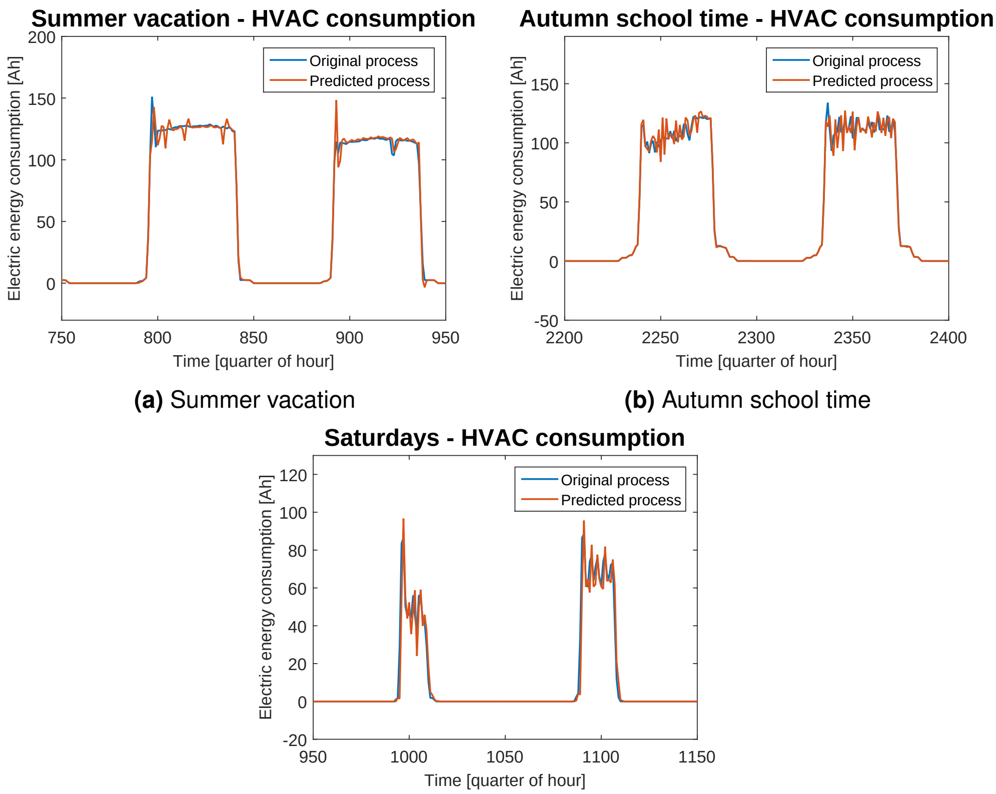
	$\hat{E}[(\tilde{y}_s^i(k+1\|k))^2\|O^k]$	
	Normal	HVAC
Summer vacation	15.0443	40.7040
Autumn school time	24.7879	55.3629
Saturdays	3.1651	9.1049
Sundays	1.4129	$-$

Table 5.3: Estimates of the variance prediction error for each segment

The presented estimates present a mean deviance of approximately 0.18%, maximum of a approximately 0.58%, and a minimum of 0% from the corresponding ones in Table 5.2. These estimates are in accordance with what was expected.

The final step for the prediction of each segment's process is to add the normal and HVAC consumption parts. The result is presented in Figure 5.11.

To show the advantages of assuming that the electrical energy consumption processes are generated by an ARMA model and a seasonal adding filter, over assuming that these values depend solely on the reading on the previous day at the same time, some predicted processed are presented



(a) Summer vacation

(b) Autumn school time

(c) Saturdays

Figure 5.10: Comparison between the original and predicted $(m = 1)$ HVAC consumption processes for each segment.

in Figures 5.12 and 5.13. The predictor used is given by,

$$\hat{y}_s^i(k) = y_s^i(k - 96).$$

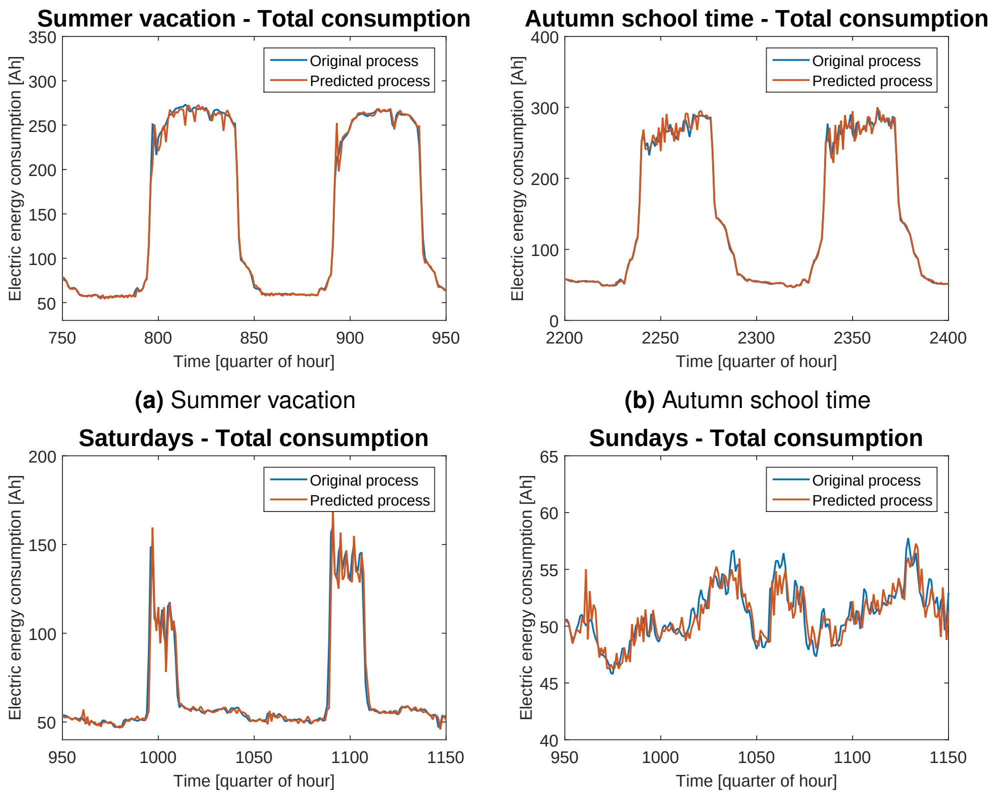
These figures show that the predicted processes do not follow the real ones as well, when compared to the predictions presented in Figures 5.9 and 5.10.

As was done before, in Table 5.3, the estimates of the prediction error were calculated using the same estimator. The results are presented in Table 5.4.

	$\hat{E}[(\tilde{y}_s^i(k+1\|k))^2\|O^k]$	
	Normal	HVAC
Summer vacation	96.8909	244.0531
Autumn school time	86.6108	219.4546
Saturdays	20.5590	113.7239
Sundays	10.1947	$-$

Table 5.4: Estimates of the variance prediction error for each segment $(\hat{y}_s^i(k) = y_s^i(k - 96))$.

These values are significantly larger than the estimated variance of the prediction error calculated



(a) Summer vacation

(b) Autumn school time

(c) Saturdays

(d) Sundays

Figure 5.11: Comparison between the original and predicted ($m = 1$) total consumption processes for each segment.

before, proving to be disadvantageous in comparison to the one step predictions calculated before.

Next is presented an analysis on the effect of the prediction horizon, m, on the actual prediction. For this, a single segment is chosen, summer vacation. Figure 5.14 shows the prediction and the real values of the stationary process for $m = 1, \ldots, 100$.

It can be seen that, as the prediction horizon increases, the predicted values tend to zero.

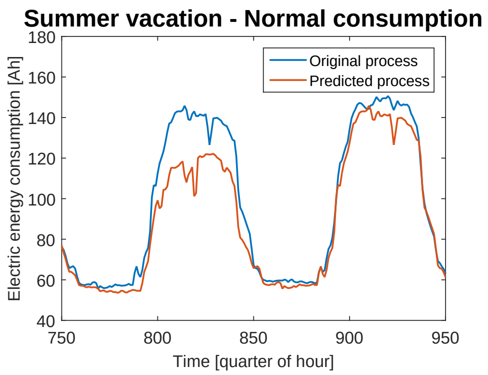
Figure 5.15 shows the prediction error for the normal and HVAC consumption during the summer vacation together with top and bottom limits of the interval $[-3\sigma_{pe}(m), +3\sigma_{pe}(m)]$, where $\sigma^2_{pe}(m)$ is the estimate of the variance of the prediction as computed using equation (3.35), for a given prediction horizon, m. This interval illustrates how σ^2_{pe} evolves as the prediction horizon increases.

As expected, the variance of the prediction error increases with the increase of m. It increases more rapidly for the first prediction horizons and then it seems to converge to a given value.

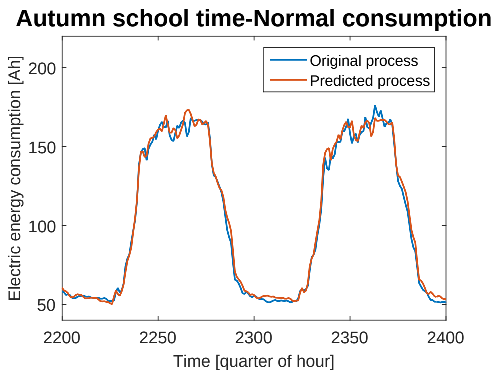
Figure 5.16 exhibits the original and predicted processes shown in Figure 5.14 after adding the seasonality, and also the interval that shows how the variance of the prediction error evolves, only in this case, it is centered on the original process, $[y(k+m) - 3\sigma_{pe}(m), y(k+m) + 3\sigma_{pe}(m)]$

For a sufficiently large prediction horizon, there is no advantage in computing the predictions of the stationary processes, as it tends to zero, and the seasonal prediction processes depend almost solely on the consumption value from the previous day.

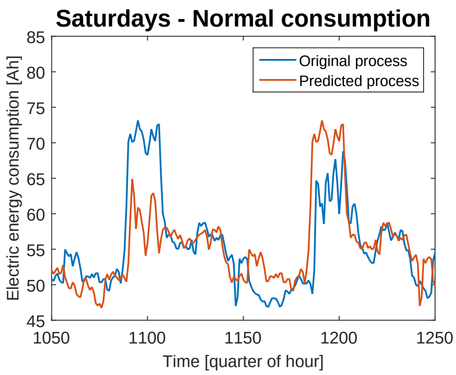
There are two ways to approach these predictions when it comes to the initial conditions of the predicted processes. The first is to, at each loop, update these initial conditions with the actual values



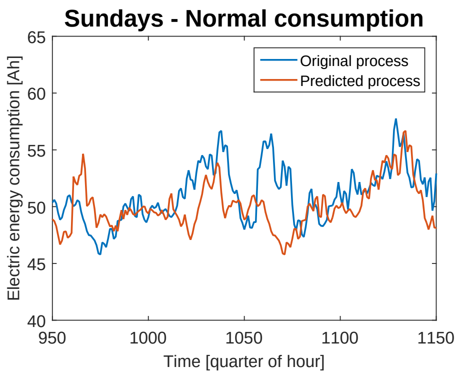
(a) Summer vacation



(b) Autumn school time



(c) Saturdays



(d) Sundays

Figure 5.12: Comparison between the original and predicted ($\hat{y}_s^i(k) = y_s^i(k-96)$) normal consumption processes for each segment.

of the process, and the second is to not do that, the initial conditions in a certain loop are updated accordingly to the prediction values computed on the previous loops. The results from the study of the variance of the prediction error for $m = 1$, were obtained using the second approach, whereas, the results of the study of the prediction horizon on the variance of prediction error were computed using the first approach.

