

Calibration and validation of physics-based data-driven models for simulating the thermal behavior of indoor spaces in an assisted living facility

Italo Aldo Campodonico Avendano^{1,*}, *Farzad Dadras Javan*², *Behzad Najafi*², and *Amin Moazami*^{1,3}.

¹Department of Ocean Operations and Civil Engineering, NTNU, 6009 Ålesund, Norway

²Dipartimento di Energia, Politecnico di Milano, Via Lambruschini 4, 20156 Milano, Italy

³Department of Architectural Engineering, SINTEF Community, Børrestuveien 3, 0373 Oslo, Norway

Abstract.

A case study represented by an assisted living facility in Norway is modeled utilizing physics-based data-driven digital twin (DT) of the indoor thermal spaces with indoor temperature. Autoregressive Distributed Lag (ARDL), Machine Learning (ML), and Non-linear Autoregressive (NARX) models with time-series and sliding-window cross-validation are compared. Results show that NARX models have the highest accuracy, with a MAPE score of 0.03%. In addition, the sliding-window enhanced the models' accuracy and reduced the cyclical pattern for the autocorrelated values. The HVAC systems in this study case are representative of those found in Norwegian buildings, making the digital twin calibration applicable to other facilities.

1 Introduction

The clean energy transition has become a key factor toward the carbon neutrality expected in the European Union by 2050 [1]. From one side, a high adoption of renewable energy sources (RES) in European countries have reduced the dependency on fossil fuels [2]. However, RES are highly unpredictable, and their generation is susceptible to seasonal environmental conditions [3]. Therefore, the implementation of new technologies that allow flexibility of the grid, such as energy storage or demand-side management, are now a priority to successfully achieve the climatic targets [4]. Worldwide, buildings are one of the main energy users, reaching in the UE up to 42% in 2021 [5]. Following this fact, the electrification of the sector is expected as one of the pillars of the energy transition, and Norway is one of the examples of what can be expected in other countries shortly. Currently, 81% of the residential building sector in the country is already electrified [6], while 98% of the on-grid generation is clean. However, in 2023 is expected to be a shortage in energy production, which is paving and prioritizing the entrance of demand-side management as a real alternative to deal with energy congestion, balancing, and restoring the grid [7].

*Corresponding author: italo.a.c.avendano@ntnu.no

1.1 Digital twin for energy flexibility in buildings

Through the years, different techniques have been used to model the physics and the control of indoor spaces. However, several of these applications are mostly related to the design process and do not apply to real-time control scenarios due to the intrinsic physical background that needs highly computational resources. Therefore, the proliferation of sensor and smart meter data have enhanced the application of data-driven modeling techniques such as grey-box and black-box or machine learning (ML) modeling.

Developing a digital twin of the thermal zones of the buildings by using a data-driven approach gives the advantage of adapting the model when newer data is available. Grey-box modeling based on lumped models represents one of the most used approaches for creating DT since these models can deliver an easy understanding of the simplified physics of the thermal zones. In order to increase the accuracy of the models, Madsen and Holst [8] proposed the modeling of the zones with the implementation of stochastic differential equations (SDE), where the deterministic section, based on lumped systems, is associated with a stochastic process. Based on the assumption that indoor temperature sensors and energy meters are the most common sensors in indoor space, Bacher and Madsen [9], utilizing lumped systems based on SDE, conducted a wide study to analyze how the addition of hidden states helps in understanding how the different dynamics of the thermal zones need to be modeled.

Starting from the base previously mentioned, different works have been developed in the area utilizing this methodology. Nevertheless, the utilization of SDE and hidden states requires high computational costs that notably increase with larger training datasets. Also, passing from a continuous to a discrete process given by the space state equations makes the system lose the interpretability of the lumped systems due to the need to apply transform functions [8]. Following the same philosophy, Palmer Real et al. [10] defined the transformed capacitances and resistors of the lumped system are equivalent to the time constants of the spaces, which can be independently calculated by the natural temperature decay in the nighttime, when the outdoor temperature is constant and no heating and solar radiation is present. Utilizing this approach, Askeland et al. [11] modeled residential dwellings for flexibility estimation. The results of this work are then compared with white-box simulations of the dwellings utilizing IDA ICE.

