Energy Conversion and Management: X 24 (2024) 100777

Contents lists available at ScienceDirect

Energy Conversion and Management: X

journal homepage: www.sciencedirect.com/journal/energy-conversion-and-management-x

Time of the week AutoRegressive eXogenous (TOW-ARX) model to predict thermal consumption in a large commercial mall

Iñigo Lopez-Villamor [a,b,*], Olaia Eguiarte [a], Beñat Arregi [a], Roberto Garay-Martinez [c], Antonio Garrido-Marijuan [a]

[a] TECNALIA, Basque Research and Technology Alliance (BRTA), Astondo bidea 700, 48160 Derio, Spain
[b] University of Deusto, Faculty of Engineering, Av. Universidades 24, 48007 Bilbao, Spain
[c] University of Deusto, Institute of Technology, Faculty of Engineering, Av. Universidades 24, 48007 Bilbao, Spain

ARTICLE INFO

Keywords:
Thermal building model
ARX
Time of the week
Consumption
Shopping mall

ABSTRACT

This paper proposes a procedure to build a Time of the Week AutoRegressive eXogenous (TOW-ARX) model, indexed with respect to time and day of the week, to characterize heat consumption in tertiary buildings. Models for building heat load characterization and prediction are crucial to enhance energy efficiency. The proposed model can be used for different purposes, e.g., control of indoor climate, or characterization of the thermal response of the building. A case study is described where the TOW-ARX model is used to characterize the energy consumption of a large retail building in Madrid. In order to discard the risk of model overfitting, cross validation is applied using the k-fold technique. The performance of the TOW-ARX model is compared with a set of different models: a reduced version of the model where similar segments are clustered using the k-means method (R-TOW-ARX), a general ARX model, a linear regression steady-state TOW model (TOW-LR), a version of the latter reduced through clustering (R-TOW-LR), and a general multiple linear regression model (LR). The results reveal that ARX-based models notably outperforms the rest. The TOW-ARX model shows the best metrics, but also outnumbers the number of coefficients of the other models by far. The selection of the most suitable model is not straightforward and should depend on the purpose of such model: the TOW-ARX model would arguably be the best for control purposes due to its low mean absolute error, but the ARX model would be preferable for an efficient characterization of the thermal response of a building due to its reduced number of parameters.

1. Introduction

1.1. Context

Improving the energy performance of buildings is crucial as they account for 40 % of global energy consumption and one-third of greenhouse gas emissions [1]. In the EU, about 75 % of buildings have a poor energy performance [2], with a low replacement rate of existing buildings, accounting for approximately 1–3 % per year [3]. Within this group, large retail centers and shopping malls are buildings with very relevant specific energy consumption. With variations for type and size of buildings, several sources specify energy loads in the range of 164 to 585 kWh/m^2 [4], which can be up to 1000 kWh/m^2 [5] for food stores.

Advanced building energy management and optimal controls are low-intrusive technologies that can be used to improve building energy performance. With real-time monitoring and optimization of Heating, Ventilation and Air Conditioning systems (HVAC) substantial savings, ranging from 15 to 50 %, can be achieved [6].

A reliable characterisation and forecasting of building energy consumption is crucial for understanding the causal relationships between energy use, boundary conditions, and occupant behavior. This knowledge enables the development of more effective energy management and control strategies.The development of accurate models requires comprehensive data from building management systems. Over the last few years, remarkable progress has been made in integrating SCADA systems with IoT technologies and cloud-based solutions. At the same time, the deployment of smart metering infrastructure [7] has made readily available time series energy consumption data. Altogether this progress results in having a wealth of data [8 9] available for the development of data-driven models and predictive analytics [10 11 12], unlocking new possibilities for real-time monitoring, data analysis, and remotely controlling of building systems.

* Corresponding author at: TECNALIA, Basque Research and Technology Alliance (BRTA), Astondo bidea 700, Derio 48160, Spain.
E-mail address: inigo.lopez@tecnalia.com (I. Lopez-Villamor).

https://doi.org/10.1016/j.ecmx.2024.100777
Received 2 August 2024; Received in revised form 11 October 2024; Accepted 27 October 2024
Available online 28 October 2024

1.2. State of the Art

Originally, data-driven methods were heavily influenced by the lack of IT infrastructure for the measurement, communication, and processing of data. In this context, Energy Signature (ES) methods were developed, where the correlation between (monthly) interval data recorded by utilities to heating-degree-days was exploited. PRISM, an initial version of such methods dates back to 1986 [13]. Later on, this approach was further exploited to develop more advanced approaches such as the ASHRAE changepoint model [14], where the correlation of heating and cooling loads to heating and cooling degree-days is modeled. The underlying assumption of these methods is that there is a stable energy consumption baseload, and that additional loads are introduced to meet heating (higher at low temperatures) or cooling (higher at high temperatures) needs.

ES models are a (quite simplistic) type of data-driven model, which can be expanded into more complex multiple-input models that include solar radiation [15 16] and plug-loads [17] as explanatory variables. The well-known dependencies of energy loads with regards to weekday-weekend patterns are incorporated in so-called Time-Of-Week temperature (TOW) models [18]. This approach has been exploited to develop TOW-segmented ES models in [16 19] that were able to characterize the variations of energy performance due to the daily and weekly fluctuation in the occupancy and usage patterns of buildings. Further approaches have also been developed where this pattern analysis is performed to identify typical days [10 20] and reduce the number of segmentations required.

All the aforementioned models perform well in delivering energy performance assessments at aggregated levels (i.e., weekly, monthly, yearly). However, they fail to capture the actual dynamics of buildings, leading to inaccuracies in granular predictions (i.e., hourly resolution), which are essential for short-term performance forecasting and applications with variable energy costs. Existing literature [16] has demonstrated the ability to identify the impact of environmental conditions and usage patterns on building energy performance. Nonetheless, these models require further refinement to adequately address the transient behavior of buildings.

Short-term transient models can be formulated through time series models such as the general ARMAX model [21], where the Auto Regressive (AR) terms account for the system dynamics, the Moving Average (MA) corrects for forecasting errors and eXogenous inputs (X) can be used to incorporate present and past values of boundary conditions (e.g. climate, occupancy) and control inputs (e.g. energy provided by HVAC systems). Several works have proven the capacity of ARX models to provide short-term forecasting models for buildings [22 23]. These models take the well-known linearity of the thermal dynamics in buildings to outdoor conditions as a ground truth, while some of them have explored the possibility to incorporate non-linearities associated to solar geometry by means of engineering equations [24] or spline functions [23]. Other works have worked on mixed approaches that incorporate the short (a few hours) and medium-term (preceding day) history [25].

As access to larger datasets became available, more complex data-driven models (multiple regression, artificial neural networks, decision trees, support vector machines, etc.) have been introduced. These models are mathematically powerful systems to model complex interactions in buildings [26], but are not formulated considering any physical criteria.

With extensive research on electric smart grids in the last decades and greater data availability for electricity meters at building level [16], most of the literature on data-driven models has focused on the field of electricity load analysis. Only a few studies have focused on heating load profiles and even fewer are based on actual measurements. Therefore, heat load characterization and prediction methods are relatively new, and this research field is yet to be consolidated, with research works that port the know-how on data-driven models to thermal systems [16 27].

The sensitivity to climatic and calendar variables is similar to that observed in electric systems, where thermal performance and/or occupancy patterns are deeply related to them. Andersen et al. concluded that climatic conditions highly affect final electricity consumption in dwellings [28], and Carmo et al. analysed data from more than 4500 smart meters to conclude that individual electricity loads should be differentiated by use categories (residential, industrial, etc.), weekday and weekend, and summer and winter [27]. Different formulations of TOW data [29 30 31] are used to pass on this information to models.

Overall, there is an interesting trend towards the use of data-driven models due to their lower development effort compared to other modelling approaches [32 33], but their suitability needs to be further investigated. Models should have physical consistency, even if not explicitly formulated as physical/engineering equations. These should be able to incorporate the effects of environmental variables, the impact of occupancy/usage, consider changes in operational criteria, and incorporate inertial behaviour.

