Prediction of Electrical Energy Consumption at IST

Sofia Fidalgo da Silva

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 error 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.

1. INTRODUCTION

Power system management is an important issue, given that, in a power grid that distributes energy at the national level, a balance between production and demand must be kept, 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. 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.

Forecasting, has been widely used in several areas, such as forecasts of electrical power consumption and generation (Amjady (2001)) (Hugo Costa and Semião (2015)) (Pawlowski et al. (2010)) (Hou et al. (2014)) (Moraes et al. (2013)), heat loads (Dotzauer (2002)), water demand (Liu (2010)), wind generation (Fan and Lee (2012)) (Sánchez (2006)) (Gao et al. (2009)) (Zhu et al. (2013)), and financial time series (Xia and Zhao (2009)) (Li et al. (2005)). Several different models have been utilized for prediction, including regression models (Xia and Zhao (2009)), single or multivariate, exponential smoothing Pawlowski et al. (2010), the ARMA model family (Amjady (2001)) (Hugo Costa and Semião (2015)) (Hou et al. (2014)) (Gao et al. (2009)) (Li et al. (2005)) (Li et al. (2009)) (Wei and Qun (2009)), and Support Vector Machines (Zhu et al. (2013)). There are more recent models being utilized, which are Neural Networks (Hugo Costa and Semião (2015)) (Liu (2010)) (Li et al. (2005)) (Li et al. (2009)), fuzzy logic (Moraes et al. (2013)) (Wang and Han (2010)) (Li et al. (2009)), and Grey Theory (Liu (2010)) (Li et al. (2009)). Some of these models have been combined in order to generate better forecasts (Li et al. (2005)) (Sánchez (2006)) (Wang and Han (2010)). behavior. For the parameter estimation of the model, for the most classic approaches, the Least Squares Method (Li et al. (2005)), and Maximum Likelihood (Hugo Costa and Semião (2015)) (Fan and Lee (2012)) (Gao et al. (2009)) (Wei and Qun (2009)) are the most used, for the AI approaches, machine learning methods are implemented (Hugo Costa and Semião (2015)) (Li et al. (2005)) (Li et al. (2009)). Finally, some ideas on data treatment among the revised literature include, for time series analysis, segmentation (Amjady (2001)), and also the use of the Discrete Wavelet Transform to obtain stationarity (Hou et al. (2014)) (Fan and Lee (2012)).

The prediction problem consists of the following: Given a sequence of observations of a time series, 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. 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*. To solve the problem of prediction of electric energy consumption at IST, one must consider the following steps. Establish a criterion to evaluate the predictions, which in this case is a quadratic criteria. Separate week days, weekends, holidays and vacation (it is also necessary to consider the HVAC part of the process separately from the rest, this matter will be later discussed in

Section 2), and remove the daily seasonality of the processes to obtain stationary processes. Then, the Box-Jenkins (Box and Jenkins (2015)) method, which uses ARMA models is used, and finally, after obtaining the estimates of the models, the predictions are computed.

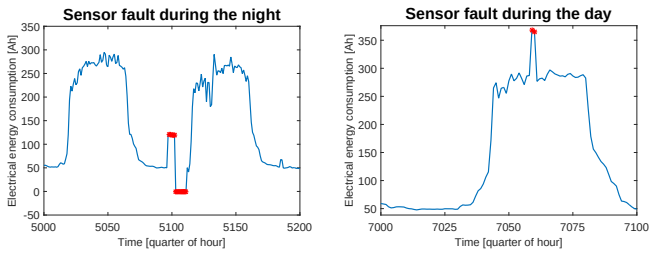
In this dissertation, the ARMA family models are used to describe the systems that generate the electric consumption processes. These models, that are to be used to make the predictions of the electrical energy consumption at IST, are identified by using traditional approaches, namely, a prediction error method and the least squares method. An adaptive strategy is also considered to estimate the parameters of the models, based on the recursive extended least squares method with exponential forgetting. Several prediction architectures are presented, which include a multiple models approach, an adaptive approach and some combinations of the previous two.

This document is organized as follows. After a brief introduction to the theme of this work, the data that is used for prediction purposes is presented in Section 2. This is followed, in Section 3, by the description of the methods that are used to solve the prediction problem. These methods are then integrated into some architectures that are presented afterwards in Section 4. Section 5 shows the experimental results that follow the implementation of said architectures. To finalize, some conclusions about the work performed and experimental results obtained are presented in Section 6.

2. AVAILABLE DATA

The acquired data files of the electric energy consumption are from the North Tower, one of the buildings of the Alameda Campus of IST. The data acquired is from August 2015 to March 2016. Also, the data provided is given in Ah. 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. 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.

Sensor faults can generate outliers in the process. Figure 1 shows probable sensor faults during the night, 1a and during the day, 1b. Besides this, there is a constant consumption (hereafter



(a) Night time sensor fault (b) Day time sensor fault

Fig. 1. Example of sensor faults

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 2 and 3 show some graphics that represent the data throughout several months, a week, and a day. In Figure 2 one

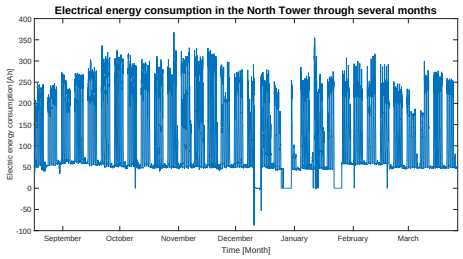


Fig. 2. Electric energy consumption in the North Tower throughout several months

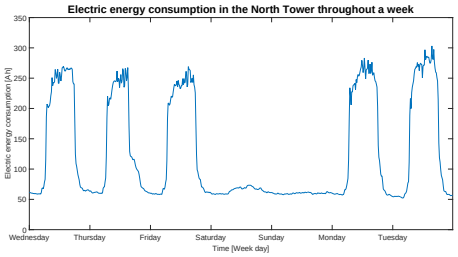


Fig. 3. Electric energy consumption in the North Tower throughout a week

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 3 shows also peaks and valleys that correspond to day and night time, and a considerable drop in consumption relative to the weekend.

