Warm-Start Meta-Ensembles for Forecasting Energy Consumption in Service Buildings*

Pedro J.S. Cardoso[1,2][0000−0003−4803−7964], Pedro M.M. Guerreiro[2][0000−0001−7272−2000], Jânio Monteiro[2,3][0000−0002−4203−1679], André S. Pedro[2][0000−0002−8359−1322], and João M.F. Rodrigues[1,2][0000−0002−3562−6025]

[1] LARSyS, Universidade do Algarve, Faro, Portugal
[2] ISE, Universidade do Algarve, Faro, Portugal
[3] INESC-ID, Lisbon, Portugal
{pcardoso - pmguerre - jmmontei - aspedro - jrodrig}@ualg.pt

Abstract. Energy Management Systems are equipments that normally perform the individual supervision of power controllable loads. With the objective of reducing energy costs, those management decisions result from algorithms that select how the different working periods of equipment should be combined, taking into account the usage of the locally generated renewable energy, electricity tariffs etc., while complying with the restrictions imposed by users and electric circuits. Forecasting energy usage, as described in this paper, allows to optimize the management being a major asset.

This paper proposes and compares three new meta-methods for forecasts associated to real-valued time series, applied to the buildings energy consumption case, namely: a meta-method which uses a single regressor (called Sliding Regressor – SR), an ensemble of regressors with no memory of previous fittings (called Bagging Sliding Regressor – BSR), and a warm-start bagging meta-method (called Warm-start Bagging Sliding Regressor – WsBSR). The novelty of this framework is combination of the meta-methods, warm-start ensembles and time series in a forecast framework for energy consumption in buildings. Experimental tests done over data from an hotel show that, the best accuracy is obtained using the second method, though the last one has comparable results with less computational requirements.

Keywords: Warm-Start Ensembles · Meta Ensembles · Decision Tree Regressors · Energy Consumption Forecasting · Time Series

1 Introduction

In 2016, the European Union presented a package of measures with the aim of providing a stable legislative framework to facilitate the transition process to renewable energy. In this context, Regulation (EU) 2018/1999 [12], required that

* This work was supported by the Portuguese Foundation for Science and Technology (FCT), project LARSyS - FCT Project UIDB/50009/2020 and by the European Union, under the FEDER (Fundo Europeu de Desenvolvimento Regional) and INTERREG programs, in the scope of the T2UES (0517_TTUES_6_E) project.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

all member states should prepare and submit to the European Commission, a National Energy and Climate Plan (NECP), with a medium-term perspective (Horizon 2021-2030). E.g., Portugal's main goal in the NECP is to become neutral in greenhouse gas emissions by 2050, which requires to comply with trajectories that lead to a reduction in greenhouse gas emissions between 85 and 90 % by 2050. To achieve this goal, the greatest reduction in emissions will have to be achieved in the current decade, with decreases ranging between 45 and 55 %. One of the several goals defined by Portugal in its NECP [14], was that until 2030, 80 % of the electric energy that is consumed, should come from renewable energy sources. Such a high percentage of self-sufficiency cannot be achieved by solely generating more energy. It also requires consuming less energy and adapting the consumption pattern with the generation levels, based on Demand Response (DR) measures. A reduction in consumption can either be accomplished using deterministic or data-driven methods [10]. Deterministic methods are mainly used in the project phase of new buildings, allowing the proper design of the building structure and materials to be used. Data-driven methods are mostly used in functional buildings, identifying the normal consumption levels in existing infrastructures, allowing the identification of its consumption targets, or the definition of a set of levels for the evaluation of the consumption profile.

In terms of DR, Energy Management Systems (EMSs) [18,40] are the equipment that normally performs the individual supervision of shiftable/power controllable loads. Optimized management decisions result from algorithms that select how the different working periods of equipment should be combined, observing the generated energy, energy tariffs etc., while complying with the restrictions imposed by users and electric circuits [24]. E.g., decisions can be made using mathematical optimization, model predictive control or heuristic control, with several methods requiring a look into the future, i.e., forecasting energy generation and the building's consumption, before deciding how loads should be scheduled to work[6,23]. Similarly, the detection of anomalies can be exposed if the real consumption deviates from the one that was forecasted, and many situations exists in which minor changes in consumption can indicate a serious problem. In any case, it might be important to have a human to judge alarms, as unpredicted values can be normal due to naturally extraordinary events or anomalous due to machine failure, malfunctioning of a sensor etc. So, either when using DR or performing an assessment of the efficiency of buildings, a prediction of the consumption is a tool to help decision makers in theirs tasks.

In this context, large service buildings, such as hotels, shopping centers, hospitals, schools, offices, public buildings etc., have great variability in their consumption of resources, such as energy or water. Their energy consumption values depend on many variables (e.g., occupation, outside temperature, solar radiation, appliances settings or building occupation) making it very difficult to predict those values with precision. Having a human, for example a specialized engineer, to analyze each of these situations is obviously unaffordable from a technical and economic point of view. The solution is to develop an artificial intelligence system, that can learn over time what is considered normal and detect what is

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

abnormal for each building. Machine learning algorithms will make it possible to combine different variables such as time of day, day of week, temperature, radiation, occupation, and electricity consumption to predict the normal consumption of the building and the respective expected deviation. Distinct methods can be used to obtain this [37].