1.3. Considerations on shopping malls

Shopping malls are buildings specifically devoted to retail and leisure facilities. Although their configuration depends on many factors, malls are commonly very large buildings, typically characterized by high ventilation rates and relatively compact aspect ratios [4 34]. Being so, energy needs to keep these buildings comfortable are usually not strongly correlated with energy transfer through the building envelope, but mainly linked to the efficiency of HVAC systems and their control (with the exception of highly glazed buildings and/or sub-zones, where solar heat gains may be significant).

Additionally, occupancy is highly dependent with regards to commercial schedules, commuting periods and the availability of leisure times. A strong intra-weekly pattern is typically observed, with increased occupancy in weekends.

In this context, useful energy models need to ensure that their predictions are accurate, both over a long integration period (i.e., year/month) as well as subsets of these (i.e., correspondence with peak tariff periods), incorporate performance variations associated with usage schedules, and incorporate the transient effects associated to inertial processes in buildings.

The current literature on energy forecasting for these buildings presents machine learning algorithms to characterize cooling loads [35 36 37], or even the complete energy consumption of a mall based on simulated data [38]. These have proven good prediction metrics for long assessment periods (i.e., by means of aggregated metrics such as RMSE), but models have not sufficiently achieved the other features identified in the previous paragraph.

To the authors' knowledge, the use of time series or other inertial models to predict heating consumption in shopping malls has not been explored. Inertial models are an interesting approach to achieve the desired features and are backed by their proven accuracy in load forecasting and their computational efficiency.

1.4. Contribution of this work

This paper presents a methodology to build TOW-segmented ARX models to predict thermal loads in buildings, and demonstrates its application in predicting heating and cooling loads in a shopping mall. While the study builds on existing works, it presents a differential approach to other existing works in the field of data-driven energy performance:

- When compared to pattern clustering works, the model incorporates a transient thermal model.
- When compared to energy signature works and TOW models, it incorporates the short-term transient phenomena through an ARX model.

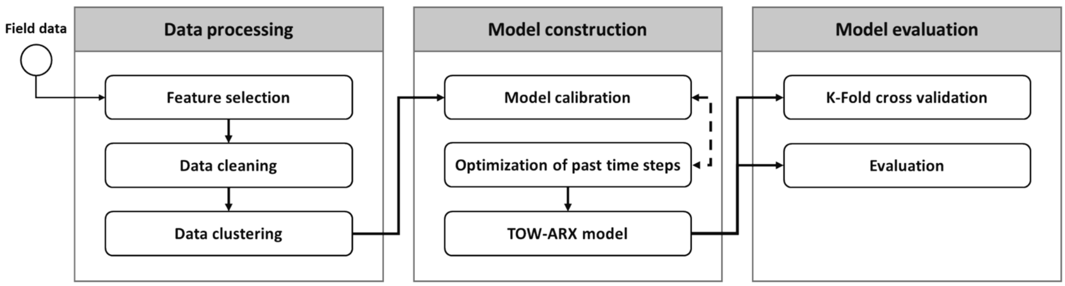

Fig. 1. TOW-ARX model development steps.

- When compared to time-series modelling, it incorporates a way to consider behavior variations over the week by means of the TOW segmentation. A reduced version of the model, where supervised clustering is applied to the previous model to reduce the complexity of the inputs, is incorporated to the study and compared with the full model.
- The model is developed and validated with real data from a shopping mall in Madrid (Spain), a substantially more complex environment than other research works developed over simulated data.

Furthermore, the proposed approach is compared against several models available in the literature, focusing on both accuracy and qualitative aspects. The performance of the TOW-ARX model is evaluated against LR, ARX, and TOW-LR. In addition, a reduced version of TOW-LR is tested, incorporating the clustering process developed for the reduced version of TOW-ARX, providing a comprehensive comparison. Accordingly, this paper is structured as follows. Section 2 illustrates the methodology to build a TOW-ARX model for predicting energy consumption in buildings, along with the evaluation metrics for validation and the framework of comparison. Section 3 describes the case study used for testing. Section 4 displays the results of applying the TOW-ARX on the dataset, presents the comparison results of the baseline/reference/comparison methods. Section 5 analyses the results of all 6 models and discusses the benefit introduced by the proposed method. Finally, Section 6 summarizes the contribution and significance of this paper, as well as potential directions for further development.

2. Methodology

This section describes the proposed methodology, resumed in Fig. 1, for the identification and evaluation of a TOW-ARX model.

2.1. Time of the week (TOW) applied to ARX models

The objective is to create a model that optimises the prediction of thermal consumption in buildings one hour ahead, improving the accuracy and robustness of currently existing models.

TOW models [30] are known to improve the accuracy of forecasting models due to their capability to characterize unknown disturbances that would otherwise not be captured in the training data. On the other hand, ARX models are a combination of autoregressive (AR) and exogenous (X) input models. In contrast to linear regression (LR) models, they are capable of representing the dynamic phenomena of the building [39], e.g., thermal inertia, by means of the inclusion of past values. The time delay or order of the autoregressive part and of each of the exogenous variables is independent and needs to be optimized for each model. ARX models are generally represented mathematically as Eq. (1).

$$y_t = \sum_{p=1}^{P} \theta_p y_{t-p} + \sum_{n=1}^{N} \sum_{m_n=0}^{M_n} \beta_{n,m_n} x_{n,t-m_n} + \varepsilon_t \tag{1}$$

being θ the coefficients of the AR part of the model, β the coefficients of the exogenous variables of value x , P the number of past time steps in the AR part, N the number of exogenous variables, M_n the number of past time steps of each exogenous variable n, and ε_t the error term representing the unexplained part of the model.

The proposed novel methodology identifies a new kind of ARX model that combines the strengths of both ARX and TOW models. As shown in Fig. 2, while in a usual ARX model all the dataset is used to calibrate a single model that works for the whole period, in the case of TOW-ARX models the data is clustered by its TOW (168 clusters with hourly resolution) and each cluster is used to train an independent ARX submodel that works for its corresponding TOW. This implies that different time delays and sets of coefficients are used in the predictive equation for each TOW.

2.2. Data preprocessing

2.2.1. Variable selection

Since the proposed model's objective is to predict the thermal consumption of buildings, that variable is adopted as the autoregressive term of the model. Regarding the possible exogenous variables, the literature cites several possible variables depending on the available data:

• **Weather data:** based in previous works [40], outdoor temperature and solar radiation have been identified as parameters that significantly affect building energy consumption.

• **Occupancy data**: as several studies highlight [41,42] occupancy rates are directly connected with buildings consumption values.

• **Indoor or set point temperature data**: indoor temperature is used in several studies [43] as an input of the model to represent the indoor

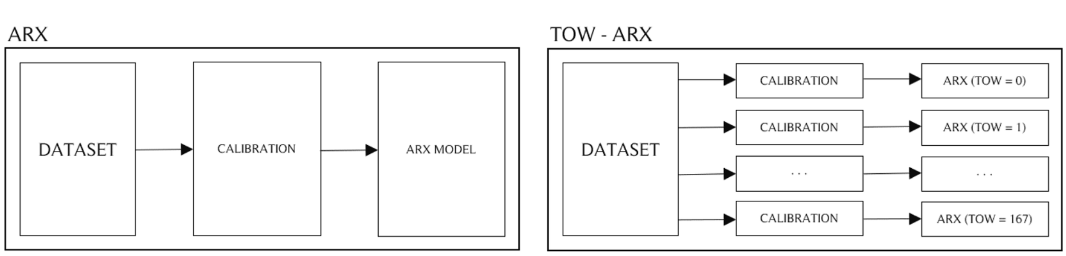

Fig. 2. Model identification procedure for ARX (left) and TOW-ARX (right) models.