Figures 4 and 5 represent the normal and HVAC consumption, respectively, throughout the week depicted in Figure 3. It

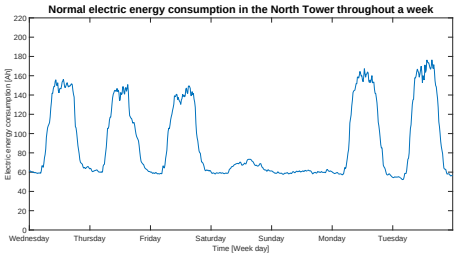


Fig. 4. Normal electric energy consumption in the North Tower throughout a week

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 and end of the school days.

3. METHODS

This section presents the proposed methods to address all the parts required to solve the prediction problem, namely, data processing, model identification and prediction methods.

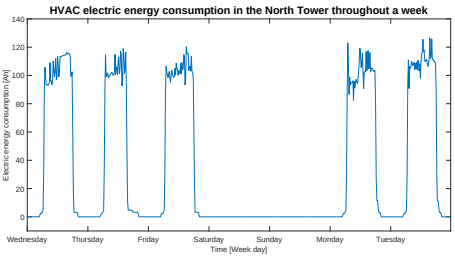


Fig. 5. HVAC electric energy consumption in the North Tower throughout a week

3.1 Data processing

In terms of data processing there is the need to deal with outliers, and seasonality. 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 (1)

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

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. 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 (2)

$$x(k) = \begin{cases} \frac{1}{M}\sum_{j=1}^{M}|x(k-j) - x(k-1-j)| \\ if\ x(k) - x(k-1) \geq 0 \\ \\ -\frac{1}{M}\sum_{j=1}^{M}|x(k-j) - x(k-1-j)| \\ if\ x(k) - x(k-1) < 0 \end{cases}. \qquad (2)$$

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.

A seasonal differencing filter (Box and Jenkins (2015)) 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

$$y(k) = (1 - q^{-T})x(k). \qquad (3)$$

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, the method to estimate the ideal number of parameters, p, is described.

This work assumes that the stationary systems are generated by ARMA models (Åström and Wittenmark (1997)). Given a stationary signal, $y(k)$, it can be seen as the output of a linear system to which is fed white noise (Åström and Wittenmark (1997)), $e(k)$, as input. Equation 4, depicts how $y(k)$ is generated

$$y(k) = -a_1 y(k-1) - \ldots - a_n y(k-n) + e(k) + c_1 e(k-1) + \\ + \ldots + c_n e(k-n) = \frac{C^*(q^{-1})}{A^*(q^{-1})}e(k), \qquad (4)$$

where the a_i and c_i coefficients are the quantities to be estimated, and $q-1$ is a backward shift operator. The degrees of the polynomials $A^*(q^{-1})$ and $C^*(q^{-1})$, are n_a and n_c, respectively.

The first method to be used is the one implemented in the already existing MATLAB function *armax* that is based on a prediction error method (Ljung (1987)). This function returns the identified model, sys, given a data set, x, and the values of n_a and n_c.

Another method to implement is a modified version of the Least squares method (Ljung (1987)). The regular Least squares method does not guarantee that the estimates are not biased when the noise is colored ($n_c \neq 0$), and it assumes that the model parameters are the same throughout the entire series. The objective is to have, not only, parameter estimates that are not biased, but also an algorithm that adapts to the changes in the system parameters throughout the series. In order to accomplish that, the equations for the Recursive extended least squares method are (Sanoff and Wellstead (1983)) (Ljung (1987))

$$\boldsymbol{\varepsilon}(k) = y(k) - \hat{\boldsymbol{\theta}}_{\boldsymbol{es}}^T(k-1)\boldsymbol{\varphi_{es}}(k), \qquad (5)$$

$$K(k) = \frac{P(k-1)\boldsymbol{\varphi_{es}}(k)}{\lambda(k-1) + \boldsymbol{\varphi}_{\boldsymbol{es}}^T(k)P(k-1)\boldsymbol{\varphi_{es}}(k)}, \qquad (6)$$

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{es}}(k) = \hat{\boldsymbol{\theta}}_{\boldsymbol{es}}(k-1) + K(k)\boldsymbol{\varepsilon}(k), \qquad (7)$$

$$\lambda(k) = 1 - [1 - \boldsymbol{\varphi}_{\boldsymbol{es}}^T(k)K(k)]\boldsymbol{\epsilon}^2(k)/\varepsilon_0, \qquad (8)$$

$$If\ \lambda(k) < \lambda_{min} \implies \lambda(k) = \lambda_{min}. \qquad (9)$$

$$P(k) = [I - K(k)\boldsymbol{\varphi}_{\boldsymbol{es}}^T(k)]P(k-1)/\lambda(k), \qquad (10)$$

where $\theta_{es}(k)$ denotes the vector of parameter estimates at a certain time instant k, $P(k)$ is the covariance matrix associated to said estimates, $K(k)$ is the Kalman gain associated to k, and $\lambda(k)$ is called the forgetting factor, $\lambda(k) \in]0,1[$, that gives less weight to past data preventing that, in case the parameters change, the data from the past influences te estimates so much that the estimates do not converge to the new values, that should not go below λ_{min}, in order to have a sufficient amount of information for identification. The estimates of the process $e(k)$ are denoted by $\varepsilon(k)$, ε_0 denotes the mean value of the prediction error, and the vector $\varphi_{es}(k)$ is given by

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

that is composed of known values, only.