This paper proposes and compares three new meta-methods for the forecast associated to real-valued time series, namely, energy consumption in buildings. With a short and a full memory variant, the meta-methods will be supported on well known regression methods to implement a sliding window solution which will forecast the energy consumption of an hotel at certain instants. The first method, called Sliding Regressor – SR, is somehow a standard method which uses a single regressor given as a parameter, being fitted with the latest data just before forecasting is required. Similar in the fitting moment, the second method, called Bagging Sliding Regressor – BSR, uses an ensemble of regressors with no memory of previous fittings. The third method, called Warm-start Bagging Sliding Regressor – WsBSR, also uses an ensemble of regressors, however, it is distinct from the BSR by maintaining the previous regressors in memory. Preserving the ensemble idea, the latter allows to fit less regressors, a step with high computational cost, while using a broad number of regressors to make the forecasts. The use of the WsBSR is therefore a possibility as it achieves slightly worst results but with a fraction of the computational requirements. The paper's main contribution to the state of the art is the combination of the proposed meta-methods, warm-start ensembles, and time series in a forecast framework for energy consumption in buildings.

The remaining paper is structured as follow. Section 2 describes the problem and presents a brief summary of the state of the art. The third section explains the proposed methods. Experimental results are given in Sec. 4 and the last section presents a conclusion and future work.

2 Preliminary considerations

This section describes the problem and presents a brief state of the art in the resolution of forecasting real-valued time series problems.

2.1 Problem Description

Time series regression are methods for forecasting a future numeric value based on historical responses. Time series regression can help to understand and predict the behavior of dynamic systems from observed data, being commonly used for modeling and forecasting economic, financial and biological systems [26].

This paper proposes a set of meta-methods to predict the energy consumption of buildings. Independently of the number and type of parameters that might differ significantly depending on the available data (e.g., the number and type of sensors that equip the building), it is assumed that we are given a set of N_F dependent variables or features, $(x_1, x_2, \ldots, x_{N_F})$, and want to forecast

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

an independent variable or target, y. Furthermore, for experimental purposes but generalizable, it will be assumed that observations are indexed with a timestamp and the target will be the energy consumption of a building. For training and fitting purpose, it is also assumed that observations will be made inside a time window $\mathcal{W} = [t_s, t_f]$, between the initial, t_s, and final, t_f, instants. Since the observations are made in discrete instants, N_O observations are assumed in the interval $\mathcal{W}$, at moments $\mathcal{T} = \{t_s = t_1, t_2, \ldots, t_f = t_{N_O}\} \subset \mathcal{W}$, being $\Omega = \{(X_t, y_t) : t \in \mathcal{T}\}$ the set of observations, where $X_t = \left(x_{t,1}, x_{t,2}, \ldots, x_{t,N_f}\right)$ are the features values and y_t the corresponding target value. To simplify the description, it is also assumed that observation are taken at regular intervals of time, δ, i.e, $\mathcal{T} = \{t_1, t_1 + \delta, t_1 + 2\delta, \ldots, t_1 + (N_O - 1)\delta\} \subset \mathcal{W}$.

In the training phase, methods will forecast future consumption from a specific moment in time, t_p, and the following setup and goals are considered: (a) Methods will forecast n consumptions with a granularity δ in the period $\Delta_{t_p} = [t_p + \delta, t_p + n\delta]$, i.e., the methods will do forecasts for instants $\mathcal{F}_{t_p} = \{t_{p+1}, t_{p+2}, \ldots, t_{p+n}\} = \{t_p + \delta, t_p + 2\delta, \ldots, t_p + n\delta\}$. (b) During the tuning phase, developers can adjust $\mathcal{F}_{t_p}$ to known (future) values of the dependent (e.g., temperature, occupation) and independent variables, included in the observation set Ω. This allows to use metrics to identify the best conjunction of parameters (see Sec. 2.3). In the production phase, methods will be fed with forecasted values for the independent variables and predict the dependent one. (c) In the fitting and training phases, methods will have available a time window of historical data, with timestamps in $[t_p - \delta_W, t_p]$. Depending on the size of the time window, i.e., the interval of data used to fit the regressors, two types of methods can be considered, namely: the ones that use all historical data, called full memory (FM) methods, and the ones which only use "recent" data, called short memory (SM) methods. δ_W is a parameter which allows to define how long should we go into the past. Obviously, the FM methods can be considered as a sub case of the SM methods as δ_W can be as big as desired.

2.2 Forecasting Time Series Methods

In many energy efficient situations, the data obtained from water consumption, electricity, outdoor temperature, occupation or solar radiation, refer to the same time and place are characterized as multimodal data in the field of Machine Learning. In many of these cases, it is difficult to determine how these different types of data relate or how one modality relates to another. Distinct methods have been defined to perform the prediction of consumption. Globally these methods can be classified in time series versus non-time series techniques.

Time series techniques analyse the data, recorded over equal intervals of time, extracting statistical information and other characteristics from it. Time series solutions can be classified in univariate versus multivariate methods. It is considered an univariate time series if there is a single sequence of values in an observation, while if exists multiple sequences of values in an observation, we have a multivariate time series. Several methods specified for univariate time series analysis were used in energy prediction, for instance the Autoregressive

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

Integrated Moving Average (ARIMA) [37], Case-Based Reasoning [22], Support Vector Machines (SVM) [8,25], Artificial Neural Networks (ANN) [13,20], Grey prediction models [9], Moving Average [16,32], Exponential Smoothing [5] or Fuzzy Time Series [19,29].