Figure 5.17 shows two sets of the predicted values for different prediction horizons. One was obtained using the first approach, and the other the second approach.

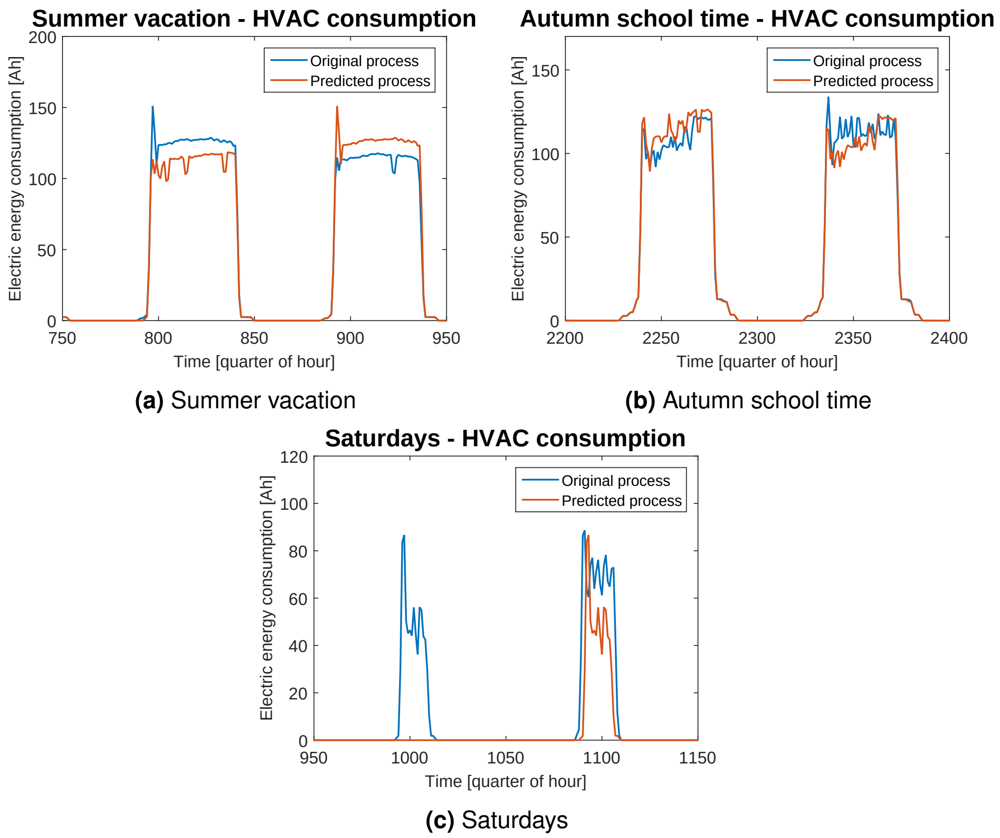
It can be seen that the predicted process that used the second approach, the initial conditions in a certain loop are updated accordingly to the prediction values computed on the previous loops, present strong oscillations.

Notice the identified models presented in section 5.2. Whenever they are not purely AR models, the polynomial $C^*(q^{-1})$ has a root on the unit circle, or inside it, but very close to the unit circle.

The system that generates the predictions, represented by equation (3.34), will have poles on, or very close to the unit circle, which can lead to high oscillations.

Probably, what is happening in this particular case is that the output of the predictor is influenced by the prediction error. When the initial conditions of the prediction processes are updated with the real values that prediction loop starts with a null prediction error, whereas, when they are not updated the prediction error is not equal to zero, which lead to the oscillations that can be seen in Figure 5.17.

Notice that, when the initial conditions of the predicted process are updated at each loop with the

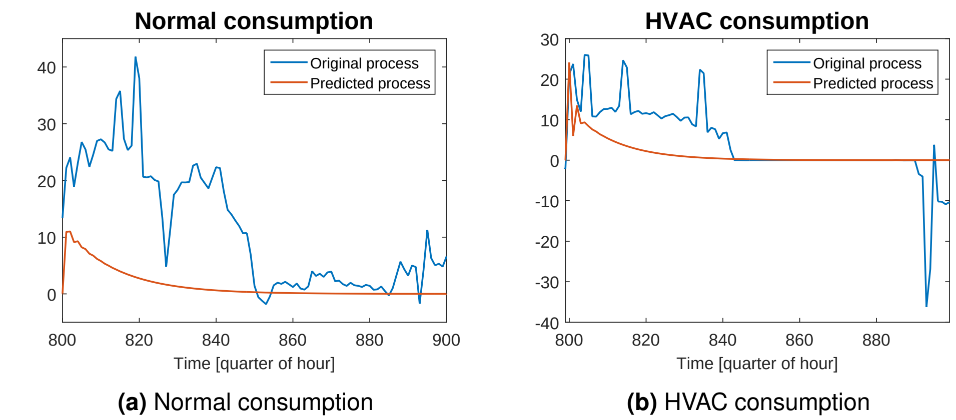


(a) Summer vacation

(b) Autumn school time

(c) Saturdays

Figure 5.13: Comparison between the original and predicted ($\hat{y}_s^i(k) = y_s^i(k-96)$) HVAC consumption processes for each segment.



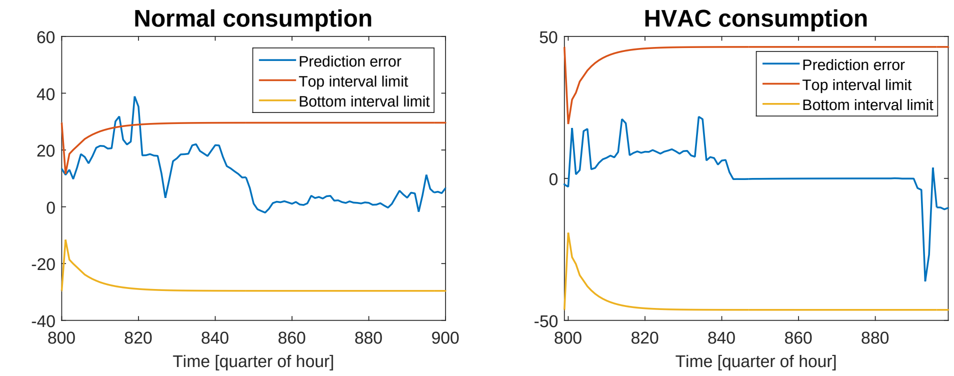
(a) Normal consumption

(b) HVAC consumption

Figure 5.14: Comparison between the original and predicted $(m = 1, \ldots,\ 100)$ stationary consumption processes for the summer vacation segment

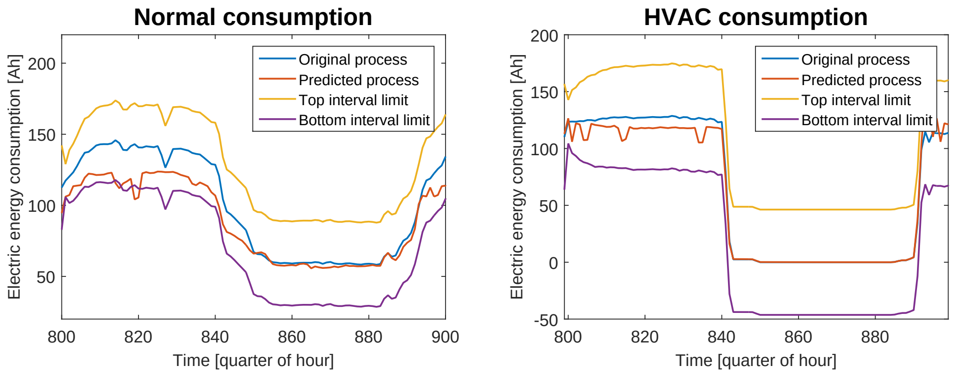
real values of the process, the variance of the prediction error does not necessarily have to be equal to the value given by equation (3.35). To avoid these high oscillations, for the remainder of this work, the initial conditions of the predicted processes will be updated at each loop with the real values of the corresponding process.

In order to solve the issue of $C^*(q^{-1})$ having zeros on the unit circle, there is a proposed method that involves the use of Kalman filtering theory to determine the optimal predictor, which is a *time-varying system*, if the polynomial $C^*(q^{-1})$ has, in fact, zeros on the unit circle [20]. This approach will



(a) Normal consumption **(b)** HVAC consumption

Figure 5.15: Prediction error of the consumption processes for the summer vacation segment, for $m = 1, \ldots,\ 100$.



(a) Normal consumption **(b)** HVAC consumption

Figure 5.16: Comparison between the original and predicted $(m = 1, \ldots,\ 100)$ consumption processes for the summer vacation segment.

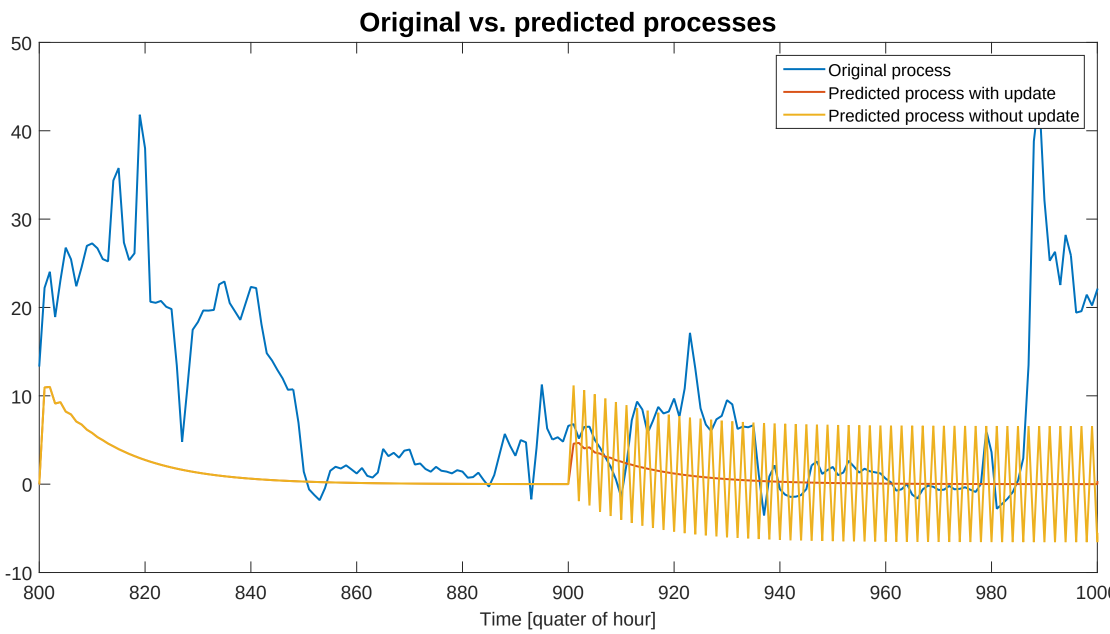


Figure 5.17: Comparison between the original and two predicted $(m = 1, \ldots,\ 100)$ consumption processes for the summer vacation segment, one with updated initial conditions, and the other without updates.

not be addressed in this dissertation.