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 equation (11),

$$n_a*, n_c* = \arg\min_{n_a, n_c} \frac{1}{M}\sum_{k \in test\ set}(y(k-1) - \hat{y}(k|k-1))^2, \qquad (11)$$

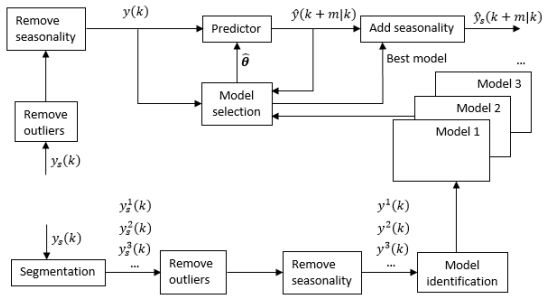


Fig. 6. Architecture for prediction using multiple models

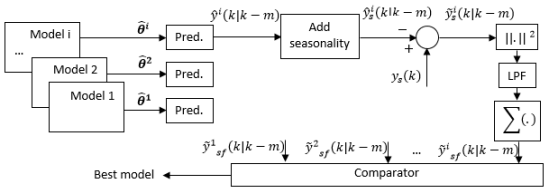


Fig. 7. Architecture for the model selection box from Figure 6

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. This method is used to avoid overfitting.

3.3 Prediction

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

$$J = E[(y(k+m) - \hat{y}(k+m|k))^2 | O^k] \tag{12}$$

be the variance of the prediction error in steady state, which is the value to minimize, where $E[\ ]$ denotes the expected value, O^k are the observations until time instant k (Åström and Wittenmark (1997)). The process $y(k+m)$ can be described in the following way,

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

where polynomials $G_m^*(q^{-1})$ and $F_m^*(q^{-1})$ result from the long division of polynomials $C^*(q^{-1})$ and $A^*(q^{-1})$. Substituting $\hat{y}(k+m|k)$ in equation (12),according to (Åström and Wittenmark (1997)), follows that the optimal predictor is given by

$$e(t) = \frac{A^*(q-1))}{C^*(q-1)}y(k), \tag{13}$$

and the variance of the prediction error can be written as

$$E[(y(k+m) - \hat{y}(k+m|k))^2 | O^k] = (1 + f_1^2 + \ldots + f_{m-1}^2)\sigma_e^2, \tag{14}$$

where σ_e^2 is the variance of the signal $e(k)$, and the $f_i, i = 1, \ldots, m-1$, terms are the coefficients of the polynomial $F_m^*(q^{-1})$.

4. ARCHITECTURES

4.1 Multiple models

The first proposed architecture is based on multiple models, as illustrated on Figure 6. 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, $\hat{\boldsymbol{\theta}}$, 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, from the available ones. After obtaining the predicted values of the stationary process, seasonality, must be added by using the inverse of the seasonal differencing filter that was used to remove it, and according to the chosen segments.

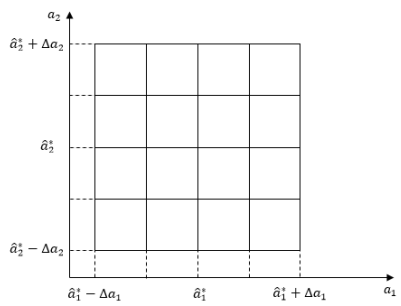


Fig. 8. Grid of possible variations of the main model, with estimated parameters a_1^* and a_2^*

Figure 7 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 $\tilde{y}_{sf}^i(k|k-m)$ goes above a certain threshold T. When that occurs, the model with the lower filtered prediction error is chosen. 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.

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 that are presented.

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 7, after the best main model is selected, the two approaches presented compute variations in said estimates for each main model.

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 8 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.

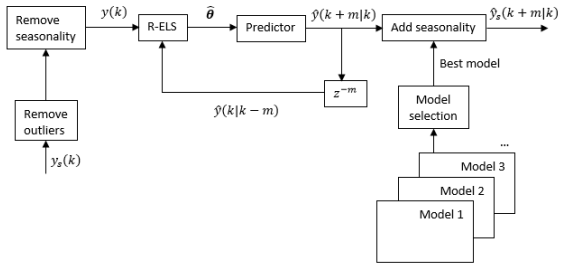


Fig. 9. Architecture using adaptive prediction - recursive extended least squares method

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. Equations (5) to (10) provide the necessary steps to compute the estimates. 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, they 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 7, 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.