Energy consumption normally vary according to other variables like date-time, outside/inside temperatures, humidity, solar radiation and building occupancy. When several time series variables are evaluated together, multivariate time series should be used. Some of the models used in multivariate analysis include the Vector Auto-Regressive [7] method, the vector ARIMA [35], Vector Autoregressive Moving Average [17] and the Bayesian Vector Autoregression [15]. While all these models have been used in predictions, some have not been used in energy consumption forecast.

Besides time series methods, other methods have also been used in energy prediction. These include Regression Analysis [33], Decision Trees [34], and k-Nearest Neighbours (KNN) algorithms [36]. For instance, in [34] a comparison is made between Regression Analysis, Decision Trees and Neural Networks when predicting electricity energy consumption. In [36] the KNN method was applied to the prediction of energy consumption in residential buildings.

Many studies have emphasized the superior performance of Ensemble and Hybrid models [28], as for instance [1,38], which led to the development of different solutions in energy [4,31]. For instance, in [31] an evolutionary multi-objective ensemble learning solution was tested for the prediction of Electricity Consumption, in [4] an ensemble learning framework was created for anomaly detection in energy consumption of buildings, and in [11] a stacking ensemble learning was proposed for short-term prediction of energy consumption. In [21] the authors compared ANN and SVM with an Hybrid Method that combines both, concluding the later achieves the best accuracy. A survey of time series prediction applications using SVM is presented in [30]. A tree-based ensemble method with warm-start gradient for short-term load forecasting is proposed in [39].

Nevertheless, to the best of the authors' knowledge, no meta-method framework which combines bagging, warm-start, and forecasting of energy time series was ever proposed.

2.3 Scoring Methods

To compare the different models, some metrics must be calculated. These metrics measure the distance between the prediction of the model and the real observations [2]. In regression this might be relatively straightforward as we are comparing real numbers, but several metrics are available, each one with different strengths. Considering a set of observation $\mathcal{O} = \{(X_t, y_t) : t \in \{1, 2, \ldots, m\}\} \subset \Omega$ and a function $pred_M$ associated to model M, that predicts y_t given X_t, $\hat{y}_t = pred_M(X_t)$, it is possible to define several performance metrics. For instance, $MAE = \frac{1}{m}\sum_{t=1}^{m} |y_t - \hat{y}_t|$ defines the Mean Absolute Error, $MSE = \frac{1}{m}\sum_{t=1}^{m} (y_t - \hat{y}_t)^2$ the Mean Square Error, $MAPE = \frac{1}{m}\sum_{t=1}^{m} |y_t - \hat{y}_t|/|y_t|$ the Mean Absolute Percentage Error, and $R^2 = 1 - \sum_{t=1}^{m} |y_t - \hat{y}_t| / \sum_{t=1}^{m} |y_t - \bar{y}_t|$, where $\bar{y}_t = \frac{1}{m}\sum_{t=1}^{m} y_t$, defines the Coefficient of Determination score.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

Algorithm 1 Forecasting with the Sliding Regressor method – SR

Require: Forecasting instant (t_p); Unfitted regressor (R); Set of observations (Ω); Size of the training window (δ_W); Set of features values ($\mathcal{X} = \{X_t : t \in \mathcal{F}_{t_p}\}$).
1: $\mathcal{D} \leftarrow$ select data from Ω with timestamp in $[t_p - \delta_W, t_p]$.
2: Fit the regressor R using $\mathcal{D}$.
3: **return** $\{(t, pred_R(X_t)) : X_t \in \mathcal{X}\}$ ▷ *Predict target using regressor R*

3 Proposed Methods

This section presents the three proposed meta-methods, namely: Sliding Regressor, Bagging Sliding Regressor, and Warm-start Bagging Sliding Regressor. The objective was to build a method independent framework, combining bagging and warm-start, for the forecasting of energy consumption time series.

3.1 Sliding Regressor method – SR

The Sliding Regressor (SR) meta-method, in its broad sense, is a traditional regression method that fits a model using given data and a mandatory parameterized regressor (e.g., Decision Tree regressor or Lasso regressor) [2]. The name was chosen thinking that the method will use a sliding time window when applied to the forecast of time series. Two variants can be considered: the full memory method (SR-FM) will use all available data to train the model, while the short memory model (SR-SM) will use a interval of data (the size is a parameter of the method) previous to the instance in time when predictions/forecasts are to be made.

In operation phase, given an initial instant t_p from which predictions are to be made, the SR method will forget any previous fits and refit the regressor using data in the interval $[t_p - \delta_W, t_p]$ (as previously stated, δ_W is a parameter which allows to define how long should we look in the past for data). Then, the fitted method will forecast the consumption in a set of future instants, $\mathcal{F}_{t_p} = \{t'_1, t'_2, \ldots, t'_n\}$. To make the forecast, the fitted model needs the values of the features in those instants $\mathcal{X} = \{X_t : t \in \mathcal{F}_{t_p}\}$. In this case, features might be known in advance or be forecasted themselves. For instance, it will be necessary to forecast the temperature or occupation of the building for the next predicting period (if those are features to be considered). Algorithm 1 sketches the procedure. This method was implemented in order to have a base line for the methods proposed in the next sections (Sec. 3.2 and 3.3).

