Building Electrical Load Forecasting through Neural Network Models with Exogenous Inputs

Cristina Nichiforov*, Grigore Stamatescu*†, Iulia Stamatescu*, Ioana Făgărăşan* and Sergiu Stelian Iliescu*
*Department of Automatic Control and Industrial Informatics
University Politehnica of Bucharest, Romania
Email: cristina.nichiforov@aii.pub.ro
†Institute of Technical Informatics, Graz University of Technology, Austria
Email: gstamatescu@tugraz.at

***Abstract*—As buildings become key actors in the economic and sustainable operation of future electrical grids and smart cities, reliable models which capture the underlying electrical energy consumption become an important factor for robust control algorithms. Current ubiquitous field devices supported by complex data infrastructures allow generation, storage and online analysis of large quantities of data for deriving usable black-box models of building energy patterns. The paper presents an approach to model the energy consumption of medium and large sized buildings using Non-linear Autoregressive Neural Networks with eXogenous Input (NARX). We show that the chosen network architectures offers good performance for time series prediction from historical values and external input signals such as outdoor temperature in comparison to a baseline approach. Model evaluation and validation are carried out on public dataset for replicable research outcomes.**

***Index Terms*—neural networks, computational intelligence, smart buildings, energy forecasting.**

I. INTRODUCTION

Challenges related to the efficient energy management of dense electrical grids, especially in cities, stand to benefit from robust data-driven approaches based on validated prediction models of electrical demand. Many computational intelligence and statistical learning methods are now under development and in use by academic and industry researchers alike to capture non-linear consumption patterns of built environment entities. Increased availability and decreased cost of computing resources allows the rapid testing and validation of multiple methods on different reference data sets. Drawbacks still persist with regard to the ability of a black-box model to serve as a robust proxy for the underlying energy patters without periodical retraining and online model updates and the arbitrary parametrisation of such models which impedes replicable deployment for various scenarios.

In particular neural network (NN) models are seen as an important tool for developing black-box models based on data given their universal approximator characteristics suitable for non-linear behavior. The types of architectures that have been successfully implemented range from basic feed-forward structures with limited numbers of hidden layers, up to complex recurrent networks with many hidden layers. The applicability of such models lays for example in a predictive control algorithm, where the underlying optimization problem focuses on the mixed cost-comfort objective function. This can occur through simulations for assessing the impact of a predictive control strategy upon the controllable consumption ratio of a target building, with regard to comparable performance metrics. The underlying time series forecasting problem is a good match the electrical energy load forecasting task given that good quality datasets are available for training to compensate for the somewhat empirical and iterative nature of the design process.

With regard to design, implementation and on-line modeling of commercial building electrical loads using neural networks there are many suitable options. These range from both proprietary software packages towards open-source libraries and programming languages. Deployment on embedded real-time hardware or by using cloud based infrastructure is currently also a feasible solution given advances in computing, communication and standardized data access. Decision support systems can be designed and implemented for small scale renewable energy microgrids using the predictions [1].

Our work focuses on short term load forecasting (STLF) of electrical energy in large commercial buildings as key drivers of energy consumption in strained urban grids. The foreseen benefit of improved data-driven models for prediction is dual: providing good enough forecasts that enable reliable demand response schemes operation as well as replacing the need for fine grained energy monitoring at the building level. Finally the better accuracy obtained by the building operator in deploying such models offers significant leverage in negotiation supplier energy prices as well as provides for improved environmental impact. Accounting for the specific context: building design, energy source, usage patterns, several external factors potentially influence the building load curves. The dominant ones can be traced to outdoor climate and temperature variations and tiered energy pricing schemes. The inherent daily, weekly and seasonal periodic patterns are also dominant factors which are captured by the NN model.

Main contributions are argued below:

- the development of a NARX model for electrical energy consumption of large commercial buildings under outdoor temperature effects; extended modelling while considering additional discrete inputs to the model such as working hours, days of the week and weekends;
- the evaluation of the approach on a reference public large dataset of electrical energy usage with regard to

representative error metrics and to conventional Nonlinear Autoregressive Neural Network (NAR) models.