For the remainder of this work, it is assumed that the HVAC consumption values, similarly to what happens on Sundays, are always zero. This is because it simplifies the process of implementing the algorithm of prediction for the complete consumption process, given that the ways of identification and prediction are different from the other processes.

5.4 Exogenous inputs

In order to improve the predictions of the processes, the incorporation of exogenous inputs to the system model was considered, namely the outside temperature and solar radiation. The initial approach was to describe the models, in ARMAX (the X comes from exogenous) form, and not ARMA. The stationary consumption process $y(k)$ depends on the noise process, $e(k)$, and the temperature and solar radiation according to

$$y(k) = \frac{C^*(q^{-1})}{A^*(q^1)}e(k) + \frac{B_1^*(q^{-1})}{A^*(q^{-1})}u_1(k) + +\frac{B_2^*(q^{-1})}{A^*(q^{-1})}u_2(k), \tag{5.3}$$

where $u_1(k)$ and $u_2(k)$ are the temperature and solar radiation processes, respectively.

To identify this kind of model, another MATLAB function was used, *pem*. In order to evaluate the model and decide how many parameters were necessary to describe it, the Akaike's Final Prediction Error (FPE) [21] provided by the *pem* was used. The FPE gives an idea of how well the model fits the data penalizing at the same time an elevated number of parameters. The lower the FPE the better. For simplicity reasons, the method used before for estimating the number of parameters was not used.

This approach was discarded because, for as long as the number of parameters was increasing, the FPE kept decreasing. This led to models with hundreds of parameters, which is not very good for practical purposes.

The second approach was to assume that $y(k)$ was given by

$$y(k) = \frac{C^*(q^{-1})}{A^*(q^1)}e(k)+b_{11}u_1(k)+b_{12}u_1(k-96)+b_{13}u_1(k-192)+b_{21}u_2(k)+b_{22}u_2(k-96)+b_{23}u_2(k-192). \tag{5.4}$$

This approach assumes that the values of the stationary process depend on the temperature and solar radiation at present time, one day before and two days before.

The main idea was to compute the predictions one step ahead, assuming that the processes are generated by an ARMA model, as done before, the the prediction error is computed and the estimates of the b_{ij} are the least squares estimates, as given by the following expression

$$\hat{b}_{ij} = \arg\min_{b_{ij}} \sum_{k=1}^{M} (\tilde{y}(k|k-1) - \sum_{i,j} b_{ij}u_i(k))^2,$$

where M is the number of samples used for identification.

The incorporation the information provided by the temperature and solar radiation this way proved to have no advantage when compared with the results obtained by assuming that the consumption

process is generated solely by an ARMA model, because the variance of the prediction error is approximately the same. A theory that came to mind is that, somehow, the information regarding the temperature and solar radiation is already incorporated in the ARMA models parameters, so incorporating those two processes does not provide any new information.

5.5 Multiple models architecture

The following results were obtained by using an algorithm based on the architectures presented in Figures 4.1 and 4.2.

For the smoothing step, the last 10 prediction error samples are weighted (older samples have smaller weight) and summed according to the following expression

$$\tilde{y}_{s_{smooth}}(k|k-1) = \sum_{i=0}^{9} \omega^i \tilde{y}_s(k-i|k-i-1)$$

where ω represents the weighting factor. In this case, it was chosen $\omega = 0.75$ As for the integration step, the 10 last samples of the smoothed process are summed.

Figures 5.18 and 5.19 show the normal and HVAC consumption predictions with $m = 1$, $\hat{y}_s(k+1|k)$, throughout a week, compared to the real values of the corresponding processes.

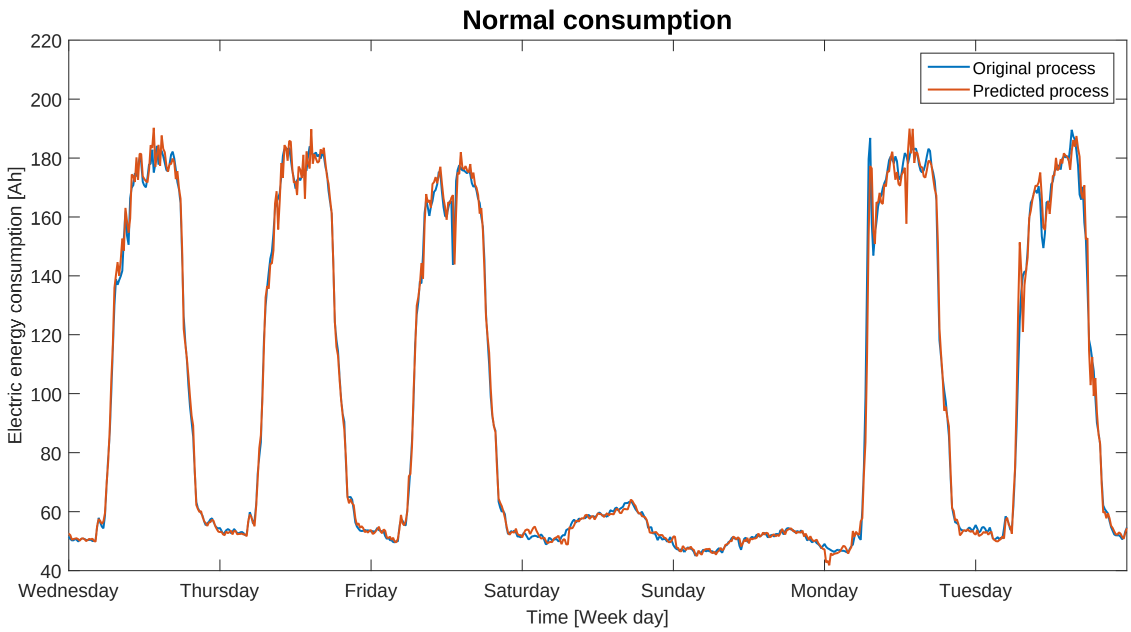


Figure 5.18: Comparison between the original and predicted ($m = 1$) normal consumption process throughout a week - Multiple models.

Figures 5.20 and 5.21 depict the sum of the normal and HVAC consumption processes over a week, for a more detailed representation, and over the complete process, for a broader view.

The estimates of the variance of the prediction error, as calculated by using the estimator from equation (5.2), for the normal, HVAC, and total processes are

- Normal Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 17.1761$,

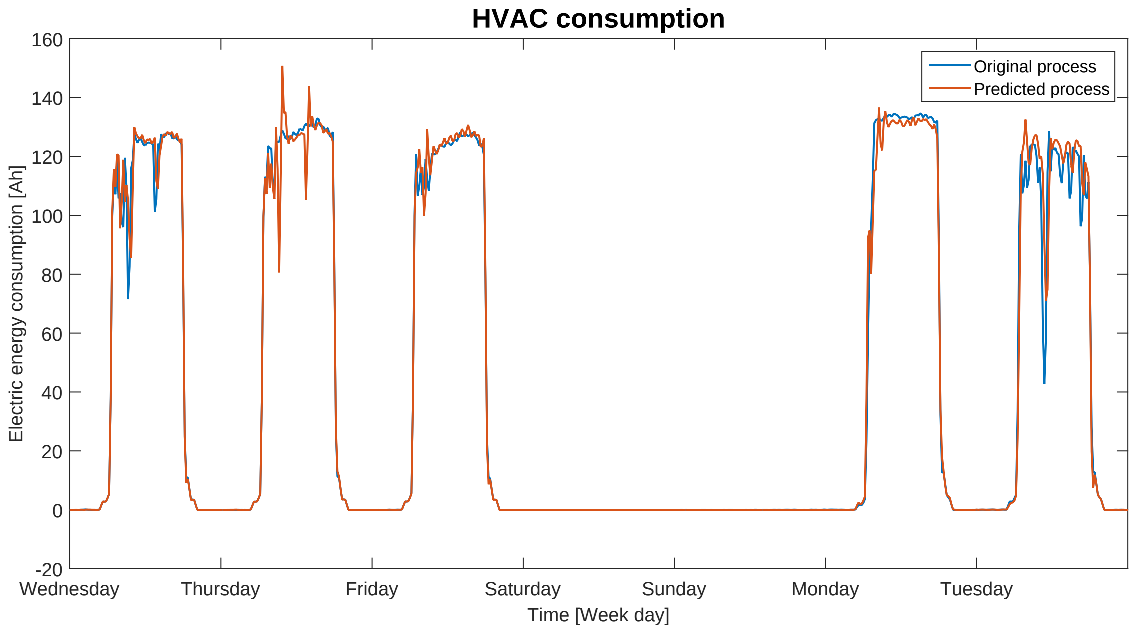


Figure 5.19: Comparison between the original and predicted $(m = 1)$ HVAC consumption process throughout a week - Multiple models.

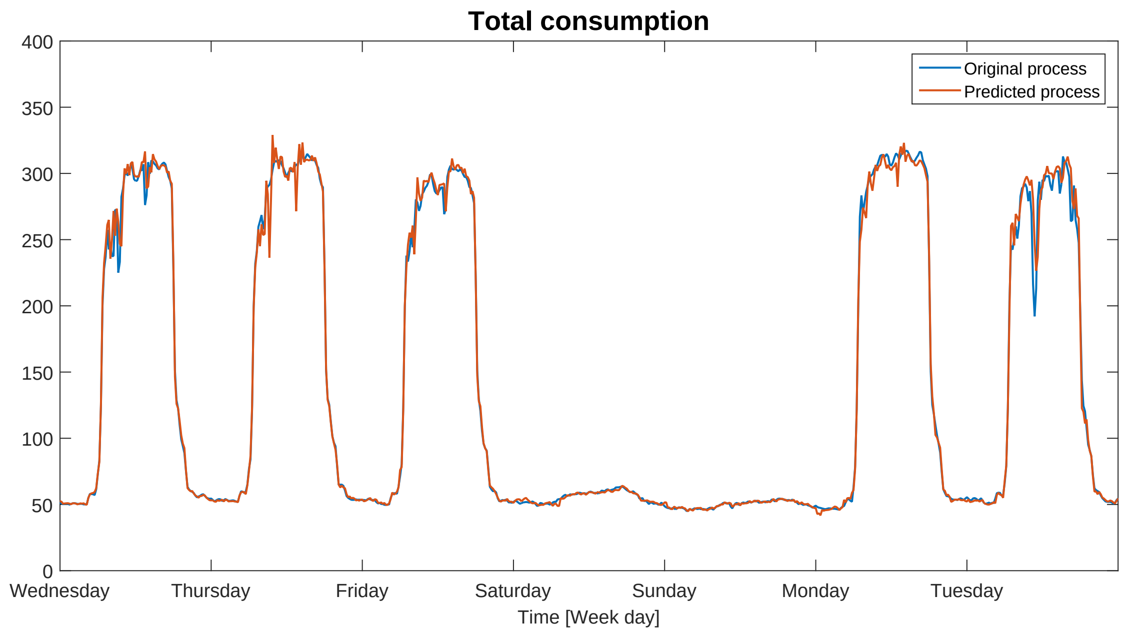


Figure 5.20: Comparison between the original and predicted $(m = 1)$ total consumption process throughout a week - Multiple models.