A second possible path for developing data-driven DT is utilizing autoregressive models. A discretized first-order lumped ODE system that models the indoor temperature is indeed an autoregressive model of first order with exogenous components (ARX) [12]. The exogenous variables represent the influence of external entities such as ambient temperature, solar radiation, heating, and cooling usage, and others, while the first-order autoregressive constant shows the influence of the previously measured temperature in the next time step. In the latest years, ARX models have been widely utilized in model predictive control (MPC) [13, 14] due to its simplicity, high accuracy, lower computational costs, and the capacity for dynamic modeling, which can be properly integrated into convex optimization programming.

The modeling of thermal zones' states as well as energy usage in buildings was previously studied using black-box technics as Machine Learning (ML) and Deep Learning (DL). Papadopoulos et al. [15] showed in their work how ML, and in specific Random Forest (RF) Regressor algorithm with a sliding-window training scheme was utilized to manage indoor temperature logging failure at different acquisition frequencies with the minimum use of data. Similarly, other works have used RF and the same training scheme for energy forecasting in large buildings [16, 17].

A middle point between autoregressive and black-box modeling can be found in non-autoregressive exogenous (NARX) models. The hidden non-linear relations between the endogenous and exogenous components can be found through polynomial functions or by

using Artificial Neural Networks (ANN) with NARX architecture. Related works in MPC for building applications have been conducted utilizing this approach [18, 19].

In the scope of this work, Autoregressive Distributed Lag (ARDL) models, which corresponds to an extension of ARX models by including lagged values in the exogenous inputs, and ML time-series forecasting are utilized in the development of dynamic digital twins of an assisted living facility building. To complement this study, a predictive NARX pipeline based on ARDL and Extreme Gradient Boosting (XGBoost) Regressor will be assessed. In addition, two training approaches are considered: time-series cross-validation and sliding-window cross-validation, where the choice of one or the other will mainly differ in the amount of historical data needed for training the models. This study represents the beginning of the modeling process of the building to achieve a fully data-driven dynamic DT, which is expected to be utilized in the first instance for energy flexibility quantification and the test of energy efficiency measures.

2 Case study

The present work utilized an assisted living facility in Bærum, Norway, as a pilot. The complex is composed of three buildings with a total of $5,678 \text{ m}^2$, and the use of its inner thermal zones varies between bedrooms, shared spaces, storage, and offices, providing a variety of thermal uses with particular characteristics that can fall between residential and commercial buildings. The weather of the region is characterized by mild temperatures in summer and notably low temperatures in the winter season, creating a primary need for heating over cooling services. As well, the zone where is located the facility possesses a district heating system based on heat pumps; thus, the cost of the heating services is related to the day-ahead market electricity price. For this building, the heating system is fed by the previously mentioned DH, while the demand is coped by using an air handling unit (AHU) and hydronic radiators. For each thermal zone, information regarding indoor temperature, setpoints, and heating valve position is provided with a sub-hourly frequency of ten minutes. For the hydronic heating systems, energy meters are present in each main circuit at the building level. Forward temperature, return temperature, and volumetric flow is presented with the same sub-hourly frequency. Common spaces are ventilated, with CO_2 measurements, flows and supply and return temperature are given in an hourly basis.

In this work, eight thermal zones of the building are modeled based on the indoor temperature and utilizing the time-series data with an extension of the year calendar (2022). Two sampling frequencies are represented in the results: 10 minutes, and 1 hour. All the zones correspond to ventilated common spaces at neutral temperatures, while the remaining heating needs are provided by a radiator. In the first instance, the ventilation setpoint was set to 22°C , while the radiator setpoint can be set individually in each room and manually overwritten.

3 Methodology

In the present section, the methodology applied for the development of dynamic DT for energy flexibility purposes is explained.