The rest of the paper is structured as follows. Section II presents closely related work which is relevant to the scope of the current contribution. Section III discusses the NARX models theoretical background in conjunction with the load forecasting application. Key experimental results are presented in Section IV for a reference commercial building dataset with multiple error metrics. Section V concludes the paper with outlook of future work to integrate the developed models for building microgrid energy flow control.

II. RELATED WORK

The context of the current work is mainly defined by the topic of Short-Term Load Forecasting (STLF) as it appears in the recent scientific literature and its application to automation of energy systems. Broad discussion of STLF techniques is carried out in [2]. The authors identify the main methods applicable for this task, ranging from conventional time series and polynomial modelling, towards computational intelligence inspired methods such as neural networks (NN) and support vector machines (SVM). A case study is provided to select the best modelling approach on an university building and finally an auto-regressive model (AR) is selected as offering the best performance in terms of mean absolute percentage error (MAPE).

A more specific application is covered by [3] where the authors compare NARX and support vector regression (SVR) for building energy forecasting with. The reported results yield a prediction accuracy between 93% and 85% for the three periods considered: day, week and month ahead. Input data is preprocessed according to standard primitives for outlier detection and handling of missing values and the load forecasting algorithm is designed for deployment on top of an existing IT infrastructure. Three types of features are identified as main determinants of the data-driven model: autoregressive features with different lags, outdoor temperature as well as contextual information which can relate to work schedules and other occupancy patterns. Several authors have successfully applied long short-term memory networks (LSTM) which can capture the inherent time dependencies corresponding to energy consumption modelling such as [4], [5]. For the first case the reported accuracy is 82.5% for day ahead forecasting while in the latter LTSM yields the best results among various other methods. In [6] the authors leverage multi-source data and provide a method for quantifying environmental factors and cluster residential consumers to improve the forecasting. A weather profile generation method is introduced by [7] and compared to the use of historical weather information. Data filtering and regrouping methods are shown to increase the performance of the forecast.

With regard to the application of the STLF models in control applications, an optimization model for predictive control of building microgrid energy flows is described by [8]. The load forecast black-box models can be leveraged for accurate predictions of the building energy requirements, based on which the optimal control actions for the controllable loads with most impact on the overall building consumption. A larger scale neighborhood level model predictive control scheme (MPC) is presented in [9] which analyses the benefits on the grid stability by allowing individual control of the main consumers. Related previous work concerned ARIMA [10], NN [11] and LSTM [12] modelling of building electrical energy consumption.

III. COMPUTATIONAL INTELLIGENCE TECHNIQUES FOR SHORT-TERM LOAD FORECASTING

A. Non-linear Autoregressive Neural Network with Exogenous Input model (NARX)

Since it is suited to model nonlinear dynamic systems, for the presented work it was proposed a NARX neural network which is derived by a class of discrete-time nonlinear systems. The nonlinear autoregressive network with exogenous inputs includes feedback connections enclosing several layers of the network. The NARX model is a nonlinear generalization of the Autoregressive Exogenous (ARX), which is commonly used as a standard instrument in linear black-box system identification for time-series.

The defining mathematical equation for a NARX model is:

$$\begin{aligned} y(t) = f(y(t-1), y(t-2), ..., y(t-n_y), \\ u(t-1), u(t-2), ..., u(t-n_u)) + \epsilon(t) \end{aligned} \tag{1}$$

where where $y(t)$ and $u(t)$ are the past values of the time series and present independent (exogenous) inputs of the NARX model at a time step t. n_y, $n_u \geq 1$, $n_y \geq n_u$ are the input and output delays, respectively, $\epsilon(t)$ is the error term and $f(\cdot)$ is a non-linear function.

A NARX model can be implemented using a feed-forward neural network to approximate the function f, Fig.1. For the current study it was used a parallel architecture which is presented in Fig.1. Also, the transfer function of the hidden layer neurons is a factor that is taken into account and for this research it is found that the best results can be obtained from a sigmoid function [13] [14]. Other salient option for the activation function is currently the hyperbolic tangent (tanh) rectifier linear unit (ReLU) function.