- HVAC Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 60.9381$,
- Total Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 101.6131$.

Figures 5.22 and 5.23 show a comparison between the predicted segment for each time instant and the real correspondence. In the first one, only a week is represented for a more detailed view, and the second one is representative for the complete process. For this to be represented if a form of

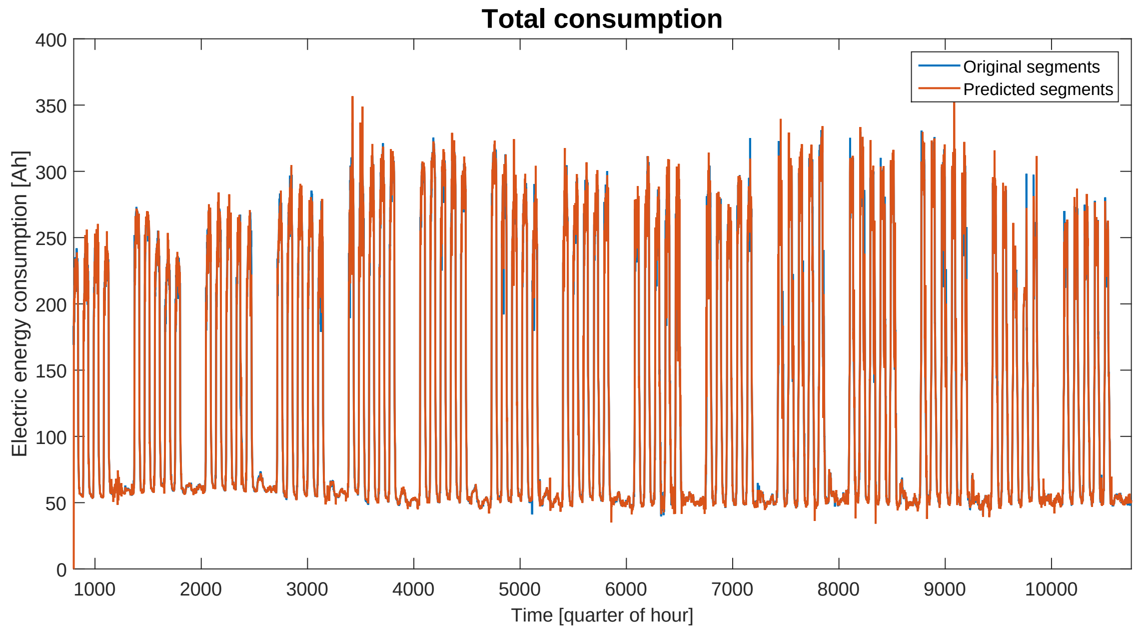


Figure 5.21: Comparison between the original and predicted $(m = 1)$ total consumption process - Multiple models.

a function, each segment corresponds to a number, as follows

$$\text{Segment} = \begin{cases} 1 & \text{if summer vacation} \\ 2 & \text{if autumn school time} \\ 3 & \text{if saturday} \\ 4 & \text{if sunday.} \end{cases} \tag{5.5}$$

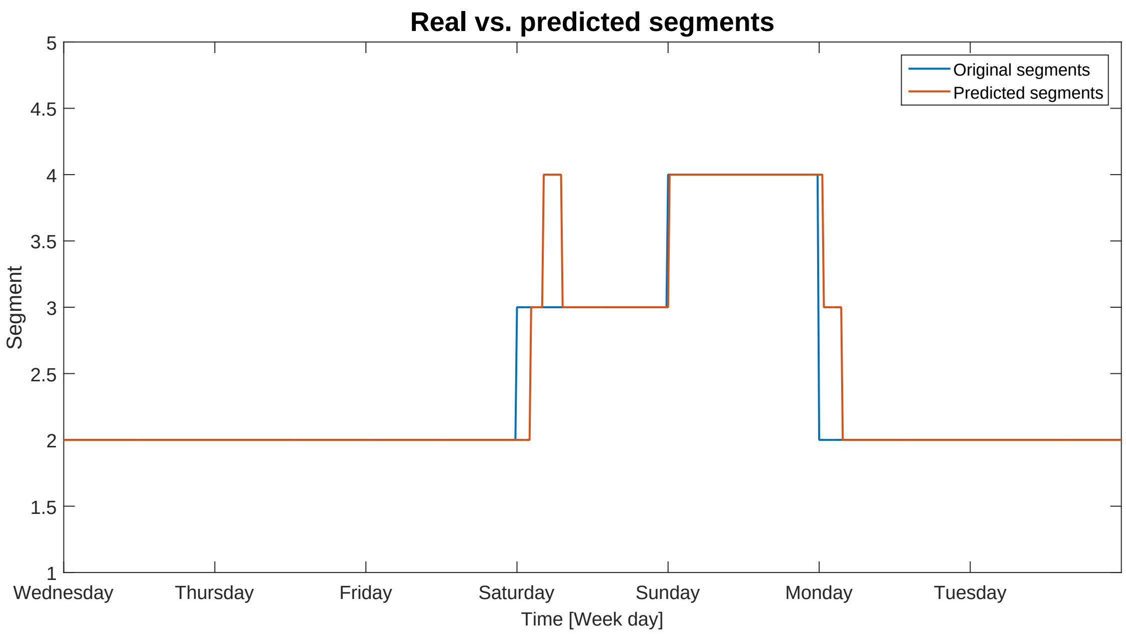


Figure 5.22: Comparison between the original and predicted $(m = 1)$ total consumption process throughout a week - Multiple models.

It can be seen in Figure 5.22 that there is sometimes a delay when the segment changes, which is probably due to the smoothing and integration phases of the algorithm. It can also be seen that, for

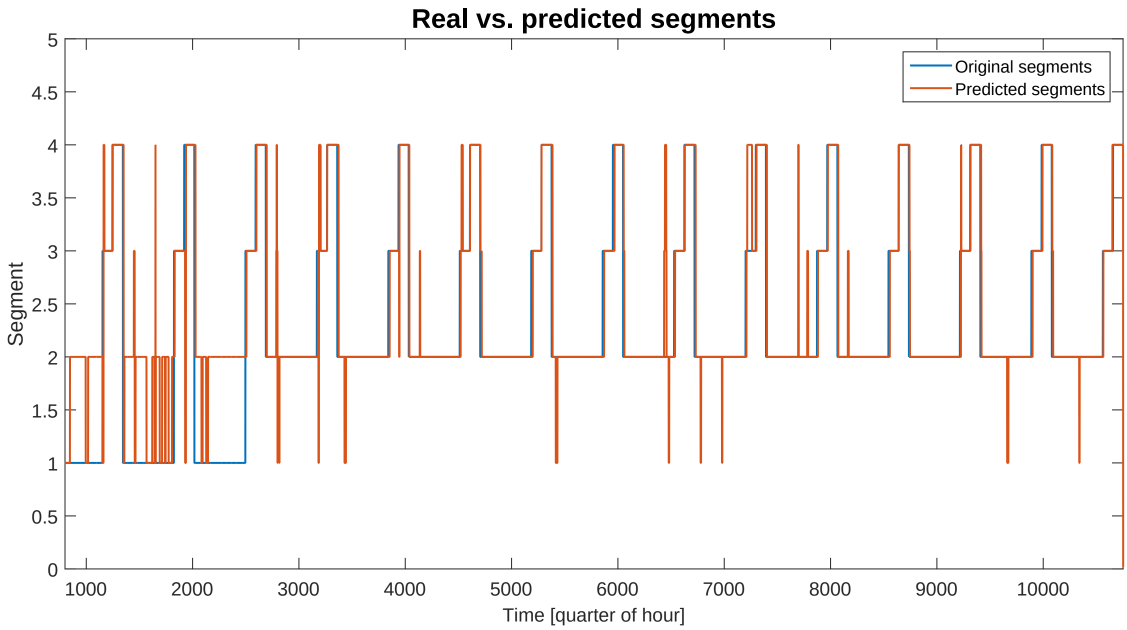


Figure 5.23: Comparison between the original and predicted ($m = 1$) total consumption process - Multiple models.

a period during Saturday, the predictor assumed that it was Sunday, that may be due to the fact that in that period the process is more similar to a Sunday.

In Figure 5.23 it is visible that most of the time during the summer vacation the algorithm predicted that it was autumn school time instead. In the beginning of the prediction process, the autumn school segment has not occurred yet, so, when adding the seasonality, it sums the previous consumption value of the segment that is the most similar, which is the summer vacation. This suggests that the identified systems for the two segments are similar and that the consumption value added when reincorporating the seasonality has a large weight on the prediction of the segments. Besides that, it can also be seen that some periods on Sundays are assumed to be Saturdays and vice-versa, and some autumn school time periods correspond to summer vacation ones. Another aspect to point out is that some periods during week days (summer vacation and autumn school time) are assumed to be weekends and vice-versa, which seems strange at first, considering that the values for consumption are considerably different, but upon further analysis it was discovered that these mis-predictions occur in instants that correspond to night time, where consumption levels are somewhat similar for all the four segments. But one can conclude that the predictions correspond to the real segments most of the time.

5.6 Multiple models with adaptive architectures

5.6.1 Parameter tunning

Consider, now, the multiple model architecture with parameter tuning as presented in Section 5.6. As mentioned before, the amount of variation applied to the parameters, is based on the covariance matrix provided by the *armax* function when computing the parameter estimates for each section.

The parameters vary in the interval $[a_i^* - 3\sigma_{a_i},\ a_i^* - 3\sigma_{a_i}]$, where a_i^* is a certain parameter estimate provided by the *armax* function, and $\sigma_{a_i}^2$ is the variance associated to the estimation of that parameter. From that interval, three values are considered for prediction purposes, the estimated parameter, and the limits of said interval. The set of values to be considered for each parameter is, then, $\{a_i^* - 3\sigma_{a_i},\ a_i^*,\ a_i^* - 3\sigma_{a_i}\}$. For each model there are 3^{na+nc} different combinations of parameters. In order to ensure stability, combinations that lead to a system with zeros or poles outside the unit circle are discarded.

Figures 5.24 and 5.25 show the normal and HVAC consumption predictions with $m = 1$, $\hat{y}_s(k+1|k)$, throughout a week, compared to the real values of the corresponding processes, for this case with parameter tunning.