3.1 Data-driven DT modeling

In the modeling of indoor temperature for indoor spaces, three approaches are taken. The first approach for developing the data-driven DT consists of defining an ARDL model (see Equation 1). In this specific case, the model will consider the previous three hours for the autoregressive component and four physics-inspired exogenous components with 3-hour lagged

values that emulate the possible heat gains in the indoor space (refer to Equation 2). The first feature corresponds to the interaction between the ambient (T_a) and the internal (T_i) temperature. The second feature considers the heating gains by the solar radiation (G_v). The third feature represents the injected air at a neutral temperature in the indoor space. Finally, the fourth feature considers the heating gains given by the hydronic radiator, as proposed by Thilker et al. [20]. At this point, the incoming flow to the radiator, the heating delivered, and the return temperature of the room (in hydronic and ventilation systems) are unknown. Similarly, the forward temperature (T_{for}) in the emitter is assumed to be the same from the generation side regardless of the losses and assuming a parallel configuration of the distribution systems.

$$\hat{y}_t = \sum_{t=0}^n Ay_{t-1} + \sum_{t=0}^n Bu_t + e_t \quad (1)$$

$$\mathbf{u} = \begin{bmatrix} T_a - T_i \\ G_v \\ \dot{V} \cdot T_{for-AHU} \\ f_{rad} \cdot (T_{for-rad} - T_i) \end{bmatrix} \quad (2)$$

The training process is carried out on a yearly basis utilizing the maximum likelihood estimation to obtain the most suitable autoregressive and exogenous parameters. As a result, vector A that contains the weights of the previous indoor temperature, and matrix B, which represents the weight of the three exogenous inputs in the same timeframe, are obtained. One of the particularities of this training process is that A and B are time-invariant, which explains its wide application on MPC. Consequently, a visualization of the modeling process is exemplified in Figure 1.

$$\begin{array}{c} \boxed{y_1 = Ay_0 + Bu_1} \\ \boxed{y_2 = A(Ay_0 + Bu_1) + Bu_2} \\ \vdots \\ \boxed{y_n = \sum_{t=0}^{n-1} Ay_t + \sum_{t=1}^n Bu_t} \end{array}$$

Figure 1: Example of the use of the time-invariant matrixes A and B for ARX models in the forecasting (and back-casting) process.

The second approach involves the utilization of the exogenous inputs, as well as the lagged values of the indoor temperature and encoded temporal features (hour of the day and day of the week) to train a machine learning (ML) model. In this specific case, XGBoost regressor is utilized. Next, the third approach consists of defining a non-linear autoregressive model (NARX). Since the model involves incoming airflow from the AHU, the indoor space is an open system; thus, the heating behavior has non-linear patterns that cannot be simply modeled with linear autoregressive models. The NARX model can be defined as shown in the following equation:

$$\hat{y}_t = f(y_{t-1}, y_{t-2}, y_{t-n}, \dots, u_t, u_{t-1}, u_{t-2}, u_{t-n}) \quad (3)$$

The linear behavior is already considered in the ARDL model. Therefore, the possible interactions between the endogenous and exogenous variables (function f) can be acquired with ML learning techniques. Consequently, the definitions of the NARX model will consist of a pipeline that first uses the ARDL model for predicting one step ahead of the indoor temperature and then uses the same XGBoost algorithm as in the second approach, but now adding the prediction gathered from the ARDL model. As a result, the ML algorithm will provide an adjusted one-step ahead prediction based on the original forecasting of the ARDL model.

3.2 Validation and model comparison

The training and validation procedure considers two cross-validation procedures but with different amounts of data in the training process. The first is time-series cross-validation (tsCV), where the training dataset incrementally increases when new data is available. The second cross-validation procedure is sliding-window cross-validation (swCV). Here the length of the training set remains constant (4 weeks of data), and new data is added while older data is discarded. In both cases, 100 folds are considered in the cross-validation, while the last 20% of the yearly time series is utilized for testing using the same approaches. After the models are trained, the residuals associated with the modeling procedure are gathered and the accuracy of the models is compared. The training and validation error is then represented utilizing the Mean Square Error (MSE), the Coefficient of Determination (R^2), and the Mean Absolute Percentage Error (MAPE).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (5)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (6)$$

Finally, autocorrelation plots and analysis of the samples of the residuals are presented for a complementary analysis of the models. It is expectable for each autocorrelation to have values the nearest possible to zero, with at least 95% of the peaks inside for the interval of confidence (5%) to be categorized as white noise. For independent and identical distributed (i.i.d) white noise, the histograms of the residuals are analyzed expecting a uniform gaussian distribution. Due to the extension of this work, only autocorrelation and histogram plots for one-hour frequency analysis are shown. A summary of the methodology applied in this work is presented in Figure 2.