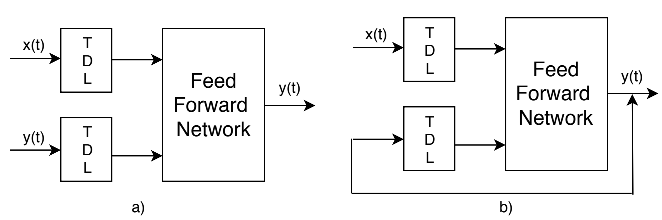

Fig. 1: Architecture of the NARX model: a) Series-parallel architecture b) Parallel architecture

B. Levenberg-Marquardt Training Algorithm

Regarding the training process, the Levenberg - Marquardt (LMBP) back-propagation learning algorithm has been used.

The LMBP algorithm is a variation of the classic Newton algorithm for discovery an optimum solution to a nonlinear minimization problem.

The LMBP update relationship is:

$$[J^T WJ + \lambda diag(J^T WJ)]h_{lm} = J^T W(y - \hat{y}), \qquad (2)$$

where J is the Jacobian matrix which contains first derivatives of the network errors with respect to network parameters, W is the the vector of weights and the values of λ are normalized to the values of $J^T WJ$ [15], [16].

The training procedure is controlled by a cross-validation technique which consists in dividing the initial dataset into three subsets. For this study, it was choose to define the three subsets as follows: 70% of the data were used for training the NARX model, 15% of the data were used for validation and the last 15% were used for testing the model.

IV. RESULTS EVALUATION

A. Choice of benchmark datasets

For the scope of the current study two data sets benchmarking datasets have been used. These stem from an online building energy data repository published by the Building and Urban Data Science (BUDS) Group at the National University of Singapore. These are part of a larger data collection effort from several hundreds of non-residential, mostly academic, buildings, proposed for performance analysis and algorithm benchmarking to the scientific community [17].

The input time series contain 8.760 data points representing the active energy consumed in the two buildings over a full calendar year. The buildings cover an indoor area of approximately 9.000 square meters and the outside temperature is provided alongside the energy data. The buildings are from two university campuses, one from Chicago (USA) and one from Zurich (Europe). The data were collected with a sampling time of one hour. These buildings were chose in conjunction to a local campus building at our university with similar size, mixed usage pattern including office, laboratory space, some classrooms, and also non-extreme temperate continental climate with four distinct seasons.

B. Choice of the NARX structures

This section describes the proposed solution: a forecasting NARX model that uses as exogenous inputs the weather conditions, more specific the outside temperature and as endogenous input the electrical load. The temperature is used as exogenous input because the literature has demonstrated that whether conditions such as wind speed or humidity have very small impact on the performance of the neural network [18], [19]. A diagram of the network is shown in Fig.2 where a two-layer feed-forward network is used for the model approximation.

Based on the data sets, there were proposed five configurations for NARX models in order to predict the energy load for each building. It was intended to have an evaluation of the forecasting performance of the various structures on the target data sets.

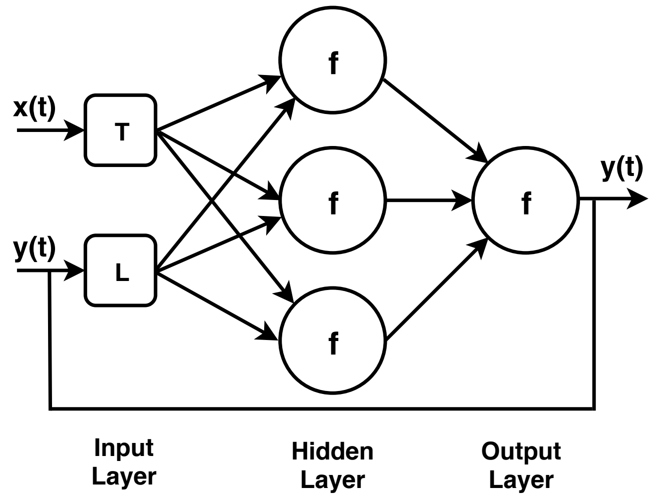

Fig. 2: Sample architecture of NARX model; inputs: temperature (T) & previous loads (L)