3.2 Bagging Sliding Regressor method – BSR

The Bagging Sliding Regressor (BSR) method extends the SR method by using a bag/ensemble of estimators. On other words, instead of using a single regressor at every prediction instant, t_p, the BSR fits N_p regressors and for each forecast returns the mean value of the predictions forecasted by each of those regressors. Like the SR, the BSR method has full memory (BSR-FM) and short memory

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

Algorithm 2 Forecasting with the Bagging Sliding Regressor method – BSR

Require: Forecasting instant (t_p); Unfitted regressor (R); Set of observations (Ω); Size of the training window (δ_W); Set of features values ($\mathcal{X} = \{X_t : t \in \mathcal{F}_{t_p}\}$); Size of the bag of regressors (N_p); Percentage of data use to fit regressors (p).

$\mathcal{R} \leftarrow \emptyset$ ▷ *Bag of regressors*
for $i \in \{1, 2, \ldots, N_p\}$ **do**
  $\mathcal{D} \leftarrow$ randomly select $p\%$ of data from Ω with timestamp in $[t_p - \delta_W, t_p]$.
  $R_i \leftarrow$ clone of regressor R
  Fit regressor R_i using $\mathcal{D}$
  $\mathcal{R} \leftarrow \mathcal{R} \cup \{R_i\}$
end for
return $\left\{\left(t, \frac{1}{N_p} \sum_{R \in \mathcal{R}} pred_R(X_t)\right) : X_t \in \mathcal{X}\right\}$

(BSR-SM) variants, which are implemented using parameter δ_W. Furthermore, the BSR method is parameterized by the percentage/fraction of data, p, used to fit each of the regressors. If $p < 1$ the regressors will be fitted with distinct sets of data, as data is selected before each of the regressors is fitted. Algorithm 2 sketches the procedure. Although the method can be used in continuous operation, it can also be applied with full potential in any fixed moment in time (becoming a traditional ensemble method), since it is independent of previous fits, which differs from the following method.

3.3 Warm-Start Bagging Sliding Regressor method – WsBSR

The Warm-Start Bagging Sliding Regressor (WsBSR) considers that the system is fitted in a continuous operation. On other words, a bag of regressors $\mathcal{R}$ is maintained and amplified in regular periods, which can coincide with the forecast moments. So, when forecast are to be made, N_r new regressors are fitted, either in full or short memory (WsBSR–FM or WsBSR–SM) variants, and those regressors are added to an existing bag of regressors. Then, to make a forecast, N_p regressors are selected from the bag using their ages as weights (younger models have higher probability of being chosen) obtaining $\mathcal{R}' \subset \mathcal{R}$. Finally, the forecast value is the mean value of the individual predictions, $\frac{1}{N_p} \sum_{R \in \mathcal{R}'} pred_R(X_t)$, for some set of feature values X_t. Algorithm 3 sketches the procedure.

This section proposed three meta-methods that computational complexity. Next section will experimentally try to conclude on the utility of using one over the others

4 Data and Experiments

For the experimental setup it was used data collected in the Alto da Colina hotel. The Alto da Colina hotel [3] is a 4-star aparthotel located in Albufeira, in the south of Portugal. It is comprised of 174 apartments and contains several facilities, including four outdoor and one indoor swimming pools, a football field,

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

Algorithm 3 Forecasting with the Warm-Start Bagging Sliding Regressor method – WsBSR

Require: Forecasting instant (t_p); Unfitted regressor (R); Set of observations (Ω); Size of the training window (δ_W); Set of features values ($\mathcal{X} = \{X_t : t \in \mathcal{F}_{t_p}\}$); List of already fitted regressors ordered by age ($\mathcal{R} = [R_1, R_2, \ldots, R_k]$); Number of regressors to fit (N_r); Number of regressors to use in the predictions (N_p, $N_p \geq N_r$); Percentage of data use to fit regressors (p).

1: **for** $i \in \{1, 2, \ldots, N_r\}$ **do**
2: $\mathcal{D} \leftarrow$ randomly select p% of data from Ω with timestamp in $[t_p - \delta_W, t_p]$.
3: $R_i \leftarrow$ clone of regressor R.
4: Fit regressor R_i using $\mathcal{D}$ and append the new model to $\mathcal{R}$.
5: **end for**
6: $\mathcal{R}' \leftarrow$ Select N_p models from $\mathcal{R}$ using their ages as weights (younger models have higher probability of being chosen).
7: **return** $\left\{\left(t, \frac{1}{N_p}\sum_{R\in\mathcal{R}'} pred_R(X_t)\right) : X_t \in \mathcal{X}\right\}$

tennis court, gymnasium, a small water park for children and several bars and restaurants, most of these, relying in electric supply. The only exception is the water heating system, that relies in a combination between solar panels (108 Solar Thermal panels) and gas, through two propane boilers.

Data was collected every 15 minutes between January 1st and October 30th, 2017, with the exception of some periods where the system failed to collect some of the features, being those observations discarded. After cleaning up the data, 23.854 observations were considered with the following features: hour, wind velocity, exterior temperature, number of registered guests, weekday and energy consumption. Fig. 1 shows the daily mean consumption (left axis) and number of guests (right axis) for the period in analysis. Although we don't have data after October (the hotel closes in the Autumn), it is observable that the hotel suffers from a estival effect with its higher consumption in the summer months.