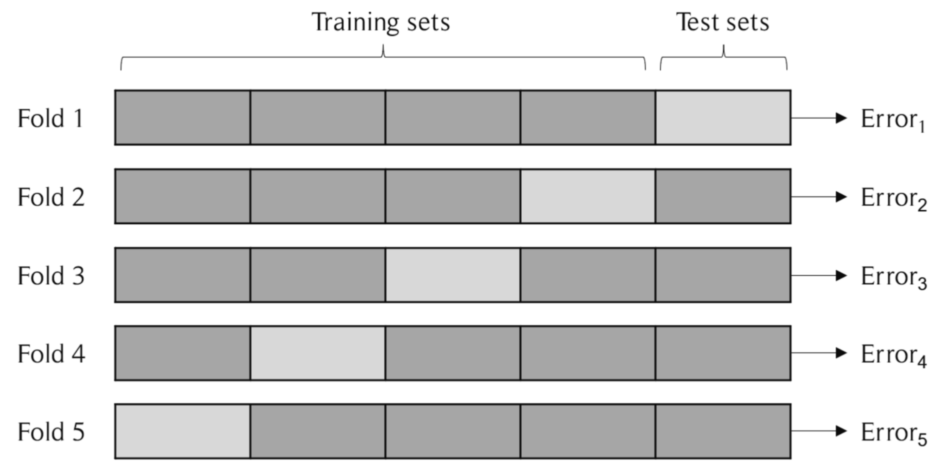

Fig. 3. Example of cross-validation with k = 5.

conditions of the modelled zone or system. In cases where indoor temperature is not available, the set point temperature of the system can be used as it is directly related to it [44,45].

2.2.2. Data cleaning

Raw datasets often contain missing values or reading errors. In this work the following procedure is proposed in order to guarantee the cleanliness of the data used to train the TOW-ARX model, and to identify any wrong values.

Outliers are detected by defining thresholds for each variable based on the expected values in normal operation of the building and its systems. In addition, any period with a missing value in any of the variables described in the previous section is removed from the dataset.

2.3. Model calibration

For each of the 168 submodels, the set of optimal parameters θ and β (Eq. (1)) is obtained by minimizing the sum of the squares of the residuals through the linear regression function of the ScikitLearn [46] Python library.

2.4. Model selection

Statistical tests are used to select the most suitable model, defined as the smallest model able to describe the main aspects of the a given dataset. When comparing complex models, methods based on information criteria can be used. Various of the most popular information criteria for the determination of the most adequate model order are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [47 48]. Both criteria are comparative metrics. This implies that they cannot be used to evaluate a model on its own, but rather to compare models with each other. The lower the value of the criteria, the more appropriate the model is among the compared alternatives. In this particular case, due to the large number of submodels, BIC is chosen because of its tendency to reward simpler models. In Eq. (2) the formulation of the BIC is given:

$$\boldsymbol{BIC} = -2\mathbf{L} + \mathbf{k}\cdot\mathbf{ln}(\mathbf{n}) \tag{2}$$

where L is the log likelihood function of the model evaluated at the maximum likelihood estimate, k is the number of estimated parameters, and n is the number of datapoints.

During the model selection procedure, a maximum order of 6 is set for all AR and X parameters based in other works [22]; for each time of the week, the model with the lowest BIC is selected.

2.5. Evaluation metrics

There are multiple metrics which can be used to evaluate the accuracy of building energy models. The ones chosen in this work are commonly used for validating and comparing models [49–51]:

$$\boldsymbol{MAE} = \frac{1}{\boldsymbol{n}}\sum_{i=1}^{\boldsymbol{n}}|\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i| \tag{3}$$

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{\boldsymbol{n}}\sum_{i=1}^{\boldsymbol{n}}|\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i|^2} \tag{4}$$

$$\mathbf{R}^2 = 1 - \frac{\sum_{i=1}^{n}(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)^2}{\sum_{i=1}^{n}(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^2} \tag{5}$$

Both the mean absolute error (MAE, Eq. (3)) and the root mean square error (RMSE, Eq. (4)) quantify the difference between predicted and measured values. While the MAE treats all errors equally, the RMSE penalizes larger errors to a greater extent, making it more sensitive to outliers. R^2 (Eq. (5)), also known as coefficient of determination, measures how much of the variance within the measured parameter is captured by the model. The higher the value of R^2, the closer predicted values are to target values.

2.6. Model validation

One of the most important parts of a modelling procedure is a robust validation method to ensure that the identified model is reliable not only within the training conditions but also beyond. For this purpose, the data is normally divided into two sets: 1) a training set and 2) a validation or test set. The training set is used to tune the model and coefficients. The model is then applied using the test dataset to check if it captures the real behavior of the building when using different data than that used in the training. In the case of the TOW-ARX models this procedure is critical because of the possibility of overfitting the model due to the great number of parameters of the 168 submodels. To guarantee that this does not happen, a cross-validation method is implemented [52,53].

In cross-validation the training dataset is divided in k groups or folds, as shown in Fig. 3. For k times, the model is trained with the data of k-1 folds, and it is tested against the remaining fold until k set of results are obtained. If the standard deviation of the evaluation metrics of the folds is negligible, it can be assumed that the model does not overfit. Conversely, if the accuracy of one or more folds is much higher than the rest, it might suggest that the model is overfitted. Alternatively, if the accuracy of all the folds is less than acceptable, the model may be

Table 1
Example of application of the submodel reduction methodology with k = 3 in the k-means clustering technique.

Datapoints grouped by TOW	% of data classified as Cluster 1	% of data classified as Cluster 2	% of data classified as Cluster 3		Final cluster
TOW = 1	100%	0%	0%	Cluster 1 > 80% →	1
TOW = 2	13%	87%	0%	Cluster 2 > 80% →	2
TOW = 3	17%	22%	61%	No cluster > 80% →	On its own
...	...	...	...	...	...
TOW = 168	5%	18%	77%	No cluster > 80% →	On its own

underfitted and it does not accurately capture the relationship between input and output variables.

2.7. Model comparison

For the purpose of assessing the accuracy of the model, its results are compared against the results of five different models of decreasing complexity. The data used to evaluate these models is the same as for the TOW-ARX model.

1. **Reduced TOW-ARX model (R-TOW-ARX):** a TOW-ARX model with a reduced number of TOW submodels. The reduction is done applying a k-means clustering technique [54] to the whole data. This method clusters together several time-of-the-week groups based on their equivalent operational pattern, substantially reducing the number of submodels by identifying representative hourly profiles. To ensure the interpretability of the clustering, all data points belonging to the same time of the week are grouped in the same cluster. The following clustering criteria is applied: when 80 % or more of data points with the same TOW belong to the same cluster, the cluster is applied to all the data of that TOW, otherwise, that TOW it is classified as a cluster on its own. As it was done for the TOW-ARX model, the BIC is used to optimize the time delay of the variables. A practical example of this submodel reduction methodology is shown in Table 1.
2. **ARX model (ARX):** an ARX model where the BIC is applied to optimize the time delay of the variables.
3. **Time of the week linear regression model (TOW-LR):** a time-indexed model conformed by 168 linear regression models. It is mathematically equivalent to the TOW-ARX model with a time delay equal to 0.
4. **Reduced TOW-LR model (R-TOW-LR):** a time-indexed model conformed by n linear regression models, where n is the number of clusters calculated applying the k-means clustering technique and the same criteria as in the R-TOW-ARX model. This model is mathematically equivalent to the R-TOW-ARX model with a time delay equal to 0.
5. **Linear regression model (LR):** a general linear regression model. This model is mathematically equivalent to the ARX model with a time delay equal to 0.

3. Case study

This section describes the selected case study where the presented approach is applied.

3.1. Description of building

The selected building is one of the biggest shopping centres located in Madrid (Spain), with 166 different shops. The building has a floor area of 85,000 m^2 divided into two floors. The first one is devoted to shops while the second houses bars and restaurants. The shopping centre is open from Monday to Sunday only closing on bank holidays. Most shops open from 10:00 to 22:00, nevertheless, restaurants and cinemas are open until midnight. In the common areas, rooftop units using natural gas boiler for heating and electricity for cooling provide the required thermal needs through an AHU connected to each of them.