4 Results and discussion

In the present section, the overall results of this work are presented. Aiming for data-driven dynamic digital twins of thermal spaces, three models that correspond to ARDL, ML (XG-Boost), and a hybrid pipeline for a NARX model are utilized for one-step ahead indoor temperature forecasting of eight thermal zones. These spaces have in common the use of ventilation to supply neutral heat and hydronic radiators to comply with the temperature requirements of the users. The overall average results considering yearly data of the spaces are shown in Table 1. Two sampling frequencies of 10 and 60 minutes and two methodologies

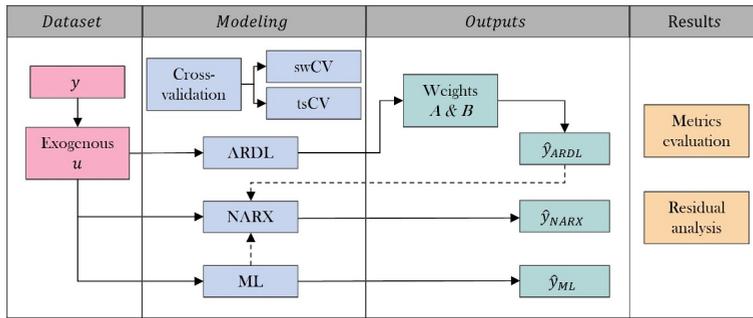


Figure 2: Simplified scheme of the methodology applied to this work.

for cross-validation: time-series and sliding-window, are analyzed for the proposed dataset. Starting with the sampling frequency of 10 minutes, it is observable high coefficients of determination and in the same order for ARDL and ML, and for both validation methodologies. However, notable differences are seen in MSE and MAPE, where ARDL shows better performance over the ML learning model. In the case of the NARX models, the coefficient of determination reaches the maximum (0.9999) between the studied models, while MSE and MAPE scores are two and one orders of magnitude lower, respectively, than the other models, showing consequently a better performance in the proposed NARX pipeline. Considering the validation and testing set, no major differences are found in the time-series cross-validation scores if the models are considered independent. Nevertheless, using four weeks of training (sliding-window cross-validation) shows a general decrease in the MSE and MAPE scores in all the models, as well as a noticeable improvement in the test dataset. Next, the increase in sampling frequency from 10 to 60 minutes also represents a decrease in the evaluation scores of the models. The reduced number of exogenous and endogenous coefficients for the 1-hour model has a visible impact on the scores of the ARDL model. Here, the MSE increases at least one order of magnitude, while the MAPE score increases four times in comparison with the 10-minute model. The coefficient of determination also was reduced from 0.99 in the validation for the time-series and sliding-window cross-validation, to 0.95 and 0.94, respectively. Even with a lower R^2 score for time-series CV, sliding-window CV reached lower scores in both MSE and MAPE, especially remarkable in the test set. For the ML model, a similar behavior to ARDL is observable for the time-series CV, but the use of sliding-window CV shows an improvement in all the scores, and in general, better performance than the ARDL model. Finally, the NARX pipeline is seen as the model that performs better, with an outstanding 0.999 and fraction in the coefficient of determination, one and two orders of magnitude lower than ML and ARDL, respectively, and an overall MAPE score of 0.03%. Similarly, to the other models, the use of sliding-window CV slightly improves the results. It is noteworthy to mention that the results MAPE of the ML modeling for 10- and 60-minute sampling frequency forecasting are in line with the results obtained by Papadopoulos et al. [15], where a similar approach was taken.

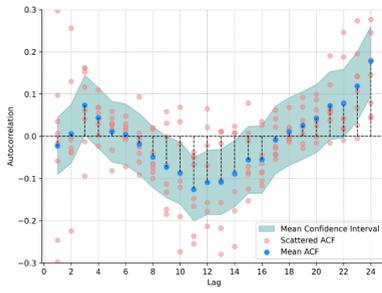
Due to the nature of this work, which is expected to use these models as the primary step for back-casting the energy usage of the building based on flexibility constraints, the models should be capable of covering the major influences of the external environment and heating equipment, leaving out the random process or white noise that cannot be modeled. To help in this process, the average autocorrelation plots of residuals of the studied cases are presented in Figure 3. Subfigures 3a and 3b present the autocorrelation plots of the ARDL model using

Table 1: Overall average results for the studied models and cross-validation procedures. The results include validation and test with 80% and 20% of the dataset, respectively.