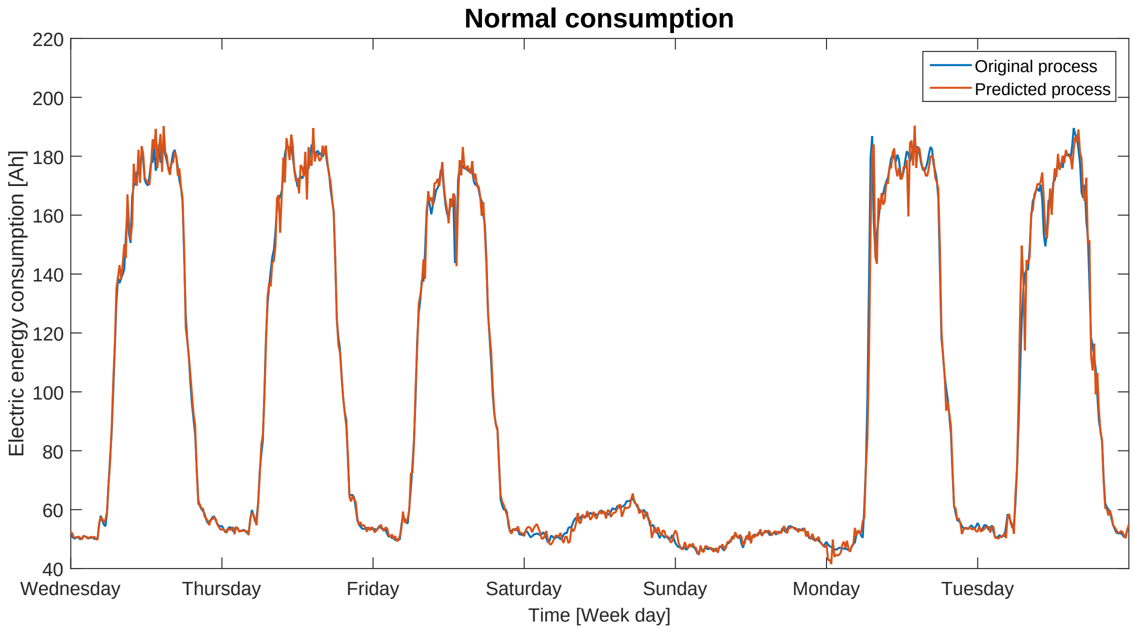


Figure 5.24: Comparison between the original and predicted $(m = 1)$ normal consumption process throughout a week - Multiple models and parameter tunning.

Figures 5.26 and 5.27 depict the sum of the normal and HVAC consumption processes over a week, and over the complete process, respectively.

It is visible that the predicted processes, in this case, present bigger variations around the original ones, when compared to predicted processes obtained using the simple multiple model approach.

The estimates of the variance of the prediction error, calculated using the estimator from equation (5.2), for the normal, HVAC, and total processes are

- Normal Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 25.3189,$
- HVAC Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 143.6848,$
- Total Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 199.3226.$

These results show that this approach with the parameter tunning does not present advantages when compared to the simple multiple models approach, in terms of variance of the prediction error.

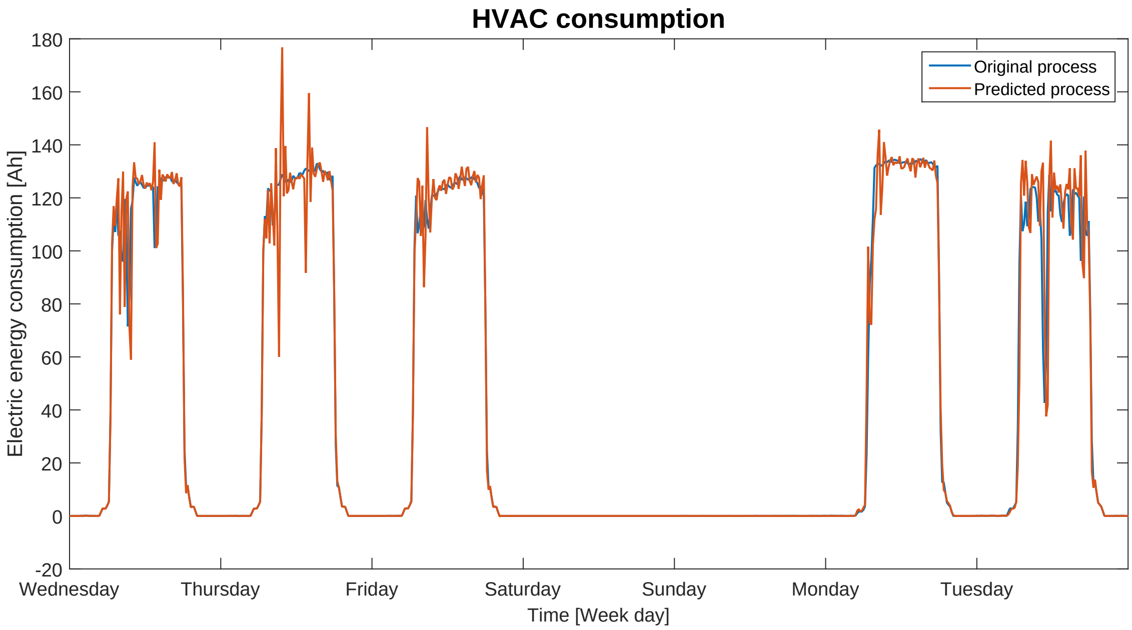


Figure 5.25: Comparison between the original and predicted $(m = 1)$ HVAC consumption process throughout a week -Multiple models and parameter tunning.

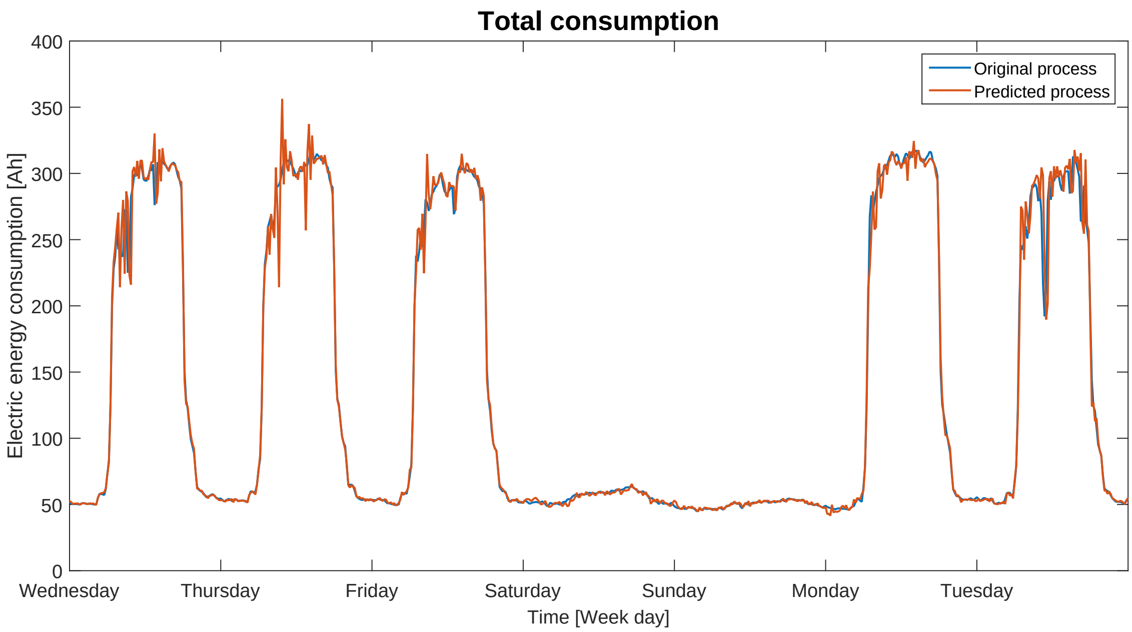


Figure 5.26: Comparison between the original and predicted $(m = 1)$ total consumption process throughout a week - Multiple models and R-ELS.

In fact, on top of being a more complicated algorithm, it provides worst results. This may be due to the fact that the models used for prediction changes very often.

5.6.2 Recursive Extended Least Squares - R-ELS

As mentioned in section 4.2, in order to use the Recursive extended least squares - R-ELS - method, there are some initial conditions that need to be computed. For each segment, the initial

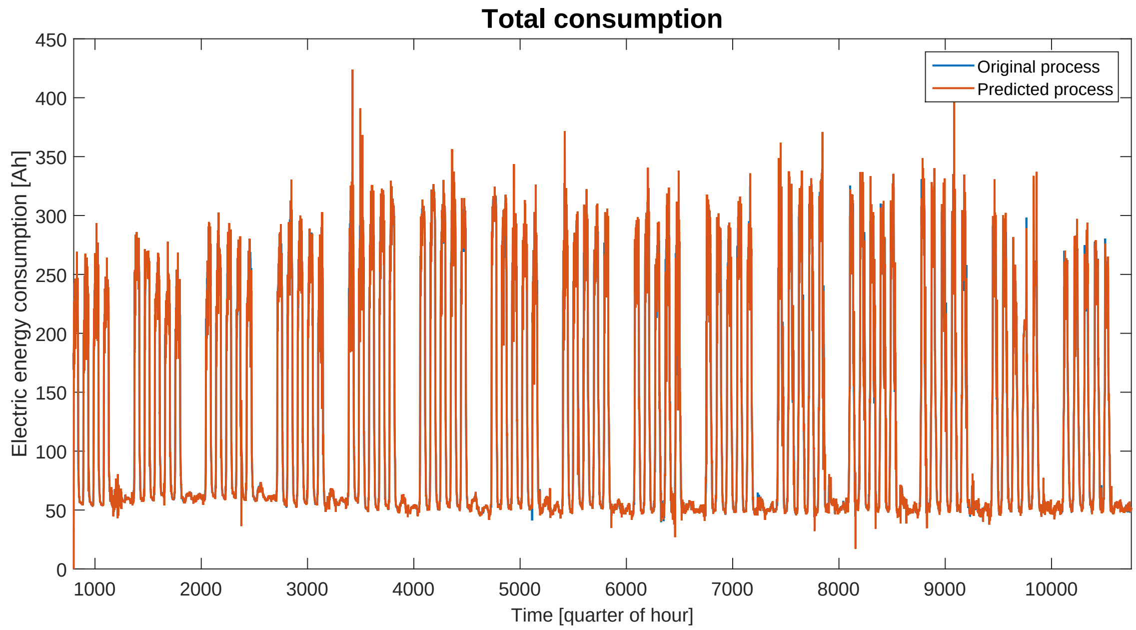


Figure 5.27: Comparison between the original and predicted ($m = 1$) total consumption process - Multiple models and R-ELS.

parameter estimates of $\boldsymbol{\theta}(k)$, are computed using the *armax* function. As for the initial values of $\varepsilon(k)$, these are computed by applying the inverse filter $(\frac{A^*(q^{-1})}{C^*(q^{-1})})$, to a set of M samples corresponding to the time instants in the interval $[k_0 - M, k_0]$, where k_0 is the time instant immediately before the one when prediction starts. The model is chosen based on the segment to which corresponds the time instant k_0. These values are computed for both normal and HVAC stationary consumption precesses.