3.2. Available data

Even if there is data available for each of the individual HVAC systems present in the building, since the purpose of the present work is to predict the thermal consumption of the building as a whole, the total consumption data is considered in the study. Specifically, the overall electric consumption is used since it represents the biggest share of the energy consumption due to the high cooling needs of the building. This

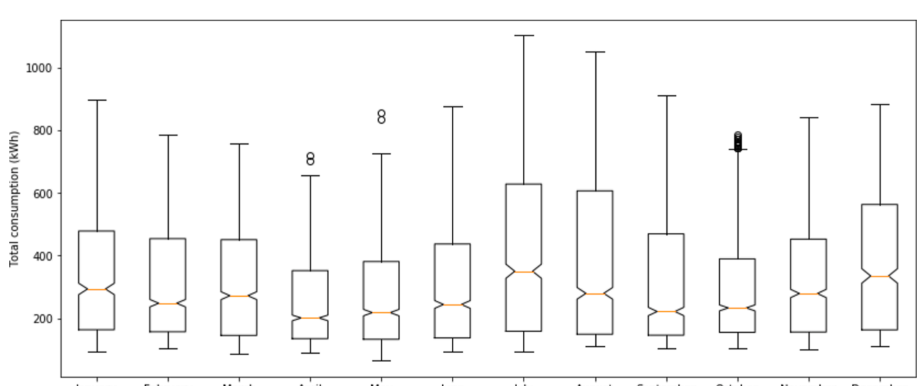

Fig. 4. Monthly distribution of the total consumption.

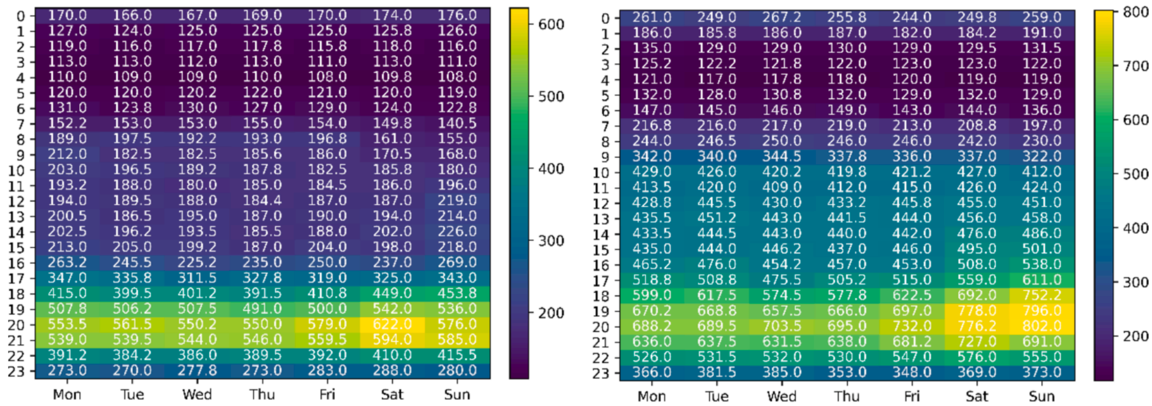

Fig. 5. Heatmap for first quartile (left) and third quartile (right) of total consumption (kWh). The columns of the heatmap represent day of the week and the rows represent time of the day.

Table 2
The coefficient values of ARX submodels when TOW equal to 10, 65 and 141.

ARX submodel TOW = 10				ARX submodel TOW = 65				ARX submodel TOW = 141			
$\beta_{1,0}$	−2.1634	θ_1	0.8453	$\beta_{1,0}$	1.8586	θ_1	1.1910	$\beta_{1,0}$	−0.4748	$\beta_{3,0}$	0.4116
$\beta_{2,0}$	0.0009	θ_2	0.9215	$\beta_{2,0}$	−0.0782	θ_2	−0.3108	$\beta_{2,0}$	2.8352	$\beta_{4,0}$	0.0069
$\beta_{3,0}$	−0.9139	θ_3	−1.1005	$\beta_{3,0}$	2.0699			$\beta_{2,1}$	1.1316	θ_1	0.9765
$\beta_{4,0}$	0.0277			$\beta_{4,0}$	0.0156			$\beta_{2,2}$	−0.5052	θ_2	−0.2576

consumption data, as well as occupancy values of the building in an hourly basis, is obtained from the BMS of the building. Moreover, as a representative indoor temperature of the whole building is not available, the set point temperature of the system is used as an input parameter for the model. On the other hand, ambient temperature and solar radiation values are obtained using Weatherbit API [55].

3.3. Data description

Two years of data, from 01/02/2021 to 17/02/2023, are used with an hourly resolution. The data cleaning process described in Section 2.2.1 was applied to the whole dataset. Thus, reading errors, missing elements and identified outliers were removed from the dataset. During this process less than 10 % of the datapoints were eliminated, with most of those corresponding to periods with missing values.

Fig. 4 depicts the monthly distribution of the total electric consumption of the shopping centre. As it can be seen, the monthly mean consumption ranges between 200 and 350 kWh, being higher during summer and winter months and lower in mild weather months. This fact reinforces the correlation of weather variables and total consumption of the building that was assumed in Section 2.2.1.

In order to analyse the spread of the consumption for each time of the week, first and third quantile values are shown as heat maps in Fig. 5. As it can be observed, the consumption significantly varies depending on the time of the day. The consumption during early hours is minimum, while at evenings the consumption is at its maximum. In addition, both heat maps show that electric consumption is distributed unevenly across the week, with somewhat different first and third quartile values for the same time of the day depending on the day of the week. In the weekends, the consumption is slightly higher during the day which could be due to a more intensive use of the building during these days. Overall, this validates the approach of including occupancy values as inputs of the model.

Table 3
K-fold cross-validation results of TOW-ARX model.

	MAE	RMSE	R^2
Fold 1	24.363	36.414	0.9656
Fold 2	23.927	36.705	0.9653
Fold 3	24.804	37.750	0.9629
Fold 4	25.063	36.790	0.9652
Fold 5	23.944	35.634	0.9675
Mean	24.363	36.414	0.9656
SD	0.508	0.762	0.0017

4. Results

4.1. Model description

The methodology presented in Section 2 is applied on the case study described in Section 3 to calibrate the TOW-ARX model. The model provides a one-hour ahead prediction of the consumption based on external and internal information of the building. Considering the number of submodels which form the model, only a subset of its parameters is presented in Table 2. The weight given to each variable in the submodels is assigned automatically to maximise the accuracy of each submodel following the process described in Section 2.4. Ambient temperature, solar irradiation, setpoint temperature, occupancy and consumption are represented by x_1, x_2, x_3, x_4 and y respectively in the model nomenclature. The equation of each of the submodels takes the form of Eq. (1), shown in Section 2.1. The total number of coefficients of the model is 1315.

4.2. Model validation

To ensure that the model does not overfit a cross-validation technique with k = 5 folds is applied. Given the results shown in Table 3, with similar metrics in both the mean value of the 5 folds and the model trained on the total dataset, it can be assumed that the model is not overfitted.

Plots of actual consumption value, and output and residuals for the model during August 2022 can be seen in Fig. 6.

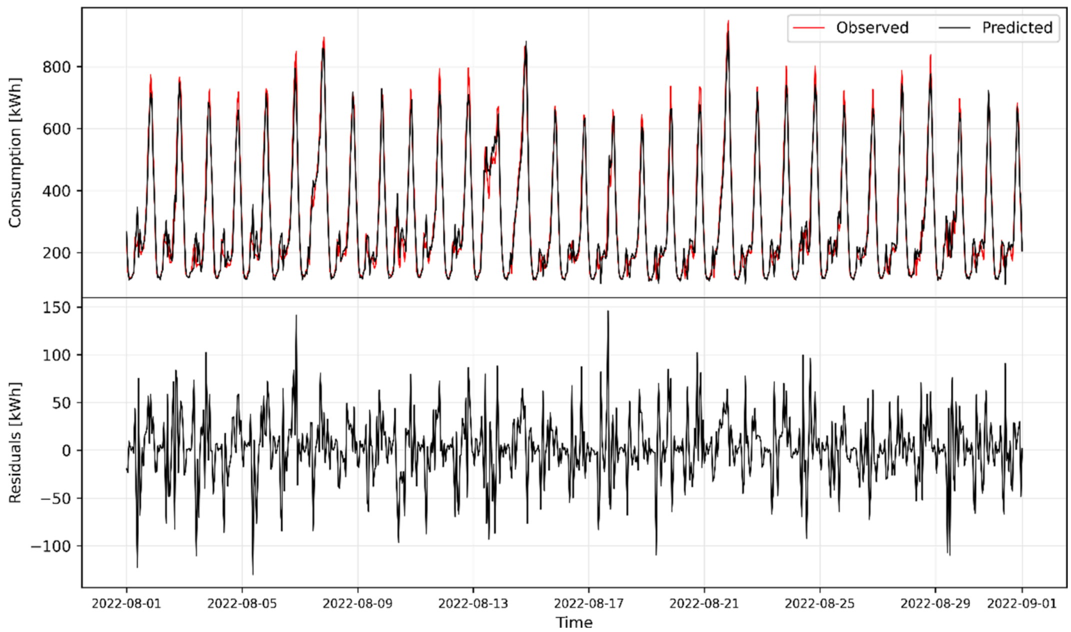

Fig. 6. Time-series plot of observed and predicted energy consumption (top), plot of the residuals for the model (bottom).