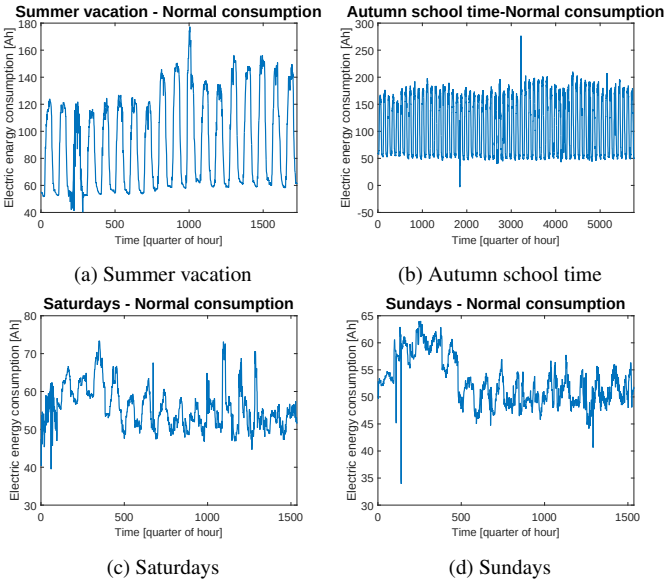
4.3 Adaptive prediction

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, based on the introduction presented in Section 3. 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. Finally, 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 7, 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

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

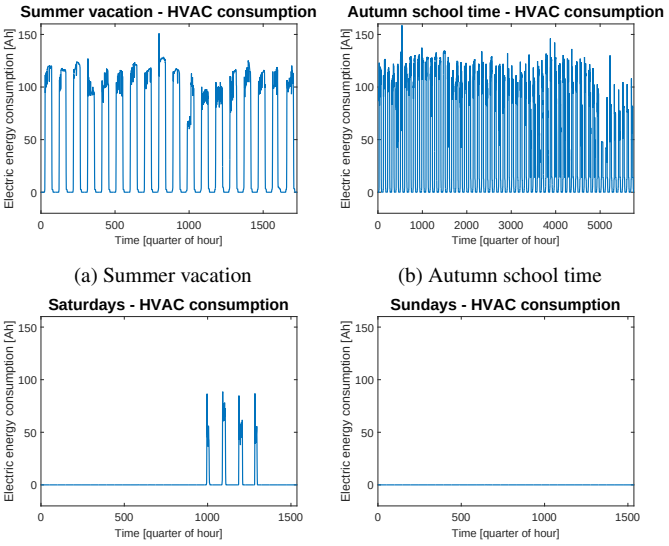
Four different segments are considered, summer vacation, autumn school time, Saturdays, and Sundays (with no differentiation between summer and autumn in the case of the latter two segments). The data corresponding to each segment is concatenated, for example, for week days, the end of a Friday is followed by the beginning of the next Monday. The normal and HVAC consumption data are separated, because it is assumed that systems that generate each part are different. Figures 10 and 11 show the normal and HVAC consumption processes, respectively. Accessing the remotion of outliers, for the normal



(a) Summer vacation (b) Autumn school time

(c) Saturdays (d) Sundays

Fig. 10. Normal consumption data for each segment



(a) Summer vacation (b) Autumn school time

(c) Saturdays (d) Sundays

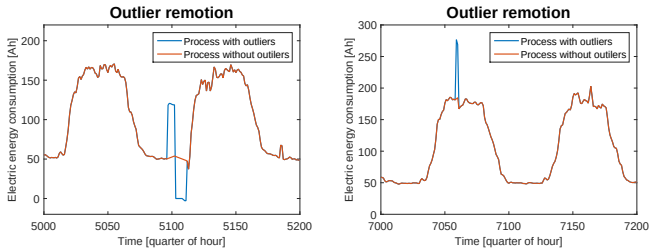
Fig. 11. HVAC consumption data for each segment

consumption process, the threshold (recall equation (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 12 shows some data segments in which outliers were detected, before and after they are removed. As for the removal



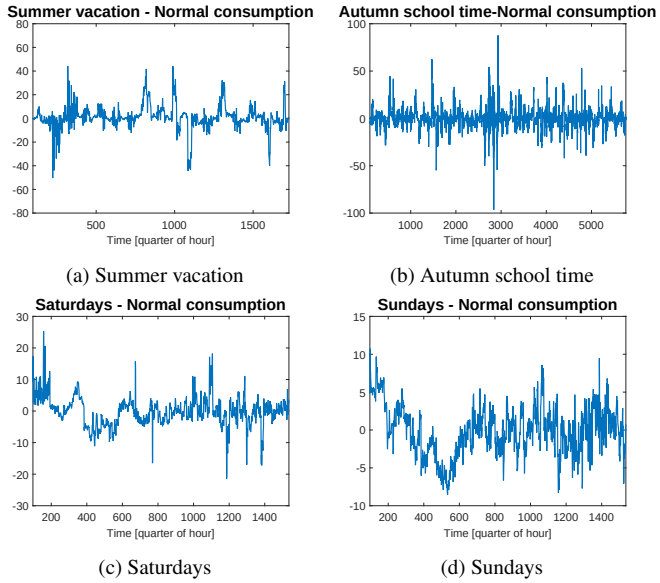
(a) Night time sensor fault (b) Day time sensor fault

Fig. 12. Comparison between segments of the processes $y_s^i(k)$ before and after outlier removal

of the daily seasonality, the data corresponding to each segment is passed through the filter presented in equation (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,

$$y(k) = (1 - q^{-96})y_s(k) \tag{15}$$

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 13 and 14 show the processes presented in Figures 10 and 11 without the seasonal part, here denoted as $y^i(k)$, result of the applying the filter from (15), with the exception for the Saturday and Sunday processes that correspond to the HVAC consumption because of the reasons mentioned above. After the data is processed, the