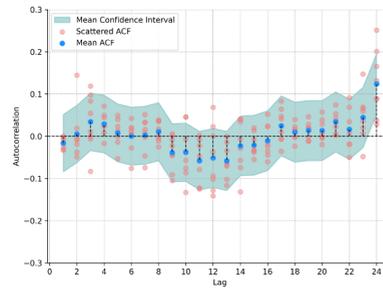
<i>Metric</i>	<i>Dataset</i>	ARDL		ML		NARX	
		<i>tsCV</i>	<i>swCV</i>	<i>tsCV</i>	<i>swCV</i>	<i>tsCV</i>	<i>swCV</i>
10 [minutes]							
MSE [10^4 °C]	<i>Validation</i>	20.918	18.765	72.99	39.072	0.701	0.6
	<i>Test</i>	18.725	8.006	68.511	19.703	0.634	0.281
R² [-]	<i>Validation</i>	0.9958	0.9944	0.9919	0.9928	0.9999	0.9999
	<i>Test</i>	0.9965	0.995	0.9928	0.984	0.9999	0.9999
MAPE [%]	<i>Validation</i>	0.11%	0.11%	0.19%	0.15%	0.03%	0.02%
	<i>Test</i>	0.10%	0.07%	0.18%	0.11%	0.03%	0.02%
1 [hour]							
MSE [10^4 °C]	<i>Validation</i>	304.609	257.8	200.616	60.267	1.131	1.133
	<i>Test</i>	273.886	129.403	231.001	39.08	1.013	0.584
R² [-]	<i>Validation</i>	0.956	0.9466	0.9745	0.9882	0.9998	0.9997
	<i>Test</i>	0.9625	0.9487	0.9723	0.9785	0.9998	0.9998
MAPE [%]	<i>Validation</i>	0.47%	0.42%	0.38%	0.21%	0.03%	0.03%
	<i>Test</i>	0.43%	0.32%	0.40%	0.17%	0.03%	0.02%

time series and sliding-window CV, respectively. In the first case, even when the average values are inside of the intervals of confidence, there exists a large dispersion in the initial and middle time lags. From this can be inferred that there is a cyclical pattern every 12 hours that the model is not capable of capturing. Therefore, it cannot be directly categorized as white noise residuals. By reducing the training data in the process with the sliding-window CV, it can be observed that the amplitude of the cyclical pattern has been reduced and just a few outliers are outside of the interval of confidence. Subfigure 3c shows the autocorrelation plot for the ML model with time-series cross-validation. Regarding ARDL with the same validation procedure, the cyclical pattern still exists but with lower amplitude. Similarly, the main outliers of the model are found in the first time-lags, showing the time dependence of the model to the previous endogenous values. By reducing the order of the training window, Subfigure 3d shows a noticeable improvement for the ML learning algorithm. Here, the cyclical behavior is reduced to its minimum and with just a few outliers outside of the intervals. Finally, Subfigures 3e and 3f shows the ACF plots for the NARX model using time-series and sliding-window CV. These two plots present, in overall, the best and most stable results and with no cyclical pattern in the autocorrelation and almost no outliers outside the average intervals of confidence. Therefore, can be determined that the residuals for both NARX model with 1-hour sampling frequency shows an independent behavior of the residuals, as well as the ML model with sliding-window CV.

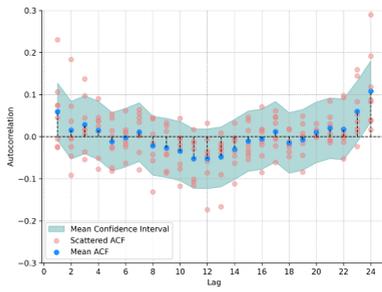
In Figure 4 are presented the histograms of the six studied cases. In general, all the cases present histograms with a gaussian distribution around mean zero. The exception is shown in Subfigure 4d, where the ML model with sliding-window CV presents some additional outliers' values in the positive tail of the histograms. For time-series CV, ML and ARDL present similar frequency and same order residuals. However, the NARX model has a narrower distribution and with outliers no further than ± 0.05 . Passing to sliding-window training scheme has help to narrow the distribution around zero in all the model. In general, the distribution of the residuals in all cases possess the characteristics of a gaussian function; thus, adding the