4.1 Experimental setup

For experimental purposes, it was decided to have a daily based forecast with the following setup and goals. To analyse the methods it was decided to run them for the full period of known data (January 1st to October 30th, 2017). More precisely, at midnight the method will predict the consumption for the following day or days, defined by a δ_T parameter. In the first phase, it was decided to set $\delta_T = 1$ which, given the gaps in the data acquired, corresponds to 237 predicted days. Since each day has 96 readings (made each 15 minutes), the forecast was made for those known values, allowing us to score our predictions (see Sec. 2.3). Regressor will have available a time window of data defined by δ_W (depending on its size, it is used the full or the short memory methods). Finally, the chosen metrics are applied to the forecasts. Algorithm 4 summarizes the procedure.

The proposed methods are meta-regressors since they require a regressor to make forecasts. In this experimental setup, it was decided to use Decision

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

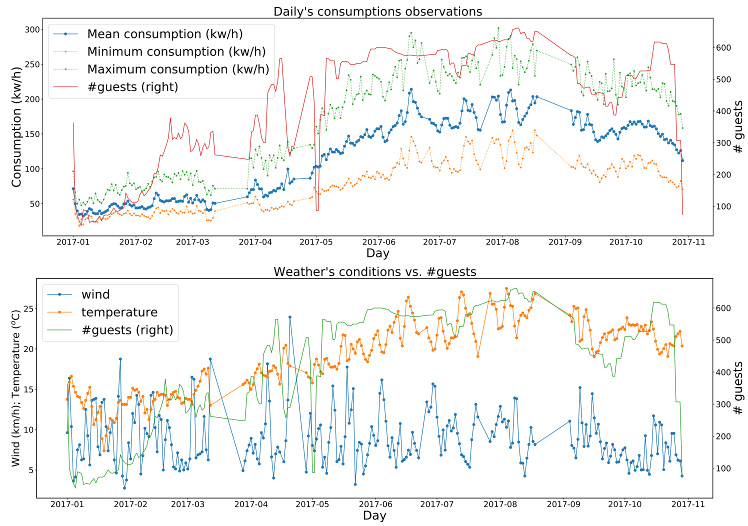

Fig. 1. Variation of the daily consumption and number of guests in the analyzed period (top); Variation in the weather conditions (mean exterior temperature, in ^oC, and mean wind velocity, in km/h) related with the number of guests (bottom).

Tree (DT) Regressors [2,34] for its precision and simplicity. This was a thought decision although it is known that these methods are not adequate for extrapolation, which might be the case whenever higher or lower values of consumption are reached. So, it was used the DT implementation in Scikit-learn [27] with two parameterizations: (a) maximum depth – with nodes expanded until all leaves are pure and (b) limited depth – with a maximum depth of 5 and minimum number of samples required to split an internal node equal to 10. Furthermore, before applying the methods it was also decided to analyse two transformation to the data, namely: polynomial features with degree, $p_d \in \{1, 2, 3\}$, and the scaling of the features to specific ranges, $\mathcal{I} = \{None, [-1, 1], [0, 1], [0, 100]\}$, where $None$ means no transformation. Other parameters are summarized in Tab. 1. Finally, experiments were run on a Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz with 16Mb or RAM and the Kubuntu 20.10 operating system.

4.2 Computational Results

Considering the use of DT with maximum depth, Tab. 2 presents for each of the three methods the 10 best R^2 means (μ_{R^2}) and corresponding standard deviations (σ_{R^2}) values, computed considering the 237 forecasted days (with $96 = 24 \times 4$ forecasts *per* day). Best results are achieved by BSR method with the 10 higher mean R^2 values standing between 0.884 and 0.900. The second best method is WsBSR, followed by SR in third. Regarding parameters, it is observable that using polynomial features with degree 2 seems to provide the best results. The scaling to some interval does not show to have big influence

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

Algorithm 4 Slide-fit-score Algorithm

Require: Set of observations (Ω) in the full time window ($\mathcal{W} = [t_s, t_f]$); Scoring metric (ϕ); Full or short memory (FM or SM) parameter; Forecast period (δ_T); Size of the training window (δ_W).

```
Scores ← ∅
s ← t_s
t ← t_s + δ_W                    ▷ Defines a minimum set of data for the first fitting
while t ≤ t_f do
    Apply the chosen algorithm (see Sec. 3.1, 3.2 and 3.3) to obtain the forecasts,
    Ŷ, for the following δ_T days.
    Extract from Ω the real target values, Y, for the same period as before.
    Scores ← Scores ∪ φ(Y, Ŷ)
    if Short memory then
        s ← s + 1                ▷ Moves s one day forward
    end if
    t ← t + 1                    ▷ Moves t one day forward
end while
return Scores
```

Table 1. Parameters used in the experimental phase (†– used in SR, ‡- used in BSR, and * - used in WsBSR).