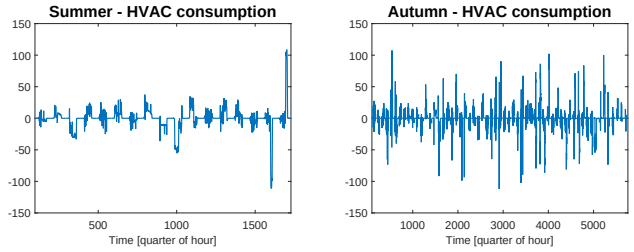


(a) Summer vacation (b) Autumn school time

(c) Saturdays (d) Sundays

Fig. 13. Normal consumption data without seasonality for each segment

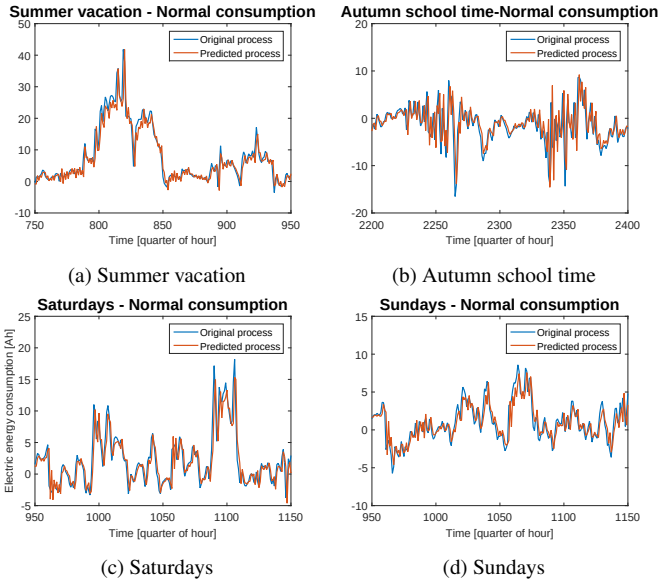
parameters of the models for each segment can be estimated



(a) Summer vacation (b) Autumn school time

Fig. 14. HVAC consumption data without seasonality for the appropriate segments

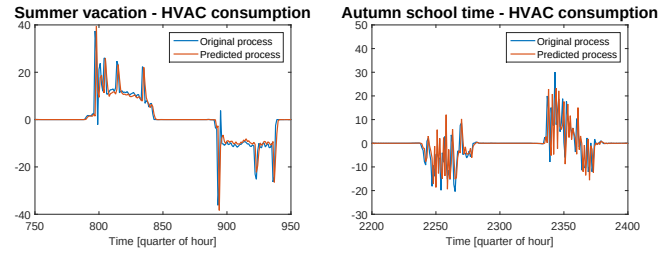
and used for prediction. Figures 15 and 16 show a comparison between the stationary processes, $y^i(k)$, from Figures 13 and 14, and the one step ahead ($m = 1$) predictions, $\hat{y}^i(k+1|k)$.



(a) Summer vacation (b) Autumn school time

(c) Saturdays (d) Sundays

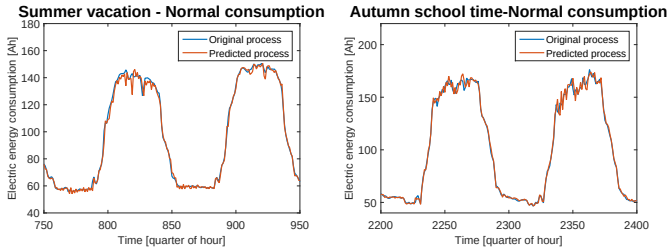
Fig. 15. Comparison between the original and predicted ($m = 1$) stationary normal consumption processes for each segment.



(a) Summer vacation (b) Autumn school time

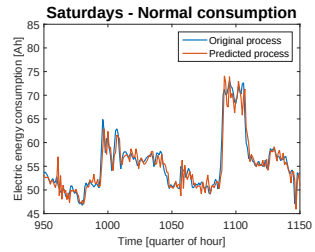
Fig. 16. Comparison between the original and predicted ($m = 1$) stationary HVAC consumption processes for each segment.

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. The prediction is calculated using the process from Figure 11c. The result is shown in Figures 17 and 18, for the normal and HVAC consumption, respectively. Recalling

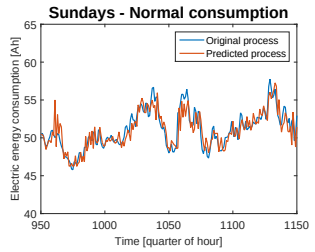


(a) Summer vacation

(b) Autumn school time

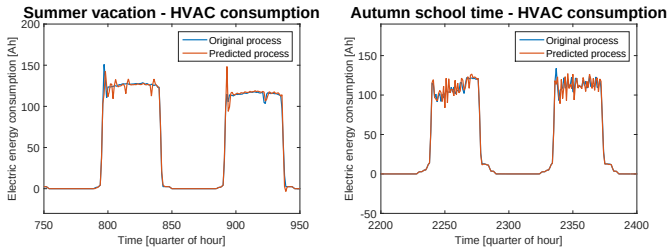


(c) Saturdays



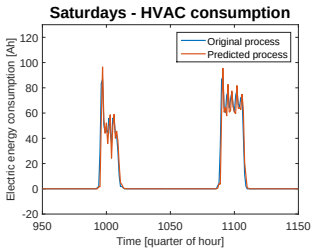
(d) Sundays

Fig. 17. Comparison between the original and predicted ($m = 1$) normal consumption processes for each segment.