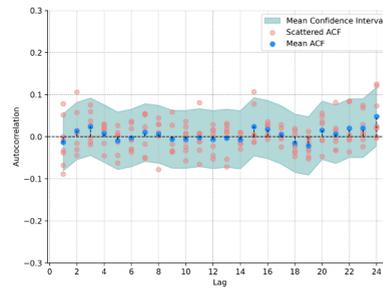
(a) ARDL model with Time-Series CV



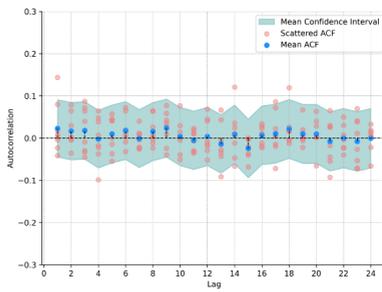
(b) ARDL model with Sliding-Window CV



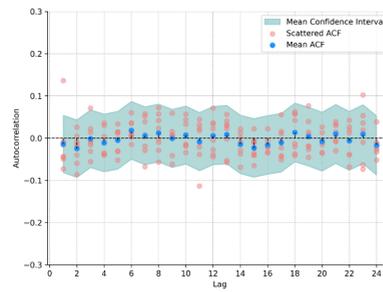
(c) ML model with Time-Series CV



(d) ML model with Sliding-Window CV



(e) NARX model with Time-Series CV



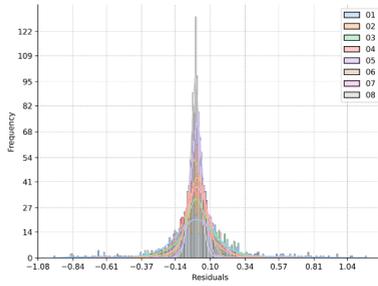
(f) NARX model with Sliding-Window CV

Figure 3: Autocorrelation plots for 1 hour sampling rate.

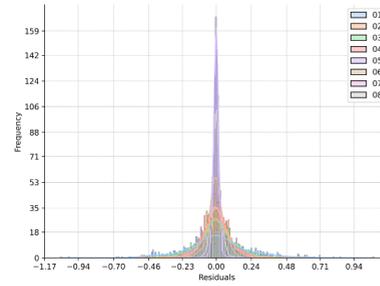
results obtained from the ACF, the noise obtained from the NARX models can be categorized as i.i.d. white noise.

5 Conclusion

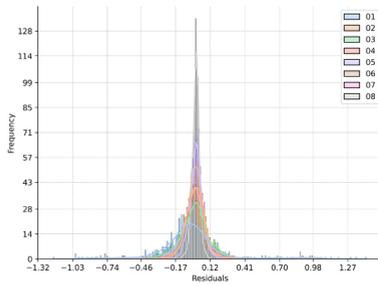
The present work compared three different models: ARDL, ML, and NARX, for the development of data-driven DT of indoor ventilated spaces in living care facilities. In addition, two training/validation procedures, corresponding to time-series and sliding-window cross-



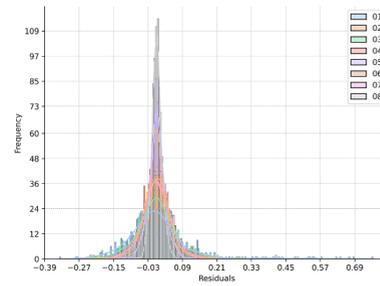
(a) ARDL model with Time-Series CV



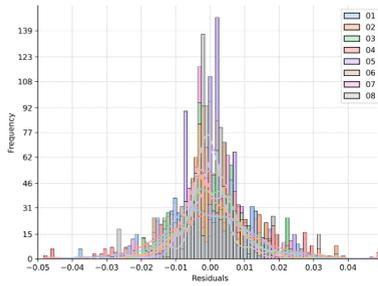
(b) ARDL model with Sliding-Window CV



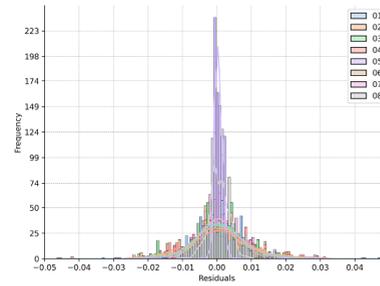
(c) ML model with Time-Series CV



(d) ML model with Sliding-Window CV



(e) NARX model with Time-Series CV



(f) NARX model with Sliding-Window CV

Figure 4: Histograms for 1 hour sampling rate.