The initialization of $P(k)$ for each segment is made by using the covariance matrices provided by the *armax* function when computing the estimates for the parameters for said segments. After some experiments were performed, the values of λ_{min} and of ε_0 for the normal and HVAC consumption processes are

$$\lambda_{min} = 0.97,$$

$$\varepsilon_0^{normal} = 0.1012,$$

$$\varepsilon_0^{hvac} = 0.1080.$$

Figures 5.28 and 5.29 show the normal and HVAC consumption predictions with $m = 1$, $\hat{y}_s(k+1|k)$, throughout a week, compared to the real values of the corresponding processes, when using the R-ELS method for adaptation.

Figures 5.30 and 5.31 depict the sum of the normal and HVAC consumption processes, real values and prediction, over a week, and over the complete process, respectively.

The predicted processes present also bigger variations around the original ones, when compared to predicted processes obtained using the simple multiple model approach, although not as accentuated as the variations that can be seen when using the multiple models with parameter tunning.

When using R-ELS together with the multiple models approach, the estimates of the variance of the prediction error, calculated using the estimator from equation (5.2), for the normal, HVAC, and

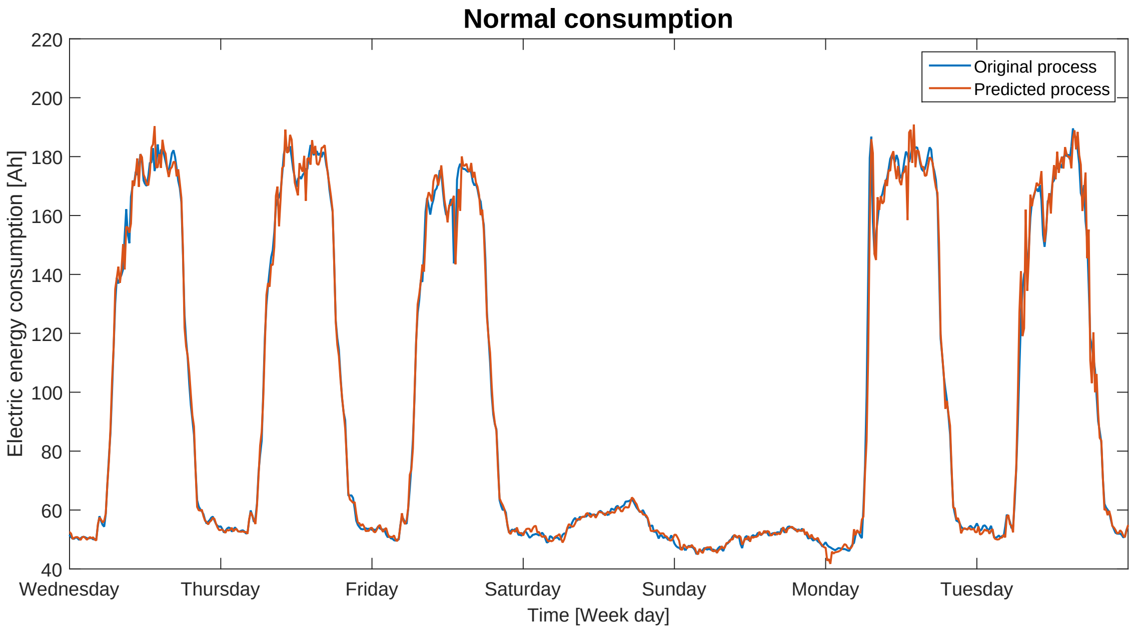


Figure 5.28: Comparison between the original and predicted ($m = 1$) normal consumption process throughout a week - Multiple models and R-ELS.

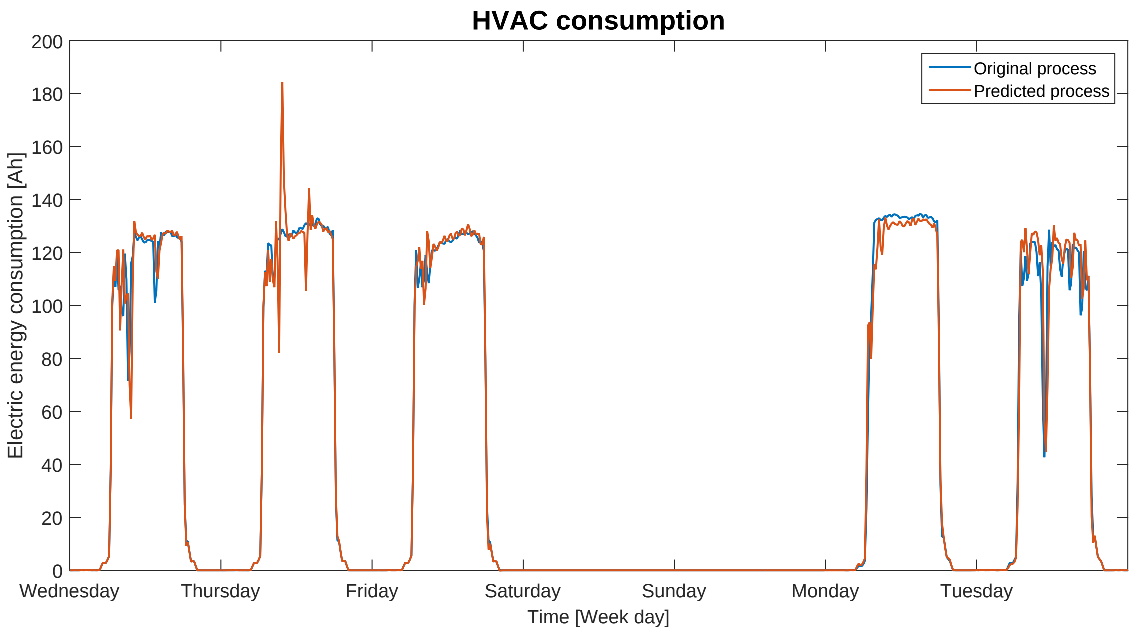


Figure 5.29: Comparison between the original and predicted ($m = 1$) HVAC consumption process throughout a week -Multiple models and R-ELS.

total processes are

- Normal Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 22.8705,$
- HVAC Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 70.1397,$
- Total Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 116.6917.$

From these results, it can be seen that this approach does not also present advantage when

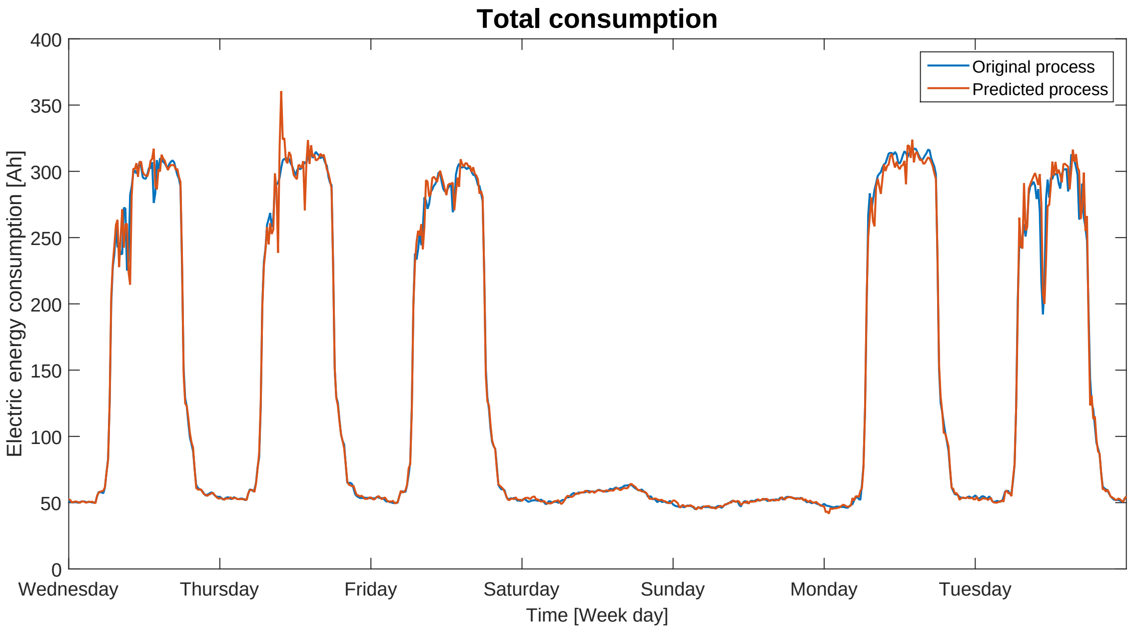


Figure 5.30: Comparison between the original and predicted $(m = 1)$ total consumption process throughout a week - Multiple models and R-ELS.

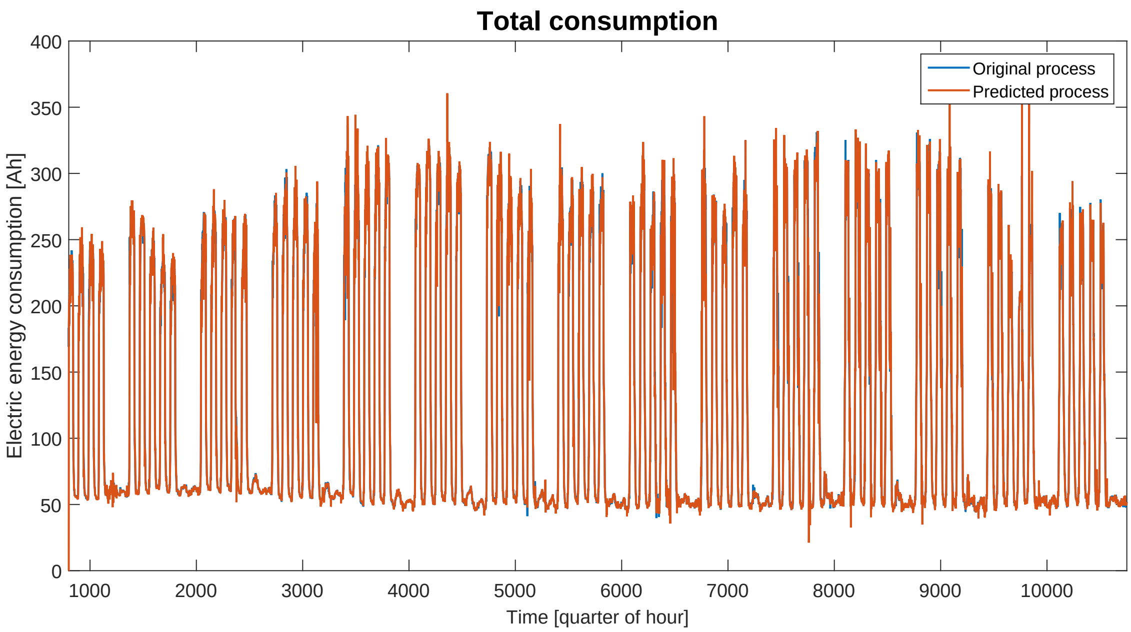


Figure 5.31: Comparison between the original and predicted $(m = 1)$ total consumption process - Multiple models and R-ELS.

compared with the simple multiple models approach. Nevertheless, when compared to the results obtained with the multiple models and parameter tunning approach, this one presents lower estimates of the variance of the prediction errors, making it more advantageous.