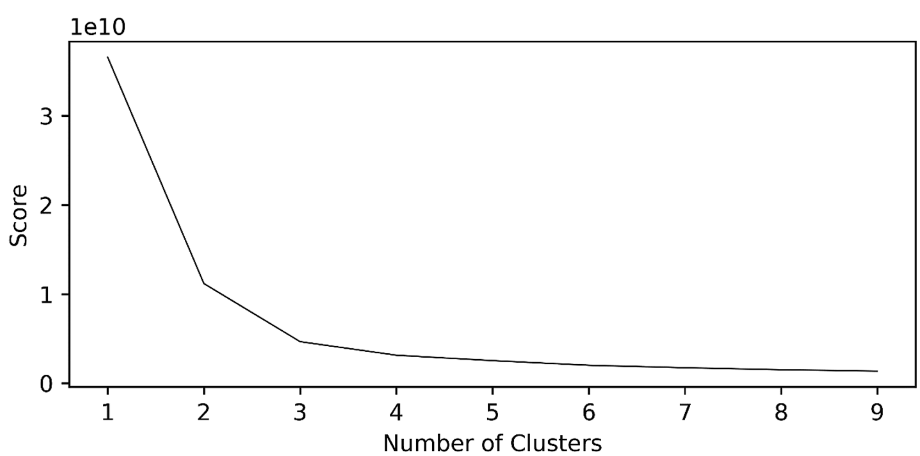

Fig. 7. Elbow curve used to determine the number of centroids (k) to use in the k-means clustering algorithm.

Table 4
K-fold cross-validation results of R-TOW-ARX model.

	MAE	RMSE	R^2
Fold 1	29.858	42.380	0.9524
Fold 2	29.614	41.621	0.9561
Fold 3	29.838	42.648	0.9535
Fold 4	29.026	41.358	0.9559
Fold 5	29.426	41.666	0.9551
Mean	29.552	41.935	0.9546
SD	0.343	0.550	0.0016

Table 5
K-fold cross-validation results of ARX model.

	MAE	RMSE	R^2
Fold 1	36.655	53.163	0.9262
Fold 2	35.409	51.440	0.9327
Fold 3	35.673	51.420	0.9318
Fold 4	35.061	50.444	0.9345
Fold 5	35.818	51.806	0.9303
Mean	35.723	51.655	0.9311
SD	0.595	0.983	0.0031

4.3. Model comparison

The model is compared to the 5 models exposed in Section 2.7.

In order to obtain the optimum number of centroids of the k-means clustering technique used to reduce the number of clusters of the TOW-ARX model, the elbow method [56] is applied. The elbow graph plots the within cluster sum of square values against different values of K. The within cluster sum of square value, as seen in Fig. 7, quickly decreases as the number of cluster K increases from 1 to 2. When K exceeds 3, the algorithm starts identifying similar clusters, and the decrease of the curve slows down. The chosen optimum number of centroids is equal to 3, resulting in a total of 32 submodels after applying the submodel reduction methodology.

The R-TOW-ARX and ARX models are checked through the k-fold cross-validation technique ($k = 5$). From the results shown in Table 4 and Table 5 it can be stated that none of these two models is overfitted.

Since the remaining models are simpler submodels of the 3 already tested, it can be assumed that they are not overfitted either. Actual

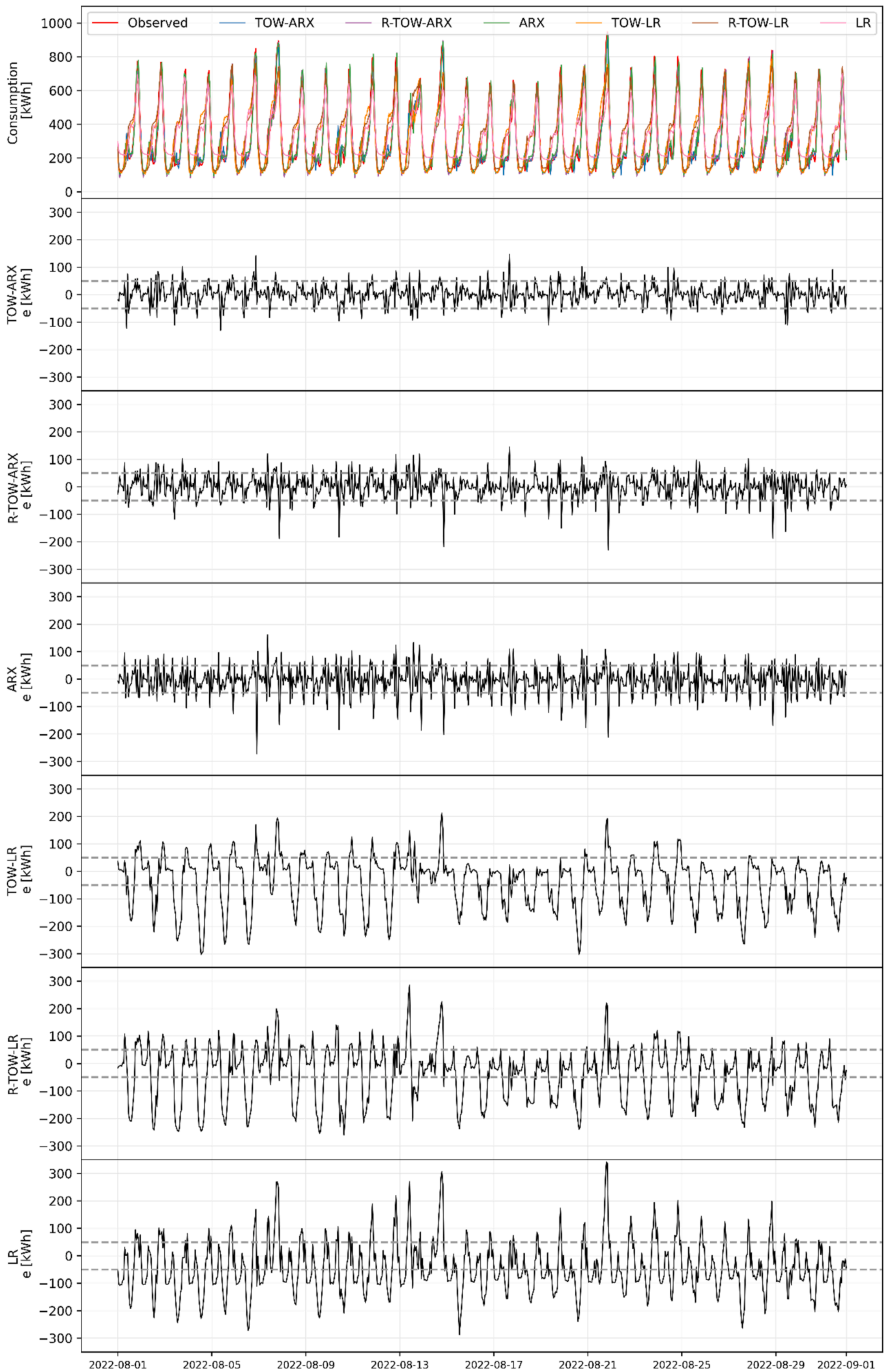

Fig. 8. The upper plot shows the predictions from each selected model versus the real consumption, and the following plots are of the residuals for each of the selected models. In order to facilitate comparison between models, a threshold is drawn at 50 and −50 kWh in the residuals plots.

Table 6
Evaluation metrics of the six models.