(a) Summer vacation

(b) Autumn school time



(c) Saturdays

Fig. 18. Comparison between the original and predicted ($m = 1$) HVAC consumption processes for each segment.

equation (14), the variance the variance of the prediction error is given by

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

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

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

Theoretically, the variance of the prediction error is equal to the variance of $e(k)$. Table 1 presents a comparison between the estimates of the variance prediction error, $\hat{E}[(\tilde{y}_s^i(k+1|k))^2|O^k]$, and the estimates of the variance of $e(k)$, for each segment.The variance of the prediction error is calculated using the following estimator

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

and the estimator used to compute $\hat{\sigma}_e^2$ is

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

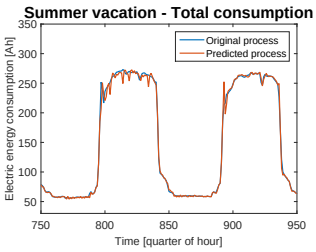
where $\varepsilon(k)$ are the estimates of the values of $e(k)$, given that they are not observable, computed using the inverse filters obtained in the model identification phase, and M is the number of samples of each process.

Table 1. Comparison between the estimates of the process $e(k)$ and the estimates of the variance of the prediction error for each segment ($m = 1$).

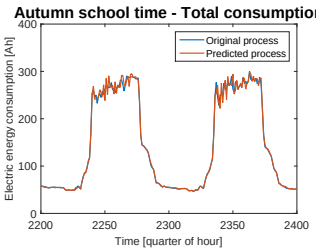
	$\hat{\sigma}_e^2$		$\hat{E}[(\tilde{y}_s^i(k+1\|k))^2\|O^k]$	
	Normal	HVAC	Normal	HVAC
Summer	14.9797	40.7040	15.0443	40.7040
Autumn	24.7887	55.3629	24.7879	55.3629
Saturday	3.1720	9.1049	3.1651	9.1049
Sunday	1.4212	–	1.4129	–

The presented estimates of the variance of the prediction error present a mean deviance of approximately 0.18%, maximum of a approximately 0.58%,and minimum of 0% from the corresponding estimates of the variance of $e(k)$. These estimates are in accordance with what was expected.

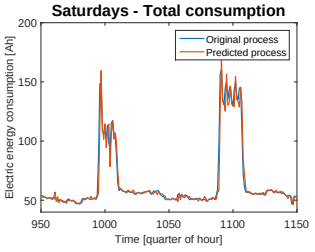
The final step for the prediction of each segment's process is to add the normal and HVAC consumption parts. The result is presented in Figure 19.



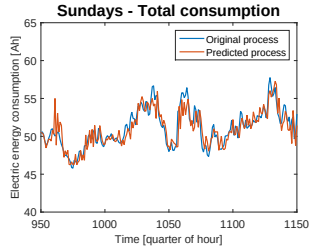
(a) Summer vacation



(b) Autumn school time



(c) Saturdays

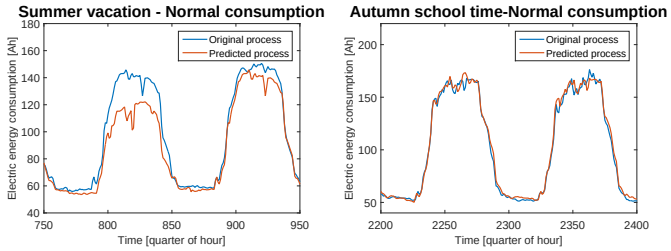


(d) Sundays

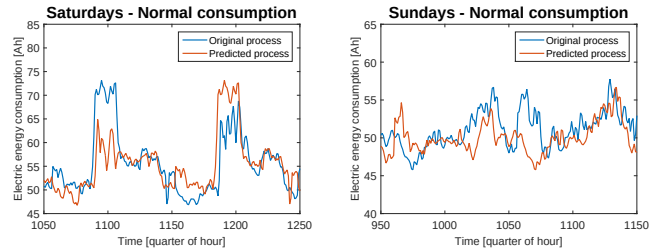
Fig. 19. Comparison between the original and predicted ($m = 1$) total consumption processes for each segment.

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 in Figures 20 and 21. The predictor used is given by,

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

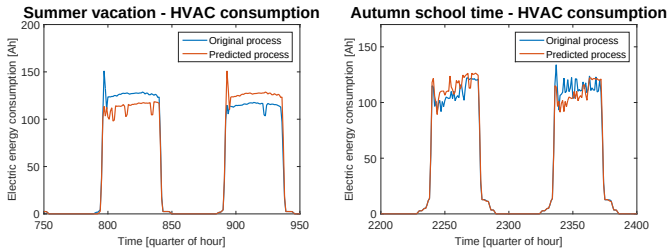


(a) Summer vacation (b) Autumn school time

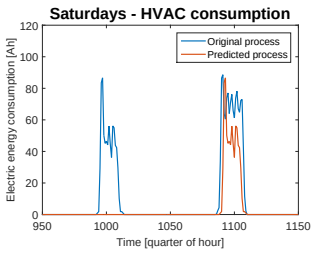


(c) Saturdays (d) Sundays

Fig. 20. Comparison between the original and predicted ($\hat{y}_s^i(k) = y_s^i(k-96)$) normal consumption processes for each segment.



(a) Summer vacation (b) Autumn school time



(c) Saturdays

Fig. 21. Comparison between the original and predicted ($\hat{y}_s^i(k) = y_s^i(k-96)$) HVAC consumption processes for each segment.

These figures show that the predicted processes do not follow the real ones as well, when compared to the predictions presented in Figures 17 and 18. As was done before, in Table 1, the estimates of the prediction error were calculated using the same estimator. The results are presented in Table 2.