validation, were applied in the different models. Overall, for the sub-hourly sampling rate, ARDL and ML models present high R^2 scores, but ARDL outperforms the ML model. The utilization of the NARX pipeline, which involves the ARDL and ML model obtained the best scores considering all the metrics, with a remarkable MAPE score of 0.023% in the validation with a sliding window approach. As it is expected, increasing the sampling rate also involves a reduction in the studied scores. The utilization of ML for modelling improves in comparison with the ARDL model. Nevertheless, the NARX model showed the best performance for one-hour step modelling. Passing from time-series to sliding-window CV showed a no-

ticeable improvement in the evaluation scores of the ARDL and ML models. Additionally, sliding-window CV helped to reduce the cyclicity of the autocorrelation of the residuals and to provide a more stable ACR among the sample zones. Finally, the NARX model outperformed the other models in both sampling rates, and its application also showed a more transversal performance between the different training processes. From the ACF analysis, can be observed that the residuals do not present any patterns and are independent. As well from the histograms, the residuals fall in a Gaussian distribution, with narrower values than other models; thus, can be inferred that the NARX model well-represents the physics of the thermal space, and the residuals can be classified as i.d.d. white noise.

5.1 Limitations and potential challenges in scalability and future work

The development of this work is still in the preliminary stages. The research delves into the specific case of ventilated spaces within living care facilities situated in Norway, where the inner thermal zones present varying thermal characteristics that would behave similarly to residential and commercial buildings. The region experiences mild summers and notably cold winters, resulting in a primary demand for heating rather than cooling services. However, the present work does not provide insight into other building types and does not cover hot and humid weather. In the present paper, only the models for ventilated spaces were presented, but there is still the need to develop models for spaces heated by only radiators and floor heating. Nevertheless, only the volumetric flows in the ventilated spaces are known, and not the specific flows to each hydronic distribution equipment. In addition, new dynamic models for the heating circuits that can account for the disaggregated heating delivered in each zone need further development to provide an overall DT of the building. This model is expected to be used for quantifying energy flexibility potential and evaluating energy efficiency strategies at the building level, which can be replicated in other study cases. Finally, the presented approach can be suggested as a tool that could seamlessly integrate with buildings already equipped with a building management system. The stored data will be used as input by the pipelines to make predictions and retrain the models. Nevertheless, additional research is necessary to optimize the flow of data, integrate computations effectively, and deliver the outcomes in a readily usable manner.

References

- [1] E.G. Deal, The european green deal, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Brussels (2020).
- [2] R. Puertas, L. Marti, Renewable energy production capacity and consumption in europe, *Science of the Total Environment* **853**, 158592 (2022). [10.1016/j.scitotenv.2022.158592](https://doi.org/10.1016/j.scitotenv.2022.158592)
- [3] A.C. Marques, J.A. Fuinhas, D.A. Pereira, Have fossil fuels been substituted by renewables? an empirical assessment for 10 european countries, *Energy policy* **116**, 257 (2018). [10.1016/j.enpol.2018.02.021](https://doi.org/10.1016/j.enpol.2018.02.021)
- [4] C. Bergaentzlé, C. Clastres, H. Khalfallah, Demand-side management and european environmental and energy goals: An optimal complementary approach, *Energy Policy* **67**, 858 (2014). [10.1016/j.enpol.2013.12.008](https://doi.org/10.1016/j.enpol.2013.12.008)
- [5] E. Commission, Energy performance of buildings directive (2023), accessed: 2023-12-11, https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en#documents