MODEL	n coeff.	R^2	MAE	RMSE
TOW-ARX	1315	0.9739	21.443	31.832
R-TOW-ARX	281	0.9570	28.732	40.815
ARX	23	0.9314	35.650	51.556
TOW-LR	672	0.7275	71.297	102.764
R-TOW-LR	128	0.6812	80.541	111.146
LR	4	0.5535	101.308	131.544

consumption, predicted output and residuals for the six models during August 2022 are presented in Fig. 8. Evaluation metrics calculated for the six models are shown in Table 6.

5. Discussion

With the exception of the LR model, the models present an overall good fit; $R^2 \approx 0.7$ in the case of TOW-LR and R-TOW-LR models, and $R^2 > 0.93$ in the models with ARX components. When comparing MAE and the RMSE of all six models, a similar trend can be observed, with the

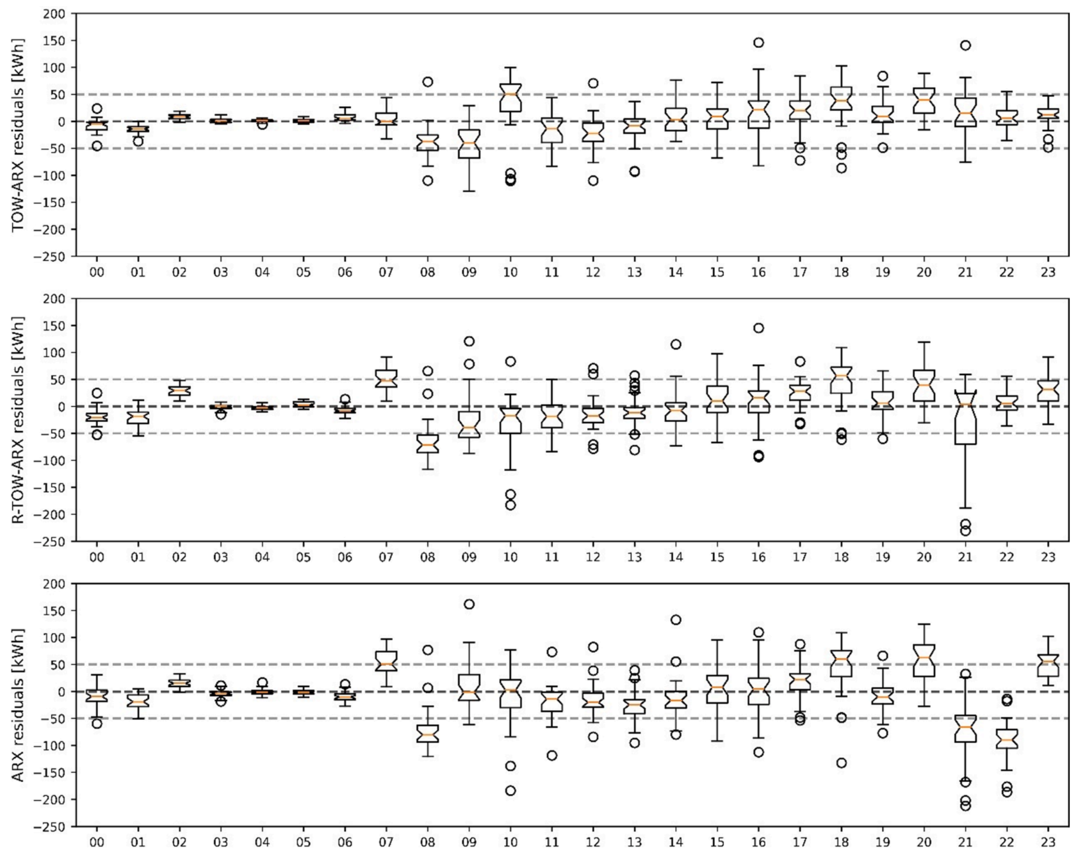

Fig. 9. Boxplot of the residuals of the TOW-ARX (top), ARX (bottom left) and R-TOW-ARX (bottom right) models grouped by hour of the day. In order to facilitate interpretability, a threshold is drawn at 50 and −50 kWh.

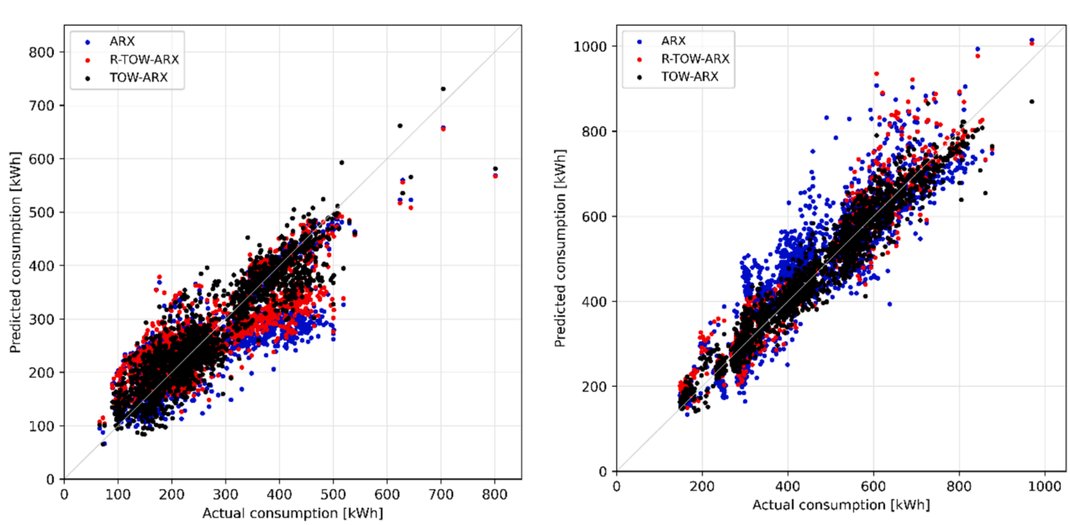

Fig. 10. Comparison between actual and predicted consumption values of the three ARX models during switch on (left) and switch off (right) periods.

three ARX-based models notably outperforming their LR counterparts. For both metrics, the error range of the LR models is approximately 2 to 5 times larger than for the ARX models, which is consistent with previous studies [31]. In addition, as shown in Fig. 8, the three ARX models greatly reduce the strong periodic bias that can be observed in the prediction errors of the three LR models. Given the number of coefficients, the precision metrics obtained, and the residuals' behaviour, it can be argued that the three ARX models are a better choice than TOW models considering the balance between computational effort and goodness of fit.

As Fig. 9 shows, the three best-fitting models have a similar accuracy through the dataset, with a poorer accuracy during the transient behaviour of the HVAC system of the mall in all three models. Moreover, the graphs show that, even if the overall accuracy of the three models is in a similar range, the TOW-ARX model's prediction error is more stable than those of the other two. Accuracy of the models during the main two transient periods of the HVAC system (08:00 to 10:00 during the switch on, and 21:00 to 23:00 during the switch off) is analysed. In Fig. 10 measured against predicted consumption values are plotted. The graphs show that, even if in those periods accuracy is lower than in the rest, TOW-ARX predictions on transient periods show a higher accuracy than those of R-TOW-ARX and ARX models.

In selecting the most suitable model, the trade-off between model size and goodness of fit must be considered. Extending a (general or time-indexed) LR model to ARX requires the inclusion of additional coefficients, but results from Table 6 appear to indicate that such additional coefficients are well worth in terms of model performance. In contrast, time-indexing a general model, using a TOW or clustered (R-TOW) indexing, notably multiplies the number of coefficients with a lower gain in model accuracy. Therefore, in formulating an efficient model, the inclusion of autoregressive and exogenous variables (ARX) should be considered before the time-indexing of the model. Certainly, the time-indexing further improves the goodness of fit, at the expense of increasing model size and complexity. The selection of the best model is not straightforward and should depend on the purpose of such model. If the aim is an efficient characterization of the thermal response of the building, arguably the simple ARX model is the most efficient in the use of resources (low number of coefficients) yet achieving a remarkable performance in characterizing dynamic thermal behaviour ($R^2 = 0{,}93$). For models with purely predictive purposes (e.g., used for model-predictive control) MAE is often used as a performance metric. The TOW-ARX method developed in this work outperforms the other five models in prediction capability, both over the overall datasets and in the cross validation.