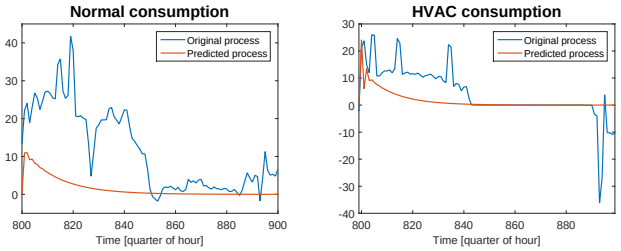
These values are significantly larger than the estimated variance of the prediction error calculated before, proving to be disadvantageous in comparison to the one step predictions calculated before.

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 22 shows the prediction and the real values of the stationary process for $m = 1, \ldots, 100$.

Table 2. Estimates of the variance prediction error for each segment ($\hat{y}_s^i(k) = y_s^i(k-96)$).

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

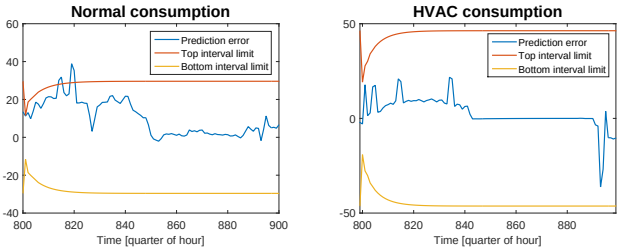
It can be seen that, as the prediction horizon increases, the



(a) Normal consumption (b) HVAC consumption

Fig. 22. Comparison between the original and predicted ($m = 1, \ldots, 100$) stationary consumption processes for the summer vacation segment

predicted values tend to zero. Figure 23 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 $\sigma_{pe}^2(m)$ is the estimate of the variance of the prediction as computed using equation (14), for a given prediction horizon, m. This interval illustrates how σ_{pe}^2 evolves as the prediction horizon increases. As expected, the



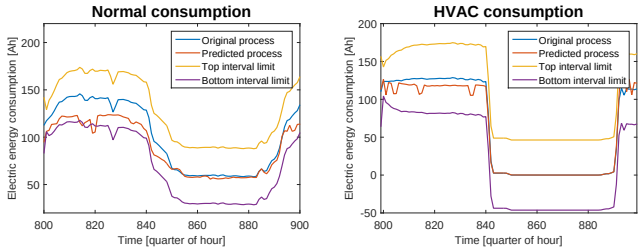
(a) Normal consumption (b) HVAC consumption

Fig. 23. Prediction error of the consumption processes for the summer vacation segment, for $m = 1, \ldots, 100$.

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 24 exhibits the original and predicted processes shown in Figure 22 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.

Next are presented the results of the implementation of the four integration architectures, the simple multiple models ar-



(a) Normal consumption (b) HVAC consumption

Fig. 24. Comparison between the original and predicted ($m = 1, \ldots, 100$) consumption processes for the summer vacation segment.

chitecture (MM), the multiple models approach with parameter tunning (MM+PT), and R-ELS (MM+R-ELS), and the solely adaptive approach (R-ELS).

To show an example of the visual representation of the results of applying these architectures, Figure 25 is presented. It depicts the total consumption predictions with $m = 1$, $\hat{y}_s(k+1|k)$ compared to the real process for the multiple models architecture. Figure 26 shows a comparison between the predicted segment

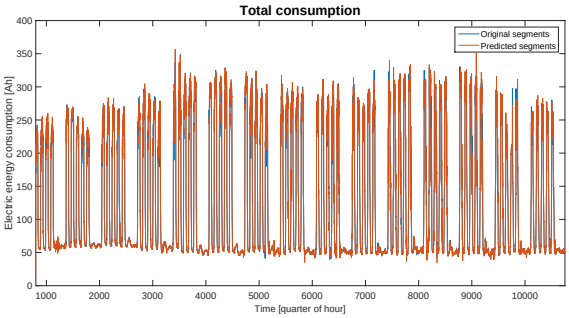


Fig. 25. Comparison between the original and predicted ($m = 1$) total consumption process - Multiple models.

for each time instant and the real correspondence. For this to be represented if a form of a function, each segment corresponds to a number, as follows

$$\text{Segment} = \begin{cases} 1 & \text{if summer vacation} \\ 2 & \text{if autumn school time} \\ 3 & \text{if saturday} \\ 4 & \text{if sunday.} \end{cases} \quad (16)$$

This applies to the architectures that are based on multiple models. As for the R-ELS approach, only three segments are considered, because of the adding seasonality phase, which are week days, Saturdays, and Sundays.

In Figure 26 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. An-

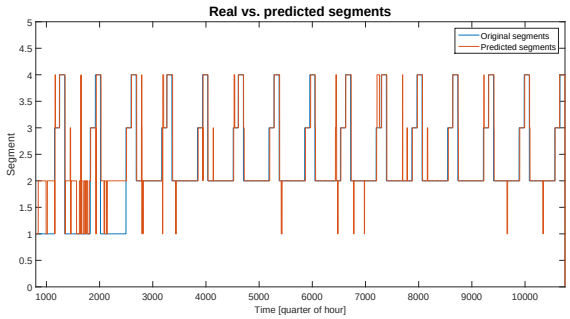


Fig. 26. Comparison between the original and predicted ($m = 1$) total consumption process - Multiple models.

other 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.

Table 3 shows a comparison between the estimates of the variance of the prediction error of the normal, HVAC and total processes, for all four architectures. From this results

Table 3. Comparison between the variance of the prediction error obtained with each of the architectures ($m = 1$).