TABLE I: Defined networks for Zurich (Z) building data set

Network	No. of hidden layers	No. of neurons in the hidden layers	No. of output layers	No. of neurons in the output layers
Z1	1	[8]	1	1
Z2	2	[8, 16]	1	1
Z3	3	[8, 8, 16]	1	1
Z4	4	[8, 8, 16, 16]	1	1
Z5	5	[8, 8, 8, 16, 16]	1	1

TABLE II: Defined networks for Chicago (C) building data set

Network	No. of hidden layers	No. of neurons in the hidden layers	No. of output layers	No. of neurons in the output layers
C1	1	[8]	1	1
C2	2	[8, 16]	1	1
C3	3	[8, 8, 16]	1	1
C4	4	[8, 8, 16, 16]	1	1
C5	5	[8, 8, 8, 16, 16]	1	1

According to Table I and Table II the proposed networks are standard feed-forward neural networks with one input layer, different numbers and configurations of hidden layers and one output layer.

Fig.3 and Fig.4 show the forecasting response by a NARX model with four hidden layers with [8 8 16 16] neurons and the actual data from both Zurich and Chicago buildings. The prediction performance for each NARX model is good because the degree of matching between real data and predicted data is accurate for one step predicted output as can be seen in the following plots.

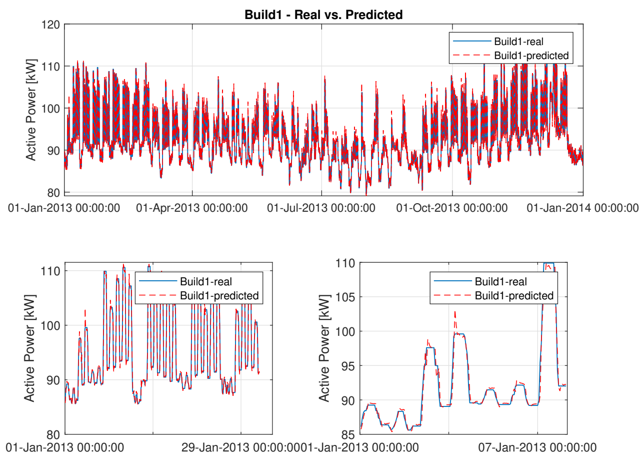

Fig. 3: Prediction result by NARX model - Zurich

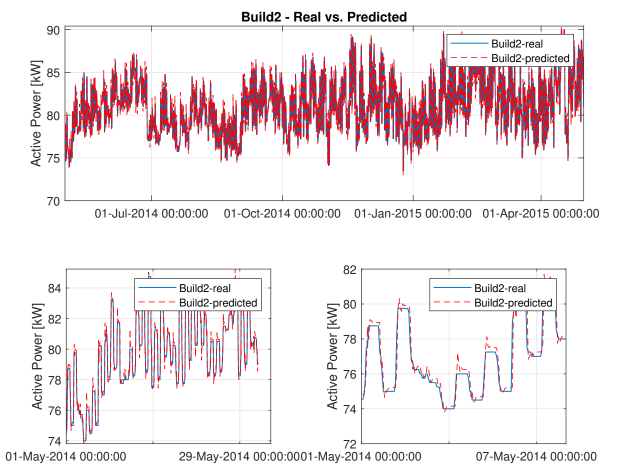

Fig. 4: Prediction result by NARX model - Chicago

C. Performance evaluation criteria

There were chosen six error indices to measure the accuracy of forecasting, which implies: Mean Absolute Error (MAE), Mean Squares Error (MSE), Mean Absolute Percentage Error (MAPE) and Mean Squared Percentage Error (MSPE). Each criteria is described by the following equations:

$$
\begin{aligned}
MAE &= \frac{\sum_1^n |Y_t - Yp_t|}{n} \\
MSE &= \sum_1^n \frac{(Y_t - Yp_t)^2}{n} \\
MAPE &= \frac{1}{n}\sum_1^n \left|\frac{Y_t - Yp_t}{Y_t}\right| 100 \\
MSPE &= \frac{1}{n}\sum_1^n \left(\left|\frac{Y_t - Yp_t}{Y_t}\right| 100\right)^2
\end{aligned} \tag{3}
$$

where n represents the number of samples, Y_t and Yp_t stand for the actual data and predicted data, respectively.

Equation 3 contains the error measures that are widely used for time series forecasting models. All these measures are grouped into two groups: absolute forecasting error (MSE, MAE) and percentage forecasting error (MAPE, MSPE). The absolute forecasting errors express average model prediction error in units of the variable of interest. This metrics can range from 0 to ∞and are indifferent to the direction of errors. Since they are negatively-oriented values, it means that lower values are better. The percentage forecasting errors are measures that indicates about the mean of the dispersion between predicted and observed value. Because absolute percentage errors are used, the problem of positive and negative errors canceling each other out is avoided. Also, in this case, the smaller the values the better the forecast.

The two following tables show the error values for different number and configuration of hidden layers.

TABLE III: Forecasting performance of each NARX - Zurich

	Z1	Z2	Z3	Z4	Z5
MAE	0.6173	0.6082	**0.5962**	0.6042	0.6155
MSE	1.2198	1.2004	**1.1607**	1.1765	1.2279
MAPE(%)	0.6472	0.6395	**0.6273**	0.6349	0.6453
MSPE(%)	1.3013	1.2799	**1.2431**	1.2582	1.3095

TABLE IV: Forecasting performance of each NARX - Chicago

	C1	C2	C3	C4	C5
MAE	0.4187	0.4147	**0.4124**	0.4159	0.4162
MSE	0.6227	0.5796	**0.5743**	0.5794	0.5837
MAPE(%)	0.5145	0.5099	**0.507**	0.5108	0.5114
MSPE(%)	0.9219	0.8584	**0.8514**	0.8572	0.8655

Because of the good results of the errors that can be seen in Table III and Table IV we can assume that the models are accurate. From the testing performance point of view, we observe that the smallest prediction errors for both the Chicago and Zurich building are obtained for the configuration with three hidden layers with 8, 8 and 16 neurons, respectively.

For this best configurations an extended NARX model has been proposed. There were proposed two models for each building with another three exogenous inputs namely: one dataset with the day number of the month, another one with the working hours intervals (08:00 - 20:00) and the third one with particular day i.e. Saturday and Sunday. The forecasting performance of these extended models can be seen in Table V. Comparing the results from Table III and IV with the ones from Table V it can be noticed that the error values are smaller in the case with three extra exogenous inputs. More specific, if we analyze Table VI it can be easily noticed that for the best network configuration the performances improved from the NAR model with one input dataset (the electric load) through the NARX with one exogenous input and NARX with four exogenous inputs. The NARX model with one exogenous input improved the MSE value with 48% and the NARX with 4 exogenous inputs with 56% for Zurich building and for Chicago Building with 24% and 36%, respectively.

TABLE V: Forecasting performance of extended-NARX for Z3&C3

	extended-Z3	extended-C3
MAE	0.5648	0.3949
MSE	0.9955	0.5076
MAPE(%)	0.5973	0.4855
MSPE(%)	1.0755	0.7533

TABLE VI: MSE values for NAR, NARX and extended-NARX models for Z3 & C3

Network	MSE - NAR	MSE - NARX	MSE - extended-NARX
Z3	2.2499	1.1607	0.9955
C3	0.757	0.5743	0.4855

In a previous work, [10] we developed several energy forecasting NAR models for Zurich and Chicago buildings. Regarding the performance comparison between NAR and NARX models, Figure 5, Figure 6 and also Table VII show that the NARX models provide smaller errors which means a more accurate prediction.