- [6] SSB, Production and consumption of energy, energy balance and energy account (2022), accessed: 2023-11-05, <https://www.ssb.no/en/energi-og-industri/energi/statistikk/produksjon-og-forbruk-av-energi-energibalanse-og-energiregnskap>
- [7] IEA, Norway 2022: Energy Policy Review. (IEA Publications, 2022)
- [8] H. Madsen, J. Holst, Estimation of continuous-time models for the heat dynamics of a building, *Energy and buildings* **22**, 67 (1995). [10.1016/0378-7788\(94\)00904-X](https://doi.org/10.1016/0378-7788(94)00904-X)
- [9] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, *Energy and buildings* **43**, 1511 (2011). [10.1016/j.enbuild.2011.02.005](https://doi.org/10.1016/j.enbuild.2011.02.005)
- [10] J.P. Real, C. Rasmussen, R. Li, K. Leerbeck, O.M. Jensen, K.B. Wittchen, H. Madsen, Characterisation of thermal energy dynamics of residential buildings with scarce data, *Energy and Buildings* **230**, 110530 (2021). [10.1016/j.enbuild.2020.110530](https://doi.org/10.1016/j.enbuild.2020.110530)
- [11] M. Askeland, L. Georges, M. Korpås, Low-parameter linear model to activate the flexibility of the building thermal mass in energy system optimization, *Smart Energy* **9**, 100094 (2023). [10.1016/j.segy.2023.100094](https://doi.org/10.1016/j.segy.2023.100094)
- [12] H. Madsen, P. Bacher, G. Bauwens, A.H. Deconinck, G. Reynders, S. Roels, E. Himpe, G. Lethé, Thermal performance characterization using time series data-iea ebc annex 58 guidelines (2015).
- [13] B. Merema, D. Saelens, H. Breesch, Demonstration of an mpc framework for all-air systems in non-residential buildings, *Building and environment* **217**, 109053 (2022). [10.1016/j.buildenv.2022.109053](https://doi.org/10.1016/j.buildenv.2022.109053)
- [14] C.A. Thilker, P. Bacher, D. Cali, H. Madsen, Identification of non-linear autoregressive models with exogenous inputs for room air temperature modelling, *Energy and AI* **9**, 100165 (2022). [10.1016/j.egyai.2022.100165](https://doi.org/10.1016/j.egyai.2022.100165)
- [15] D.N. Papadopoulos, F.D. Javan, B. Najafi, A.H. Mamaghani, F. Rinaldi, Handling complete short-term data logging failure in smart buildings: Machine learning based forecasting pipelines with sliding-window training scheme, *Energy and Buildings* **301**, 113694 (2023). [10.1016/j.enbuild.2023.113694](https://doi.org/10.1016/j.enbuild.2023.113694)
- [16] I.A. Campodonico Avendano, F.D. Javan, B. Najafi, A. Moazami, F. Rinaldi, Assessing the impact of employing machine learning-based baseline load prediction pipelines with sliding-window training scheme on offered flexibility estimation for different building categories, *Energy and Buildings* **294**, 113217 (2023). [10.1016/j.enbuild.2023.113217](https://doi.org/10.1016/j.enbuild.2023.113217)
- [17] F. Dadras Javan, I.A. Campodonico Avendano, B. Najafi, A. Moazami, F. Rinaldi, Machine-learning-based prediction of hvac-driven load flexibility in warehouses, *Energies* **16**, 5407 (2023). [10.3390/en16145407](https://doi.org/10.3390/en16145407)
- [18] F. Bünning, B. Huber, A. Schalbetter, A. AbouDonia, M.H. de Bady, P. Heer, R.S. Smith, J. Lygeros, Physics-informed linear regression is competitive with two machine learning methods in residential building mpc, *Applied Energy* **310**, 118491 (2022).
- [19] A. Erfani, T. Jafarinejad, S. Roels, D. Saelens, In search of optimal building behavior models for model predictive control in the context of flexibility, *Building Simulation* **17**, 71 (2024). [10.1007/s12273-023-1079-0](https://doi.org/10.1007/s12273-023-1079-0)
- [20] C.A. Thilker, P. Bacher, H.G. Bergsteinsson, R.G. Junker, D. Cali, H. Madsen, Non-linear grey-box modelling for heat dynamics of buildings, *Energy and Buildings* **252**, 111457 (2021). [10.1016/j.enbuild.2021.111457](https://doi.org/10.1016/j.enbuild.2021.111457)