	$\hat{E}[(\tilde{y}_s(k+1\|k))^2 \| O^k]$		
	Normal	HVAC	Total
MM	17.1761	60.9381	101.6131
MM+PT	25.3189	143.6848	199.3226
MM+R-ELS	22.8705	70.1397	116.6917
R-ELS	28.5345	60.3301	111.4326

it is verified that the architectures that combine the multiple models with adaptive approaches are not advantageous when compared with the simple multiple models approach, although, when comparing the results obtained with the MM+R-RELS with the MM+PT approach, the first one resulted in lower estimates of the variance of the prediction errors. As for the comparison of the MM and the R-ELS approaches, neither presents an advantage over the other, in terms of the variance of the prediction error of the processes. Despite that, one must take in consideration that, when using the R-ELS architecture, there is no need to estimate the parameters of the models that generate the HVAC and normal stationary processes for each 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

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

When performing the predictions for each segment separately, the results from studies on the variance of the prediction for $m = 1$ are compliant with what was expected. 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), is proven to be more advantageous considering that the variance of the prediction error is significantly lower. It was also shown that the variance of the prediction error increases as the prediction horizon increases.

During the course of this work, as an attempt to incorporate exogenous inputs into the predictions, 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.

Comparing all the integration architectures, it is concluded that the multiple model approach 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 towards 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.

REFERENCES

Amjady, N. (2001). Short-Term Hourly Load Forecasting Using Time-Series Modeling With Peak Load Estimation. *IEEE Transactions on Power Systems*, 16(4), 798–805.

Box, G.E.P. and Jenkins, G. (2015). *Time Series Analysis: Forecasting and Control*. Wiley.

Dotzauer, E. (2002). Simple model for prediction of loads in district-heating systems. *Applied Energy*, 73, 277–284.

Fan, S. and Lee, W.j. (2012). Probabilistic Short-Term Wind Power Forecast Using Componential Sparse Bayesian Learning. *IEEE Transactions on Industry Applications*, 49(6), 2783–2792.

Gao, S., He, Y., and Chen, H. (2009). Wind speed forecast for wind farms based on ARMA-ARCH model. *2009 International Conference on Sustainable Power Generation and Supply*, 1–4. doi:10.1109/SUPERGEN.2009.5348142.

Hou, Z.J., Etingov, P.V., Makarov, Y.V., and Samaan, N.A. (2014). Uncertainty Reduction in Power Generation Forecast Using Coupled Wavelet-ARIMA. *PES General Meeting — Conference & Exposition, 2014 IEEE*, 409–414. doi: 10.1109/PESGM.2014.6939528.

Hugo Costa, Cristiano Cabrita, J.M. and Semião, J.F.L.C. (2015). Mecanismo de deteção de consumos anómalos em redes energéticas inteligentes. In *Conferência sobre Redes de Computadores (CRC)*, volume 3, 1–6.

Li, A., Wang, X., Xu, H., and Li, S. (2009). Research on combination forecast method of instrument precision. *9th International Conference on Electronic Measurement Instruments, 2009.*, 1–264–1–267. doi:10.1109/ICEMI.2009.5274878.

Li, W.M., Liu, J.W., Le, J.J., and Wang, X.R. (2005). The financial time series forecasting based on proposed arma-grnn model. *2005 International Conference on Machine Learning and Cybernetics*, 4, 2005–2009. doi: 10.1109/ICMLC.2005.1527274.

Liu, J. (2010). Application of the Grey Theory and the Neural Network in Water Demand Forecast. *2010 Sixth International Conference on Natural Computation*, 1070–1073.

Ljung, L. (1987). *System Identification - Theory for the User*. Prentice-Hall.

Moraes, L.A., Flauzino, R.A., Araujo, M.A., and Batista, O.E. (2013). Development of a methodology to forecast time series using few input variables. *2013 IEEE PES Conference on Innovative Smart Grid Technologies (ISGT Latin America)*, 1–4. doi:10.1109/ISGT-LA.2013.6554376.

Pawlowski, A., Guzman, J.L., Rodríguez, F., Berenguel, M., and Sanchez, J. (2010). Application of time-series methods to disturbance estimation in predictive control problems. *IEEE International Symposium on Industrial Electronics*, 409–414. doi:10.1109/ISIE.2010.5637867.

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

Sánchez, I. (2006). Short-term prediction of wind energy production. *International Journal of Forecasting*, 22(1), 43–56. doi:10.1016/j.ijforecast.2005.05.003.

Sanoff, S.P. and Wellstead, P.E. (1983). Comments on: 'implementation of self-tuning regulators with variable forgetting factors'. *Automatica*, 19(3), 345–346.

Wang, X. and Han, Z. (2010). A new disaster monitor and forecast system based on rbf neural networks. *Proceedings of the 2010 International Conference on Electrical and Control Engineering*, 132–136. doi:10.1109/iCECE.2010.41.

Wei, S. and Qun, H. (2009). Research on network data forecast system based on non-linear time series. *International Conference on Test and Measurement, 2009.*, 1, 251–254. doi: 10.1109/ICTM.2009.5412948.

Xia, B. and Zhao, C. (2009). 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*, 405–408. doi:10.1109/WKDD.2009.50.

Zhu, S., Yang, M., Liu, M., and Lee, W.J. (2013). One parametric approach for short-term jpdf forecast of wind generation. *IEEE Transactions on Industry Applications*, 50(4), 2837–2843. doi:10.1109/IAS.2013.6682487.