The number of coefficients of the model has important implications on the required dataset size to prevent overfitting. Models with a low number of parameters (such as the general LR and ARX) can be successfully fitted with smaller datasets (spanning a few weeks) while the largest models (such as R-TOW-ARX or TOW-ARX) might require larger datasets ranging from several months to one year or more. In the present work, where two years of data have been used to fit the models, the three ARX-based models show little to no overfitting, with differences lower than 0.01 in R^2 when these are obtained from the whole dataset (Table 6) and from cross validation (Table 3, Table 4 and Table 5). In addition, it is important to remark that the proposed TOW-ARX model presents a large number of parameters when compared with the other 5 models, but it still has a low number of parameters when compared with commonly used machine learning models (e.g., LTSM and random forests models) [31 57].

6. Conclusions

This paper introduces a TOW-ARX model and applies it to predict the hourly energy consumption of a large shopping mall in Madrid (Spain). Model results are compared to 5 other models of reduced size. The goodness of fit of these models is assessed using standardized metrics such as the mean absolute error (MAE), the root mean square error (RMSE) and the coefficient of determination (R^2). In order to assess and discard the risk of model overfitting, cross validation is applied using the k-fold technique ($k = 5$). The proposed TOW-ARX model outperforms the other five presented models in all evaluation metrics. Moreover, it shows a better performance during transient periods of the HVAC system, straight after switch-on and switch-off.

The lack of extensive building datasets comprising both energy consumption and associated factors (weather, occupancy, indoor conditions) is one of the biggest barriers for the implementation of predictive techniques. The TOW-ARX method presented here achieves a very good fit ($R^2 > 0.96$) and can be used as an alternative to larger black-box models (e.g., supervised machine learning, reinforcement learning) when datasets of one year or more for the building are available (e.g., from smart meters), avoiding the use of synthetic data to train the models and the associated problems. A further advantage of the model is its capability to be applied to discontinuous datasets, with gaps in the data.

This study has been performed over data for a large commercial mall, and thus conclusions performance metrics might not be directly extrapolated to residential or non-residential buildings with different use patters (e.g., stronger weekday/weekend segmentation). The study is focused on an hour-ahead prediction ($t + 1$). Future work might look at performance over longer prediction horizons (e.g., 6, 12 or 24 h) comparing the iterative application of the methods and/or their fitting using larger timesteps.

CRediT authorship contribution statement

Iñigo Lopez-Villamor: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Olaia Eguiarte:** Writing – review & editing, Data curation. **Beñat Arregi:** Writing – review & editing, Supervision. **Roberto Garay-Martinez:** Writing – review & editing. **Antonio Garrido-Marijuan:** Writing – review & editing, Funding acquisition.

Funding

This study has been carried out in the context of SmartSPIN project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744. This publication reflects only the authors' views and neither the Agency nor the Commission are responsible for any use that may be made of the information contained therein.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

[1] "Energy efficiency – World Energy Outlook. – Analysis", IEA. Accessed: Apr. 09, 2024. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2019/energy-efficiency; 2019.
[2] *Directive (EU) 2023/… of the European Parliament and of the Council of 13 September 2023 on energy efficiency and amending Regulation (EU) 2023/955 (recast)*.
[3] Pombo O, Rivela B, Neila J. The challenge of sustainable building renovation: assessment of current criteria and future outlook. J Clean Prod Jun. 2016;123: 88–100. https://doi.org/10.1016/j.jclepro.2015.06.137.
[4] R. Bointner *et al.*, *Shopping malls features in EU-28 and Norway. Report elaborated under the FP7 project CommONEnergy: Re-conceptualising shopping malls from consumerism to energy conservation.* 2014.

[5] H. Schoenberger, M. J. L. Galvez, and D. Styles, "Best Environmental Management Practice in the Retail Trade Sector," JRC Publications Repository. Accessed: Apr. 09, 2024. [Online]. Available: https://publications.jrc.ec.europa.eu/repository/handle/JRC76036.

[6] Drgoňa J, et al. All you need to know about model predictive control for buildings. Annu Rev Control Jan. 2020;50:190–232. https://doi.org/10.1016/j.arcontrol.2020.09.001.

[7] "SAS Output." Accessed: Dec. 19, 2023. [Online]. Available: https://www.eia.gov/electricity/annual/html/epa_10_10.html.

[8] Tekler ZD, Ono E, Peng Y, Zhan S, Lasternas B, Chong A. ROBOD, room-level occupancy and building operation dataset. Build Simul Dec. 2022;15(12):2127–37. https://doi.org/10.1007/s12273-022-0925-9.

[9] Luo N, et al. A three-year dataset supporting research on building energy management and occupancy analytics. Sci Data Apr. 2022;9(1):156. https://doi.org/10.1038/s41597-022-01257-x.

[10] Lumbreras M, Diarce G, Martin K, Garay-Martinez R, Arregi B. Unsupervised recognition and prediction of daily patterns in heating loads in buildings. J Build Eng Apr. 2023;65:105732. https://doi.org/10.1016/j.jobe.2022.105732.

[11] Eguiarte O, Garrido-Marijuan A, Garay-Martinez R, Raud M, Hagu I. Data-driven assessment for the supervision of District Heating Networks. Energy Rep Dec. 2022;8:34–40. https://doi.org/10.1016/j.egyr.2022.10.212.

[12] Lumbreras M, Garay R, Marijuan AG. Energy meters in District-Heating Substations for Heat Consumption Characterization and Prediction Using Machine-Learning Techniques. IOP Conf Ser Earth Environ Sci Nov. 2020;588(3):032007. https://doi.org/10.1088/1755-1315/588/3/032007.

[13] Fels MF. PRISM: An introduction. Energy Build Feb. 1986;9(1):5–18. https://doi.org/10.1016/0378-7788(86)90003-4.

[14] J. K. Kissock, J. S. Haberl, and D. E. Claridge, "Development of a Toolkit for Calculating Linear, Change-Point Linear and Multiple-Linear Inverse Building Energy Analysis Models, ASHRAE Research Project 1050-RP, Final Report," Energy Systems Laboratory, Texas A&M University, Technical Report, Nov. 2002. Accessed: Apr. 09, 2024. [Online]. Available: https://oaktrust.library.tamu.edu/handle/1969.1/2847.

[15] Arregi B, Garay R. Regression analysis of the energy consumption of tertiary buildings. Energy Procedia Sep. 2017;122:9–14. https://doi.org/10.1016/j.egypro.2017.07.290.

[16] Lumbreras M, et al. Data driven model for heat load prediction in buildings connected to District Heating by using smart heat meters. Energy Jan. 2022;239:122318. https://doi.org/10.1016/j.energy.2021.122318.

[17] Liang X, Hong T, Shen GQ. Improving the accuracy of energy baseline models for commercial buildings with occupancy data. Appl Energy Oct. 2016;179:247–60. https://doi.org/10.1016/j.apenergy.2016.06.141.

[18] P. Price, "Methods for Analyzing Electric Load Shape and its Variability," Jun. 2010, Accessed: Apr. 10, 2024. [Online]. Available: https://escholarship.org/uc/item/8gf1w6q4.

[19] A. Hajri, R. Garay-Martinez, A. M. Macarulla, and M. A. B. Sassi, "Data-Driven Model For Heat Load Prediction In Buildings Connected To District Heating Networks," Sep. 05, 2023, *arXiv*: arXiv:2309.11504. doi: 10.48550/arXiv.2309.11504.

[20] Zhan S, Liu Z, Chong A, Yan D. Building categorization revisited: A clustering-based approach to using smart meter data for building energy benchmarking. Appl Energy Jul. 2020;269:114920. https://doi.org/10.1016/j.apenergy.2020.114920.

[21] Mustafaraj G, Chen J, Lowry G. Development of room temperature and relative humidity linear parametric models for an open office using BMS data. Energy Build Mar. 2010;42(3):348–56. https://doi.org/10.1016/j.enbuild.2009.10.001.

[22] Eguizabal M, Garay-Martinez R, Flores-Abascal I. Simplified model for the short-term forecasting of heat loads in buildings. Energy Rep Dec. 2022;8:79–85. https://doi.org/10.1016/j.egyr.2022.10.224.