TABLE VII: MSE values for NAR and NARX models for Zurich (Z) and Chicago (C) building data sets

Network	MSE - NAR	MSE - NARX
Z1	2.2546	1.2198
Z2	2.2525	1.2004
Z3	2.2499	1.1607
Z4	2.2485	1.1765
Z5	2.2510	1.2279
C1	0.788	0.6227
C2	0.777	0.5796
C3	0.757	0.5743
C4	0.738	0.5794
C5	0.766	0.5837

The mean and standard deviation of mean squared error values for the two models (NAR and NARX) were calculated. As can be seen in Table VIII, the series shows values that are tightly clustered around the mean value which leads to a low dispersion value.

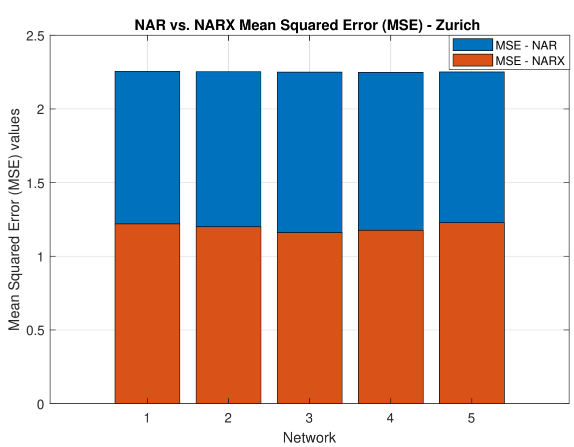

Fig. 5: NAR vs. NARX Mean Squared Error metric - Zurich

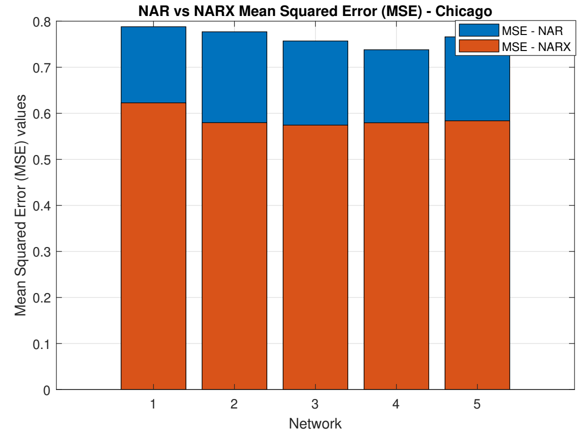

Fig. 6: NAR vs. NARX Mean Squared Error metric - Chicago

TABLE VIII: MSE values for NAR and NARX models for Zurich (Z) and Chicago (C) building data sets

Building	Mean	Standard deviation
Zurich MSE - NAR	2.2513	0.0024
Zurich MSE - NARX	1.1971	0.0284
Chicago MSE - NAR	0.7652	0.0191
Chicago MSE - NARX	0.589	0.0197

For the current research all implementation, modelling and validation of the approach has been carried out on a 3.4 GHz i7 quad core processor and 16GB RAM. Software implementation has been deployed in Matlab and can be provided to interested readers on demand along with the processed datasets for replicable research.

V. CONCLUSIONS

The nonlinear autoregressive network with exogenous input (NARX) is used to perform load forecasting for two medium to large size non-residential buildings. The outside temperature has been considered as the exogenous component in the analysis and the electrical load as endogenous input in the first stage. We have subsequently extended the model by adding further inputs for working time, day of the week and weekends. The NARX models that we proposed improve the accuracy of energy forecasting comparing to NAR models in terms of error metrics and also with regard to the complexity of the model architecture, the smallest error values were obtained with the 3 hidden layers NARX model, which means that unlike the best NAR model with 4 hidden layers, the complexity of the architecture has diminished.

Further research is currently underway to use the black-box models within predictive modeling in order to control building energy flows and implement several load management strategies e.g. by modulating chiller output power in conjunction to energy prices and weather variations or other events that can occur unexpectedly.