5.7 Adaptive prediction architecture

Finally, consider the architecture presented in section 4.3.

The first step is to determine the number of parameters, n_a and n_c, that describe both the normal and HVAC stationary consumption processes. The chosen values are

- Normal: $n_a = 7$ $n_c = 1$,
- HVAC: $n_a = 3$ $n_c = 2$.

The reasoning behind the decision to use these values was to consider all the parameter numbers for all the segments that were used in the architectures that were based on multiple models, and choose the maximum value from those sets os values. This way, there are enough parameters to describe each segment.

In this approach, the initial parameter estimates of $\boldsymbol{\theta}(k)$, and $\varepsilon(k)$, for both the normal and HVAC stationary consumption processes, are computed using the extended least squares method (without recursion) applied to a set of M samples corresponding to the time instants in the interval $[k_0 - M, k_0]$, where k_0 is the time instant immediately before the one when prediction starts.

In this case, the covariance matrices $P(k)$, for the normal and HVAC parts of the consumption process, are initialized as diagonal matrices of size $(n_a + n_c) \times (n_a + n_c)$, and the values of the diagonal are 0.1. These values were chosen after conducting some experiments.

As mentioned in section 4.3, when adding the seasonality to the processes, there is the need to have some form of segmentation. But in this case there are only three segments that will be considered, week days, Saturdays, and Sundays. It is assumed that any day can fits in one of these categories.

Similarly, based on some experiments were performed, the chosen values of λ_{min}, and of ε_0 for the normal and HVAC consumption processes are

$$\lambda_{min} = 0.97,$$

$$\varepsilon_0^{normal} = 0.0022,$$

$$\varepsilon_0^{hvac} = 0.0254.$$

Similarly to the previous approaches, Figures 5.32 and 5.33 show the normal and HVAC consumption predictions with $m = 1$, $\hat{y}_s(k+1|k)$, throughout a week, compared to the real values of the corresponding processes, when using the R-ELS without re-initialization.

Figures 5.34 and 5.35 show the sum of the normal and HVAC consumption processes, real values and prediction, over a week, and over the complete process, respectively.

A large peak can be seen in Figure 5.32. One possible explanation is that a significant change in the system parameters occurred, and the estimates did not converge rapidly enough to the new values.

When using R-ELS, the estimates of the variance of the prediction error, calculated using the estimator from equation (5.2), for the normal, HVAC, and total processes are

- Normal Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 28.5345$,

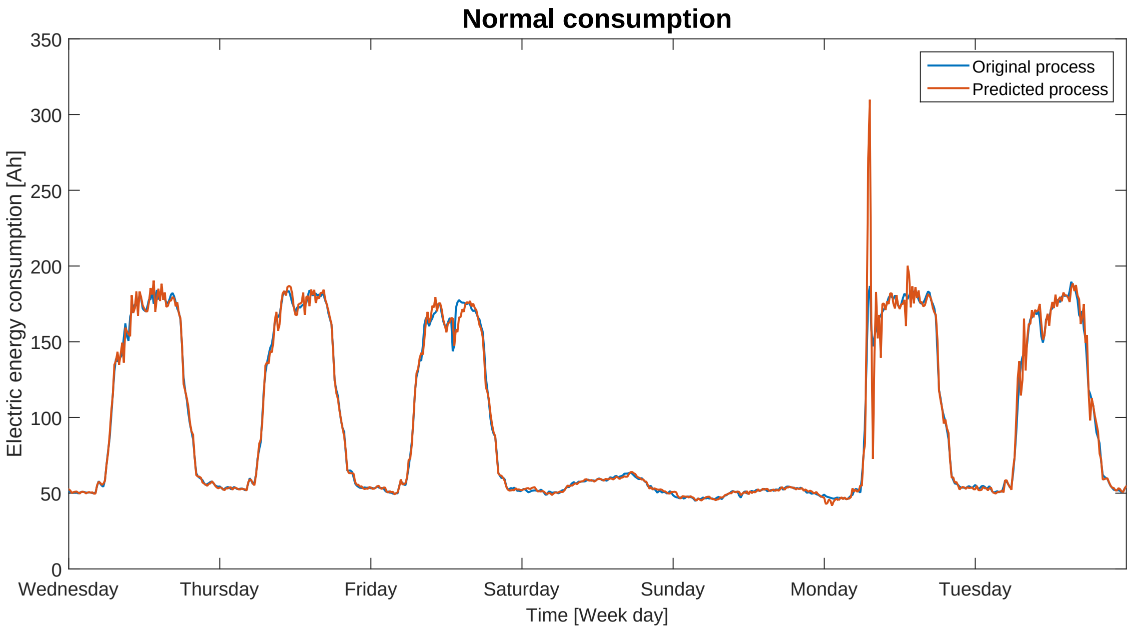


Figure 5.32: Comparison between the original and predicted ($m = 1$) normal consumption process throughout a week - R-ELS.

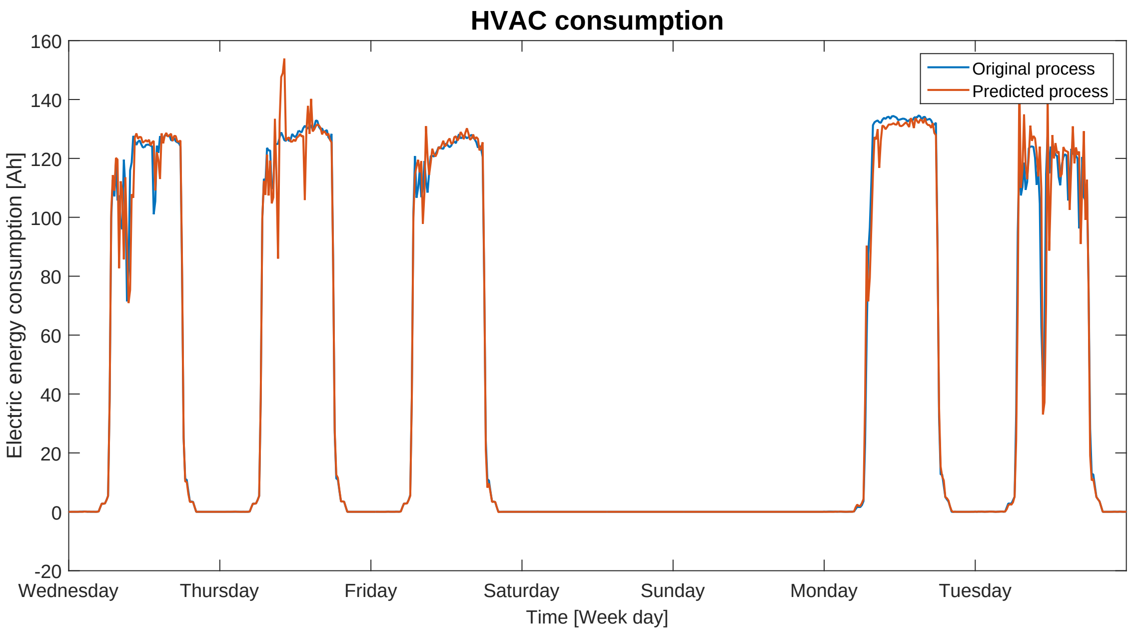


Figure 5.33: Comparison between the original and predicted ($m = 1$) HVAC consumption process throughout a week - R-ELS.

- HVAC Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 60.3301$,
- Total Consumption $\hat{E}[(\tilde{y}_s(k+1|k))^2|O^k] = 111.4326$.

From this results, it can be seen that this approach also does not present advantage when compared with the simple multiple models approach, in terms of the variance of the prediction error of the processes. Despite that, one must take in consideration that in this case there is no need to estimate the parameters of the models that generate the HVAC and normal stationary processes for each

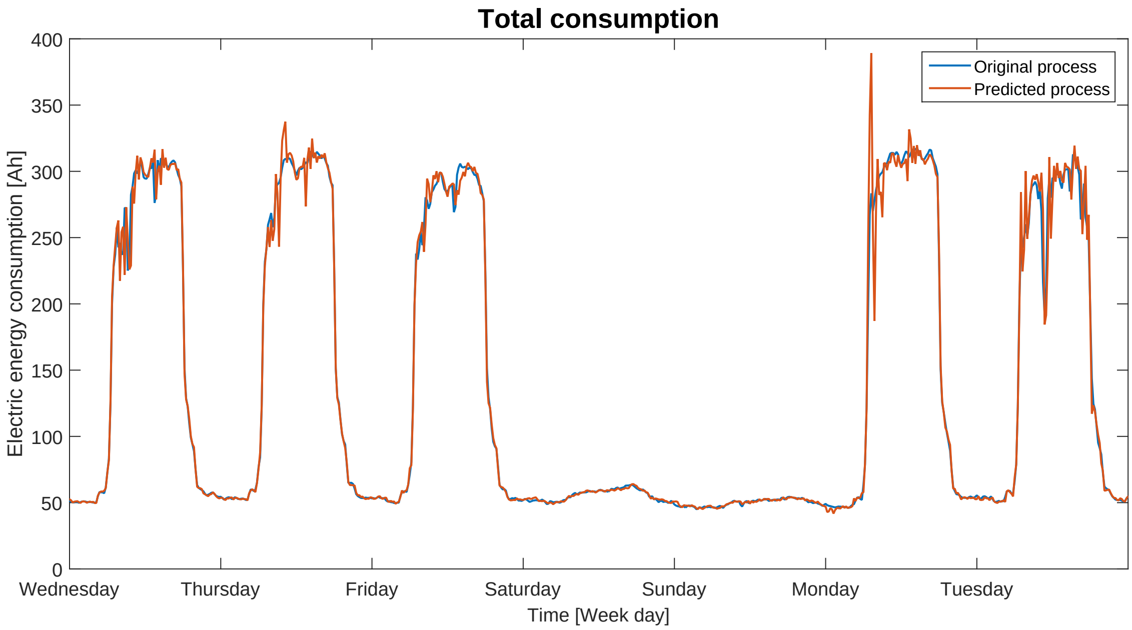


Figure 5.34: Comparison between the original and predicted $(m = 1)$ total consumption process throughout a week - R-ELS.