Parameter	Values	Observation
δ_W	7	number of days to fit in the short memory case†,‡,*
δ_T	1	number of days to score†,‡,*
p	[0.1, 0.5, 1]	percentage of data used to fit regressor‡,*
p_d	[1, 2, 3]	Polynomial feature transformation degree†,‡,*
$\mathcal{I}$	[$None$, [-1, 1], [0, 1], [0, 100]]	scaling ranges of the features values†,‡,*
N_p	[5, 10, 50]	number of regressors to make prediction‡,*
N_r	[1, 5, 10]	number of new regressors to fit each time*

as results are mixed. In terms of memory, the short memory variant looks to provide the best results, as 20 of the 30 results presented were achieved using it. Furthermore, for the tested values, it seems also advisable to use all (i.e., 100 %) available data to fit the regressor in the short memory case and 50 % of it in the full memory case. Referring to the number of regressors in the bagging methods, in the BSR it looks like "the more the merrier" while in the WsBSR using a large number of regressors in the predictions seems to be worse as, probably, this obliges to use more older regressors. As a side note, to understand the magnitude of the errors, the second best run of BSR (an unscaled case, with $\mu_{R^2} = 0.899$) had $\mu_{MAE} = 4.82$ ($\sigma_{MAE} = 3.1$) and $\mu_{MAPE} = 0.04$ ($\sigma_{MAPE} = 0.02$).

Table 3 presents the same metric values but now considering the use of DT with limited depth of 5 and minimum number of samples required to split an internal node equal to 10. As somehow expected, the value of R^2 got worse but the methods maintained the relative ranking between them. Again the usage

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

Table 2. Top 10 R^2 results for the case of the DT with full depth.

Method	μ_{R^2}	σ_{R^2}	p_d	$\mathcal{I}$	Memory	p	N_p	N_r
SR	0,771	0,288	2	[0, 100]	SM			
	0,766	0,308	2	None	SM			
	0,751	0,274	3	[-1, 1]	SM			
	0,745	0,338	2	[-1, 1]	SM			
	0,736	0,491	2	[0, 1]	SM			
	0,714	0,516	2	[-1, 1]	FM			
	0,690	0,460	3	[0, 100]	SM			
	0,679	0,422	3	[0, 1]	SM			
	0,676	0,678	2	[0, 100]	FM			
	0,675	0,782	2	[0, 1]	FM			
BSR	0,900	0,141	2	[0, 100]	SM	1	50	
	0,899	0,142	2	None	SM	1	50	
	0,898	0,142	2	[0, 1]	SM	1	50	
	0,897	0,143	2	[-1, 1]	SM	1	50	
	0,893	0,163	2	[0, 100]	FM	0.5	50	
	0,890	0,173	2	[0, 1]	FM	0.5	50	
	0,890	0,178	2	[-1, 1]	FM	0.5	50	
	0,887	0,184	2	None	FM	0.5	50	
	0,885	0,155	2	[0, 100]	SM	1	10	
	0,884	0,160	2	None	SM	1	10	
WsBSR	0,838	0,185	2	[0, 1]	SM	1	10	5
	0,833	0,174	2	None	SM	1	10	5
	0,833	0,174	2	[0, 100]	SM	1	10	5
	0,826	0,205	2	[-1, 1]	SM	1	10	5
	0,809	0,184	2	[0, 100]	SM	0.5	10	5
	0,804	0,221	2	[0, 1]	SM	0.5	10	5
	0,802	0,214	2	None	SM	0.5	10	5
	0,796	0,228	2	[0, 1]	FM	0.5	10	5
	0,795	0,277	2	None	FM	0.5	10	5
	0,793	0,263	2	[-1, 1]	FM	0.5	10	5

Table 3. Top 10 R^2 results for the case of the DT with maximum depth of 5.

Method	μ_{R^2}	σ_{R^2}	p_d	$\mathcal{I}$	Memory	p	N_p	N_r
SR	0,607	0,409	2	[]	SM			
	0,577	0,477	2	[0, 1]	SM			
	0,567	0,526	2	[-1, 1]	SM			
	0,536	0,468	2	[0, 100]	SM			
	0,519	0,461	3	[0, 1]	SM			
	0,502	0,551	3	[0, 100]	SM			
	0,492	0,478	3	None	SM			
	0,422	1,593	3	[-1, 1]	SM			
	0,351	0,751	2	[0, 1]	FM			
	0,349	0,775	2	None	FM			
BSR	0,796	0,220	2	[]	SM	1	50	
	0,795	0,223	2	[0, 100]	SM	1	50	
	0,794	0,226	2	[-1, 1]	SM	1	50	
	0,794	0,228	2	[0, 1]	SM	1	50	
	0,767	0,264	2	[-1, 1]	SM	1	10	
	0,767	0,266	2	[0, 100]	SM	1	10	
	0,766	0,277	2	None	SM	1	10	
	0,765	0,278	2	[0, 1]	SM	1	10	
	0,763	0,295	2	[0, 1]	SM	0.5	50	
	0,761	0,301	2	None	SM	0.5	50	
WsBSR	0,710	0,249	2	[]	SM	1	10	5
	0,709	0,272	2	[0, 100]	SM	1	10	5
	0,699	0,262	2	[0, 1]	SM	1	10	5
	0,691	0,286	2	[]	SM	0.5	10	5
	0,685	0,323	2	[0, 100]	SM	0.5	10	5
	0,682	0,353	2	[-1, 1]	SM	1	10	5
	0,680	0,275	2	[0, 1]	SM	0.5	10	5
	0,666	0,306	2	[-1, 1]	SM	0.5	10	5
	0,660	0,289	2	[0, 1]	SM	1	50	10
	0,652	0,307	2	[0, 100]	FM	0.1	10	5

of polynomial features of degree two seems to be the more appropriate and no definitive conclusion can be taken about the scaling of the features. The short memory solution shows to be a better option, when compared with the full memory one. Regarding the number of regressors to use in the ensemble cases, results were not conclusive for BSR but, using 50 allowed to achieve the best results. On the other hand, using only 10 regressors can be a good enough forecast for a decision maker, saving considerable computational resources. For the WsBSR method, a similar conclusion as previously can be taken, i.e., using a larger number of regressors in the predictions seems to worsen the results (probably for the obligation of using older regressors).