[23] Thilker CA, Bacher P, Cali D, Madsen H. Identification of non-linear autoregressive models with exogenous inputs for room air temperature modelling. Energy AI Aug. 2022;9:100165. https://doi.org/10.1016/j.egyai.2022.100165.

[24] I. G. de Eulate, R. Garay-Marinez, B. A. Goikolea, O. Eguiarte, and A. M. Macarulla, "Simplified Geometric Processing of Solar Radiation for Improved Data-Driven Modelling of Short-Term Energy & Comfort Performance in Buildings," in *2024 9th International Conference on Smart and Sustainable Technologies (SpliTech)*, Jun. 2024, pp. 1–6. doi: 10.23919/SpliTech61897.2024.10612489.

[25] Guo Y, Nazarian E, Ko J, Rajurkar K. Hourly cooling load forecasting using time-indexed ARX models with two-stage weighted least squares regression. Energy Convers Manag Apr. 2014;80:46–53. https://doi.org/10.1016/j.enconman.2013.12.060.

[26] Touzani S, Granderson J, Fernandes S. Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy Build Jan. 2018;158:1533–43. https://doi.org/10.1016/j.enbuild.2017.11.039.

[27] C. M. R. Do Carmo and T. H. Christensen, "Cluster analysis of residential heat load profiles and the role of technical and household characteristics," *Energy Build.*, vol. 125, pp. 171–180, Aug. 2016, doi: 10.1016/j.enbuild.2016.04.079.

[28] Andersen FM, Larsen HV, Boomsma TK. Long-term forecasting of hourly electricity load: Identification of consumption profiles and segmentation of customers. Energy Convers Manag Apr. 2013;68:244–52. https://doi.org/10.1016/j.enconman.2013.01.018.

[29] Powell KM, Sriprasad A, Cole WJ, Edgar TF. Heating, cooling, and electrical load forecasting for a large-scale district energy system. Energy Sep. 2014;74:877–85. https://doi.org/10.1016/j.energy.2014.07.064.

[30] Sarwar R, Cho H, Cox SJ, Mago PJ, Luck R. Field validation study of a time and temperature indexed autoregressive with exogenous (ARX) model for building thermal load prediction. Energy Jan. 2017;119:483–96. https://doi.org/10.1016/j.energy.2016.12.083.

[31] Yun K, Luck R, Mago PJ, Cho H. Building hourly thermal load prediction using an indexed ARX model. Energy Build Nov. 2012;54:225–33. https://doi.org/10.1016/j.enbuild.2012.08.007.

[32] Chen Y, Guo M, Chen Z, Chen Z, Ji Y. Physical energy and data-driven models in building energy prediction: A review. Energy Rep Nov. 2022;8:2656–71. https://doi.org/10.1016/j.egyr.2022.01.162.

[33] Sun Y, Haghighat F, Fung BCM. A review of the-state-of-the-art in data-driven approaches for building energy prediction. Energy Build Aug. 2020;221:110022. https://doi.org/10.1016/j.enbuild.2020.110022.

[34] Lam JC, Li DHW. Electricity consumption characteristics in shopping malls in subtropical climates. Energy Convers Manag Jun. 2003;44(9):1391–8. https://doi.org/10.1016/S0196-8904(02)00167-X.

[35] Liu X, Yu J, Zhao A, Jing W, Mi L. A hybrid WOA-SVM based on CI for improving the accuracy of shopping mall air conditioning system energy consumption prediction. Energy Build Sep. 2023;294:113186. https://doi.org/10.1016/j.enbuild.2023.113186.

[36] "2428176.pdf.".

[37] Zhao A, Zhang Y, Zhang Y, Yang H, Zhang Y. Prediction of functional zones cooling load for shopping mall using dual attention based LSTM: A case study. Int J Refrig Dec. 2022;144:211–21. https://doi.org/10.1016/j.ijrefrig.2022.07.020.

[38] Jing W, Zhen M, Guan H, Luo W, Liu X. A prediction model for building energy consumption in a shopping mall based on Chaos theory. Energy Rep Nov. 2022;8:5305–12. https://doi.org/10.1016/j.egyr.2022.03.205.

[39] Senave M, Reynders G, Bacher P, Roels S, Verbeke S, Saelens D. Towards the characterization of the heat loss coefficient via on-board monitoring: Physical interpretation of ARX model coefficients. Energy Build Jul. 2019;195:180–94. https://doi.org/10.1016/j.enbuild.2019.05.001.

[40] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy Build May 2005;37(5):545–53. https://doi.org/10.1016/j.enbuild.2004.09.009.

[41] Kim Y-S, Srebric J. Impact of occupancy rates on the building electricity consumption in commercial buildings. Energy Build Mar. 2017;138:591–600. https://doi.org/10.1016/j.enbuild.2016.12.056.

[42] Martani C, Lee D, Robinson P, Britter R, Ratti C. ENERNET: Studying the dynamic relationship between building occupancy and energy consumption. Energy Build Apr. 2012;47:584–91. https://doi.org/10.1016/j.enbuild.2011.12.037.

[43] Massana J, Pous C, Burgas L, Melendez J, Colomer J. Short-term load forecasting in a non-residential building contrasting models and attributes. Energy Build Apr. 2015;92:322–30. https://doi.org/10.1016/j.enbuild.2015.02.007.

[44] Sánchez-Zabala VF, Gómez-Acebo T. Building energy performance metamodels for district energy management optimisation platforms. Energy Convers Manag X Jan. 2024;21:100512. https://doi.org/10.1016/j.ecmx.2023.100512.

[45] Alrebei O, et al. Quantifying CO2 Emissions and Energy Production from Power Plants to Run HVAC Systems in ASHRAE-Based Buildings. Energies Nov. 2022;15. https://doi.org/10.3390/en15238813.

[46] Pedregosa F, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12(85):2825–30.

[47] Schwarz G. Estimating the Dimension of a Model. Ann Stat Mar. 1978;6(2):461–4. https://doi.org/10.1214/aos/1176344136.

[48] Cavanaugh JE, Neath AA. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. Wires Comput Stat 2019;11(3):e1460.

[49] Song J, Zhang L, Xue G, Ma Y, Gao S, Jiang Q. Predicting hourly heating load in a district heating system based on a hybrid CNN-LSTM model. Energy Build Jul. 2021;243:110998. https://doi.org/10.1016/j.enbuild.2021.110998.

[50] Chen L, Basu B, McCabe D. Fractional order models for system identification of thermal dynamics of buildings. Energy Build Dec. 2016;133:381–8. https://doi.org/10.1016/j.enbuild.2016.09.006.

[51] Vallianos C, Athienitis A, Delcroix B. Automatic generation of multi-zone RC models using smart thermostat data from homes. Energy Build Dec. 2022;277:112571. https://doi.org/10.1016/j.enbuild.2022.112571.

[52] Ding Y, Zhang Q, Yuan T. Research on short-term and ultra-short-term cooling load prediction models for office buildings. Energy Build Nov. 2017;154:254–67. https://doi.org/10.1016/j.enbuild.2017.08.077.

[53] Kavaklioglu K. Robust modeling of heating and cooling loads using partial least squares towards efficient residential building design. J Build Eng Jul. 2018;18:467–75. https://doi.org/10.1016/j.jobe.2018.04.018.

[54] T. Malatesta and J. K. Breadsell, "Identifying Home System of Practices for Energy Use with K-Means Clustering Techniques," *Sustainability*, vol. 14, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/su14159017.

[55] "Historical Weather API | Weatherbit." Accessed: Apr. 10, 2024. [Online]. Available: https://www.weatherbit.io/api/historical-weather-api.

[56] Setiawan KE, Kurniawan A, Chowanda A, Suhartono D. Clustering models for hospitals in Jakarta using fuzzy c-means and k-means. Procedia Comput Sci Jan. 2023;216:356–63. https://doi.org/10.1016/j.procs.2022.12.146.

[57] D. Durand, J. Aguilar, and M. D. R-Moreno, "An Analysis of the Energy Consumption Forecasting Problem in Smart Buildings Using LSTM," *Sustainability*, vol. 14, no. 20, Art. no. 20, Jan. 2022, doi: 10.3390/su142013358.