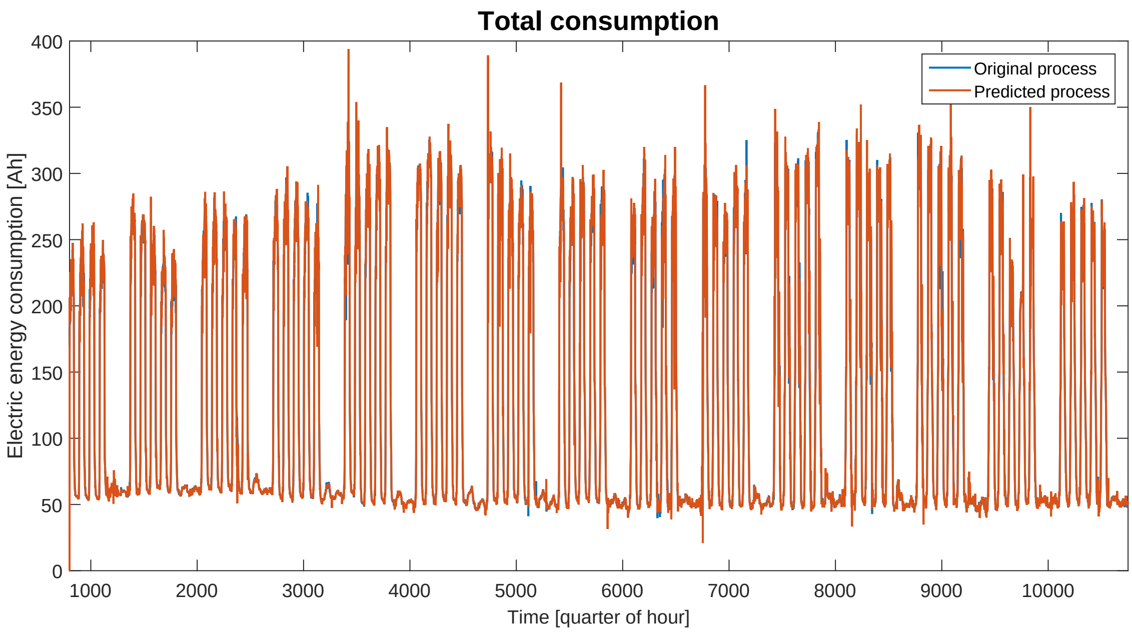


Figure 5.35: Comparison between the original and predicted $(m = 1)$ total consumption process - R-ELS.

segment, there is only the need of classifying a time instant as belonging to a week day, Saturday or Sunday. This comes as an advantage when there is the need to predict in a period that is not contemplated in the identified segments (for example, a segment during the winter).

6

Conclusions and Future Work

The purpose of this dissertation is to implement strategies in order to make predictions of the electrical energy consumption at the IST campus in Alameda by applying temporal series methods. For the practical purpose of this work, the data relative to the electric energy consumption of the North Tower is used.

These strategies include the processing of the data, model identification, and prediction.

The data processing phase includes the segmentation of the data for later identification, separation into normal and HVAC consumption processed, removal of the outliers and seasonal differencing. The data is segmented into four segments, summer vacation, autumn school time, Saturdays, and Sundays. It was decided that only the normal consumption processes are to be treated regarding the outliers, for it was considered that all the sudden variations in the HVAC consumption processes are justified. After this treatment the seasonality is removed from all the processes, except the HVAC consumption process, which is assumed to be already stationary, given that only on certain Saturdays, the HVAC was turned on.

After the data is processed, the parameters of the models for the HVAC and normal processes for each segment are estimated. The first step was is figure out how many parameters are necessary to describe the model. Once that task is performed, the parameters of the models are identified using a prediction error method.

The following step consists on computing the predictions. In order to compare one solution to another, a quadratic criterium, which is an estimator of the variance of the prediction error, is used to determine which is best. The predictions are calculated for the stationary processes, then, the seasonality is re-added to those processes, and the normal and HVAC components are summed in order to obtain the total predicted process for each segment. A study is performed in order to compute the variance of the prediction error for a prediction horizon equal to 1, and the results are compliant with what was expected. A second study is conducted regarding the advantage of identifying models for stationary processes as opposed to just say that the predicted value at a certain time instant is equal to the value of the process 96 samples ago (a day ago), and the first approach is more advantageous considering that the variance of the prediction error is significantly lower. To conclude the prediction phase, when it comes to prediction within each segment, the influence of the prediction horizon on the predicted processes is evaluated. As expected, the results show that the variance of the prediction error increases as the prediction horizon increases. During this last study, it is visible that initial conditions required when applying the prediction algorithm have great influence on highly oscillatory systems.

As an attempt to incorporate exogenous inputs into the predictions, namely outdoors temperature and solar radiation, two approaches were considered. The first one involved the estimation of the parameters of an ARMAX model. However, this first approach leads to identified models that were too complex. As a second attempt to incorporate exogenous inputs in the models, the prediction error process, that was obtained assuming an ARMA model, was used to estimate some parameters that might describe how these inputs could influence the consumption process. It was verified that the variance of the prediction error remained the same after updating the model with these newly identified

parameters. This occurrence led to the thought that the influence of these exogenous inputs may be incorporated in the parameters of the ARMA model.

After performing these experiments on separate segments, four different architectures were implemented and compared, and their purpose is to make a prediction of the complete process.

The first architecture is based on multiple models, and the goal is to chose the model that leads to a lowest prediction error for each time instant. The second and third architectures are slight variation of the first one, which are meant to provide adaptation within each model. There is adaptation using parameter tunning, in which one still considers that there are four main models and takes the parameters of each one of the models and varies them slightly, and chooses the model which leads to a lowest prediction error, and adaptation using the R-ELS method, in which every time the segment changes, the parameters need to be initialized. The last implemented architecture does not assume different segments for the stationary processes *a priori*, as it uses the R-ELS method to adapt to the circumstances, although there is the need to consider segments when re-adding the seasonality to the stationary processes.

Comparing all the architectures, it is concluded that the multiple model architectures combined with adaptation does not present any advantage when compared with the simple multiple models approach in terms of the variance of the prediction error for each case and complexity of the implementation. The one with parameter tunning provides significantly worst results when comparing the variance of the prediction error, and the one that uses the R-ELS has similar results to the simpler version, when comparing, as well, the variance of the prediction error. As for comparing the simple multiple model approach with the one that relies solely on adaptation to predict the stationary processes, when comparing the variance of the prediction error for both cases, neither one presents a clear advantage toward the other. Despite that, the second approach is considered more advantageous because it could adapt to a segment never seen before, whereas in the first approach, there would be a need to analyze the new data in order to find a new segment.

As for future work, it would be interesting to see if one could find a way to incorporate, in an explicit way, the temperature and radiation in the prediction to try to get better results, maybe even using weather forecasts. Another idea would be utilizing other identification methods, such as Maximum Likelihood or LASSO (least absolute shrinkage and selection operator), and even trying other predictors, to work around the problem of having systems with poles on the unit circle. It would also be interesting for the approaches presented in this dissertation were to be applied to other buildings of the campus, there may be different characteristics inherent to each one of the buildings, and that could reflect on the results of prediction.

Bibliografia

[1] G. H. Brundtland, "Our Common Future: Report of the World Commission on Environment and Development," *Medicine, Conflict and Survival*, vol. 4, no. 1, p. 300, 1987.

[2] N. Amjady, "Short-Term Hourly Load Forecasting Using Time-Series Modeling With Peak Load Estimation," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 798–805, 2001.

[3] J. M. Hugo Costa, Cristiano Cabrita and J. F. L. C. Semião, "Mecanismo de deteção de consumos anómalos em redes energéticas inteligentes," in *Conferência sobre Redes de Computadores (CRC)*, vol. 3, no. 1, Nov. 2015, pp. 1–6.

[4] A. Pawlowski, J. L. Guzman, F. Rodríguez, M. Berenguel, and J. Sanchez, "Application of time-series methods to disturbance estimation in predictive control problems," *IEEE International Symposium on Industrial Electronics*, pp. 409–414, 2010.

[5] Z. J. Hou, P. V. Etingov, Y. V. Makarov, and N. A. Samaan, "Uncertainty Reduction in Power Generation Forecast Using Coupled Wavelet-ARIMA," *PES General Meeting — Conference & Exposition, 2014 IEEE*, pp. 409–414, 2014.

[6] L. A. Moraes, R. A. Flauzino, M. A. Araujo, and O. E. Batista, "Development of a methodology to forecast time series using few input variables," *2013 IEEE PES Conference on Innovative Smart Grid Technologies (ISGT Latin America)*, pp. 1–4, 2013.

[7] E. Dotzauer, "Simple model for prediction of loads in district-heating systems," *Applied Energy*, vol. 73, pp. 277–284, 2002.

[8] J. Liu, "Application of the Grey Theory and the Neural Network in Water Demand Forecast," *2010 Sixth International Conference on Natural Computation*, pp. 1070–1073, 2010.

[9] S. Fan and W.-j. Lee, "Probabilistic Short-Term Wind Power Forecast Using Componential Sparse Bayesian Learning," *IEEE Transactions on Industry Applications*, vol. 49, no. 6, pp. 2783–2792, 2012.

[10] I. Sánchez, "Short-term prediction of wind energy production," *International Journal of Forecasting*, vol. 22, no. 1, pp. 43–56, 2006.

[11] S. Gao, Y. He, and H. Chen, "Wind speed forecast for wind farms based on ARMA-ARCH model," *2009 International Conference on Sustainable Power Generation and Supply*, pp. 1–4, 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5348142

[12] S. Zhu, M. Yang, M. Liu, and W.-J. Lee, “One parametric approach for short-term jpdf forecast of wind generation,” *IEEE Transactions on Industry Applications*, vol. 50, no. 4, pp. 2837–2843, Oct 2013.

[13] B. Xia and C. Zhao, “The Application of Multiple Regression Analysis Forecast in Economical Forecast: The Demand Forecast of Our Country Industry Lavation Machinery in the Year of 2008 and 2009,” *2009 Second International Workshop on Knowledge Discovery and Data Mining*, pp. 405–408, 2009.

[14] W.-M. Li, J.-W. Liu, J.-J. Le, and X.-R. Wang, “The financial time series forecasting based on proposed arma-grnn model,” *2005 International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 2005–2009, Aug 2005.

[15] X. Wang and Z. Han, “A new disaster monitor and forecast system based on rbf neural networks,” *Proceedings of the 2010 International Conference on Electrical and Control Engineering*, pp. 132–136, 2010.

[16] A. Li, X. Wang, H. Xu, and S. Li, “Research on combination forecast method of instrument precision,” *9th International Conference on Electronic Measurement Instruments, 2009.*, pp. 1–264–1–267, Aug 2009.

[17] A. K. Singh, Ibraheem, S. Khatoon, M. Muazzam, and D. K. Chaturvedi, “Load forecasting techniques and methodologies: A review,” *ICPCES 2012 - 2012 2nd International Conference on Power, Control and Embedded Systems*, pp. 1–10, 2012.

[18] S. Wei and H. Qun, “Research on network data forecast system based on non-linear time series,” *International Conference on Test and Measurement, 2009.*, vol. 1, pp. 251–254, 2009.

[19] G. E. P. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*. Wiley, 2015.

[20] K. J. Åström and B. Wittenmark, *Computer-controlled Systems: Theory and Design(3rd Ed.)*. Prentice-Hall, Inc., 1997.

[21] L. Ljung, *System Identification - Theory for the User*. Prentice-Hall, 1987.

[22] S. P. Sanoff and P. E. Wellstead, “Comments on: ’implementation of self-tuning regulators with variable forgetting factors’,” *Automatica*, vol. 19, no. 3, pp. 345–346, May 1983.