5 Conclusion and Future Work

Energy is one of the largest parcels in the operation of service buildings. The possibility to forecast the energy consumption of unmovable loads, added up with the fixed ones, gives decision makers the chance to plan ahead the positioning

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

of the movable charges. These can be later optimized taking into consideration many factors as energy production from renewable sources or energy prices.

In this paper, three meta-methods are analyzed to perform the referred forecast. Established the regressor to be used with the meta-method (e.g., Decision Tree regressor), the first (SR) can be seen as a traditional forecast method, while the second (BSR) and third (WsBSR) use ensembles of regressors to make the forecasts. The difference between them comes from the fact that the last uses a warm-start procedure, adding new regressors when required, which are complemented by the ones already fitted in the past. Over the elected parameters, the BSR method has shown a better accuracy but with higher computational cost in the fitting phase. The use of the WsBSR is therefore a possibility as it has shown slightly worse results but with a fraction of the computational cost (in a typical run, for similar set of parameters, WsBSR took approximately 25% of the time required by BSR to run the full simulation).

In terms of future work, other methods besides the decision trees are to be used. Furthermore, the usage of heterogeneous methods in the ensemble will be tried, i.e., the usage of the same method but with different parameters and also distinct methods. In this case, it is also intended that the selection of the forecasting methods (which ones to fit and which ones to use in the forecast) will be tuned on run time. Finally, integration of data assimilation techniques also seem a very promising field of research.

References

1. Ahmad, A., Hassan, M., Abdullah, M., Rahman, H., Hussin, F., Abdullah, H., Saidur, R.: A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renewable and Sustainable Energy Reviews **33**, 102–109 (may 2014). https://doi.org/10.1016/j.rser.2014.01.069
2. Alpaydin, E.: Introduction to machine learning. MIT press (2020)
3. Alto da Colina Hotel: `https://www.alfagar.com/alfagar-alto-da-colina.html`, (accessed Jan. 22, 2021)
4. Araya, D.B., Grolinger, K., ElYamany, H.F., Capretz, M.A., Bitsuamlak, G.: An ensemble learning framework for anomaly detection in building energy consumption. Energy and Buildings **144**, 191–206 (jun 2017). https://doi.org/10.1016/j.enbuild.2017.02.058
5. Bindiu, R., Chindris, M., Pop, G.: Day-ahead load forecasting using exponential smoothing. Scientific Bulletin of the "Petru Maior" University of Targu Mures **6**, 89 (2009)
6. Cabrita, C.L., Monteiro, J.M., Cardoso, P.J.S.: Improving energy efficiency in smart-houses by optimizing electrical loads management. In: 2019 1st International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED). IEEE (May 2019). https://doi.org/10.1109/synergy-med.2019.8764140
7. Chandramowli, S., Lahr, M.L.: Forecasting New Jersey's electricity demand using auto-regressive models. SSRN Electronic Journal (2012). https://doi.org/10.2139/ssrn.2258552
8. Chen, T.T., Lee, S.J.: A weighted LS-SVM based learning system for time series forecasting. Information Sciences **299**, 99–116 (apr 2015). https://doi.org/10.1016/j.ins.2014.12.031

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

9. Chiang, J., Wu, P., Chiang, S., Chang, T., Chang, S., Wen, K.: Introduction to grey system theory. Gao-Li Publication, Taiwan (1998)
10. Deb, C., Zhang, F., Yang, J., Lee, S.E., Shah, K.W.: A review on time series forecasting techniques for building energy consumption. Renewable and Sustainable Energy Reviews **74**, 902–924 (Jul 2017). https://doi.org/10.1016/j.rser.2017.02.085
11. Divina, F., Gilson, A., Goméz-Vela, F., Torres, M.G., Torres, J.: Stacking ensemble learning for short-term electricity consumption forecasting. Energies **11**(4), 949 (apr 2018). https://doi.org/10.3390/en11040949
12. Eurepean Union: Regulation (Eu) 2018/1999 of the European Parliament and of the Council (2018), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2018.328.01.0001.01.ENG
13. Gómez, J., Molina-Solana, M.: Towards self-adaptive building energy control in smart grids. In: NeurIPS 2019 Workshop Tackling Climate Change with Machine Learning. Vancouver, Canada (Dec 2019), https://www.climatechange.ai/NeurIPS2019_workshop.html
14. Governo Português: Plano nacional de energia e clima 2021-2030 (PNEC 2030) (2020), https://apambiente.pt/_zdata/Alteracoes_Climaticas/Mitigacao/PNEC/PNECPT_TemplateFinal201930122019.pdf
15. Joutz, F.L., Maddala, G.S., Trost, R.P.: An integrated bayesian vector auto regression and error correction model for forecasting electricity consumption and prices. Journal of Forecasting **14**(3), 287–310 (may 1995). https://doi.org/10.1002/for.3980140310
16. Karim, S.A.A., Alwi, S.A.: Electricity load forecasting in UTP using moving averages and exponential smoothing techniques. Applied Mathematical Sciences **7**, 4003–4014 (2013). https://doi.org/10.12988/ams.2013.33149
17. Kascha, C.: A comparison of estimation methods for vector autoregressive moving-average models. Econometric Reviews **31**(3), 297–324 (may 2012). https://doi.org/10.1080/07474938.2011.607343
18. Lee, D., Cheng, C.C.: Energy savings by energy management systems: A review. Renewable and Sustainable Energy Reviews **56**, 760–777 (Apr 2016). https://doi.org/10.1016/j.rser.2015.11.067
19. Lee, W.J., Hong, J.: A hybrid dynamic and fuzzy time series model for mid-term power load forecasting. International Journal of Electrical Power & Energy Systems **64**, 1057–1062 (jan 2015). https://doi.org/10.1016/j.ijepes.2014.08.006
20. Li, K., Hu, C., Liu, G., Xue, W.: Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. Energy and Buildings **108**, 106–113 (dec 2015). https://doi.org/10.1016/j.enbuild.2015.09.002
21. Liu, Z., Wu, D., Liu, Y., Han, Z., Lun, L., Gao, J., Jin, G., Cao, G.: Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction. Energy Exploration & Exploitation **37**(4), 1426–1451 (jan 2019). https://doi.org/10.1177/0144598718822400
22. Monfet, D., Corsi, M., Choinière, D., Arkhipova, E.: Development of an energy prediction tool for commercial buildings using case-based reasoning. Energy and Buildings **81**, 152–160 (oct 2014). https://doi.org/10.1016/j.enbuild.2014.06.017
23. Monteiro, J., Cardoso, P.J., Serra, R., Fernandes, L.: Evaluation of the human factor in the scheduling of smart appliances in smart grids. In: International Conference on Universal Access in Human-Computer Interaction. pp. 537–548. Springer (2014)
24. Monteiro, J., Eduardo, J., Cardoso, P.J.S., ao, J.S.: A distributed load scheduling mechanism for micro grids. In: 2014 IEEE International Confer-

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_26

ence on Smart Grid Communications (SmartGridComm). IEEE (Nov 2014). https://doi.org/10.1109/smartgridcomm.2014.7007659
25. Nie, H., Liu, G., Liu, X., Wang, Y.: Hybrid of ARIMA and SVMs for short-term load forecasting. Energy Procedia **16**, 1455–1460 (2012). https://doi.org/10.1016/j.egypro.2012.01.229
26. Paolella, M.S.: Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH. John Wiley & Sons (2018)
27. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
28. Rokach, L.: Ensemble Learning. WSPC (Feb 2019), `https://www.ebook.de/de/product/35671842/lior_rokach_ensemble_learning.html`
29. Sadaei, H.J., Enayatifar, R., Abdullah, A.H., Gani, A.: Short-term load forecasting using a hybrid model with a refined exponentially weighted fuzzy time series and an improved harmony search. International Journal of Electrical Power & Energy Systems **62**, 118–129 (nov 2014). https://doi.org/10.1016/j.ijepes.2014.04.026
30. Sapankevych, N.I., Sankar, R.: Time series prediction using support vector machines: A survey. IEEE Computational Intelligence Magazine **4**(2), 24–38 (2009). https://doi.org/10.1109/MCI.2009.932254
31. Song, H., Qin, A.K., Salim, F.D.: Evolutionary multi-objective ensemble learning for multivariate electricity consumption prediction. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE (jul 2018). https://doi.org/10.1109/ijcnn.2018.8489261
32. Taylor, J.W.: Triple seasonal methods for short-term electricity demand forecasting. European Journal of Operational Research **204**(1), 139–152 (jul 2010). https://doi.org/10.1016/j.ejor.2009.10.003
33. Tso, G.: A study of domestic energy usage patterns in Hong Kong. Energy **28**(15), 1671–1682 (dec 2003). https://doi.org/10.1016/s0360-5442(03)00153-1
34. Tso, G.K., Yau, K.K.: Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy **32**(9), 1761–1768 (sep 2007). https://doi.org/10.1016/j.energy.2006.11.010
35. Vu, K.M.: The ARIMA and VARIMA time series: their modelings, Analyses and Applications. AuLac Technologies Inc. (2007)
36. Wahid, F., Kim, D.: A prediction approach for demand analysis of energy consumption using k-nearest neighbor in residential buildings. International Journal of Smart Home **10**(2), 97–108 (feb 2016). https://doi.org/10.14257/ijsh.2016.10.2.10
37. Wang, X., Meng, M.: A hybrid neural network and ARIMA model for energy consumption forcasting. Journal of Computers **7**(5) (may 2012). https://doi.org/10.4304/jcp.7.5.1184-1190
38. Wang, Z., Srinivasan, R.S.: A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. Renewable and Sustainable Energy Reviews **75**, 796–808 (aug 2017). https://doi.org/10.1016/j.rser.2016.10.079
39. Zhang, Y., Wang, J.: Short-term load forecasting based on hybrid strategy using warm-start gradient tree boosting. Journal of Renewable and Sustainable Energy **12**(6), 066102 (Nov 2020). https://doi.org/10.1063/5.0015220, `https://doi.org/10.1063/5.0015220`
40. Zia, M.F., Elbouchikhi, E., Benbouzid, M.: Microgrids energy management systems: A critical review on methods, solutions, and prospects. Applied Energy **222**, 1033–1055 (Jul 2018). https://doi.org/10.1016/j.apenergy.2018.04.103