REFERENCES

[1] I. Stamatescu, N. Arghira, I. Făgărăşan, G. Stamatescu, S. S. Iliescu, and V. Calofir, "Decision support system for a low voltage renewable energy system," *Energies*, vol. 10, no. 1, 2017. [Online]. Available: http://www.mdpi.com/1996-1073/10/1/118

[2] I. Fernández, C. E. Borges, and Y. K. Penya, "Efficient building load forecasting," in *ETFA2011*, Sep. 2011, pp. 1–8.

[3] N. K. Thokala, A. Bapna, and M. G. Chandra, "A deployable electrical load forecasting solution for commercial buildings," in *2018 IEEE International Conference on Industrial Technology (ICIT)*, Feb 2018, pp. 1101–1106.

[4] N. Kim, M. Kim, and J. K. Choi, "Lstm based short-term electricity consumption forecast with daily load profile sequences," in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, Oct 2018, pp. 136–137.

[5] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on lstm recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan 2019.

[6] Y. Wang, M. Liu, Z. Bao, and S. Zhang, "Short-term load forecasting with multi-source data using gated recurrent unit neural networks," *Energies*, vol. 11, no. 5, 2018. [Online]. Available: http://www.mdpi.com/1996-1073/11/5/1138

[7] G. Zhu, T.-T. Chow, and N. Tse, "Short-term load forecasting coupled with weather profile generation methodology," *Building Services Engineering Research and Technology*, vol. 39, no. 3, pp. 310–327, 2018. [Online]. Available: https://doi.org/10.1177/0143624417740858

[8] D. Marušić, V. Lešić, T. Capuder, and M. Vasak, "Price-optimal energy flow control of a building microgrid connected to a smart grid," in *2018 26th Mediterranean Conference on Control and Automation (MED)*, June 2018, pp. 1–9.

[9] A. Mirakhorli and B. Dong, "Model predictive control for building loads connected with a residential distribution grid," *Applied Energy*, vol. 230, pp. 627 – 642, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261918312066

[10] C. Nichiforov, G. Stamatescu, I. Stamatescu, I. Făgărăşan, and S. S. Iliescu, "Intelligent load forecasting for building energy management systems," in *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, June 2018, pp. 896–901.

[11] C. Nichiforov, G. Stamatescu, I. Stamatescu, V. Calofir, I. Fagarasan, and S. S. Iliescu, "Deep learning techniques for load forecasting in large commercial buildings," in *2018 22nd International Conference on System Theory, Control and Computing (ICSTCC)*, Oct 2018, pp. 492–497.

[12] C. Nichiforov, G. Stamatescu, I. Stamatescu, and I. Făgărăşan, "Evaluation of sequence-learning models for large-commercial-building load forecasting," *Information*, vol. 10, no. 6, p. 189, Jun 2019. [Online]. Available: http://dx.doi.org/10.3390/info10060189

[13] J. Buitrago and S. Asfour, "Short-term forecasting of electric loads using nonlinear autoregressive artificial neural networks with exogenous vector inputs," *Energies*, vol. 10, no. 1, 2017. [Online]. Available: http://www.mdpi.com/1996-1073/10/1/40

[14] Z. Boussaada, O. Curea, A. Remaci, H. Camblong, and N. M. Bellaaj, "A nonlinear autoregressive exogenous (narx) neural network model for the prediction of the daily direct solar radiation," *Energies*, vol. 11, no. 3, 2018. [Online]. Available: http://www.mdpi.com/1996-1073/10/1/40

[15] A. Reynaldi, S. Lukas, and H. Margaretha, "Backpropagation and levenberg-marquardt algorithm for training finite element neural network," in *2012 Sixth UKSim/AMSS European Symposium on Computer Modeling and Simulation*, Nov 2012, pp. 89–94.

[16] K. Madsen, N. Nielsen, and O. Tingleff, "Methods for nonlinear least squares problems. technical report." Informatics and Mathematical Modelling, Technical University of Denmark, Tech. Rep., 2004.

[17] C. Miller and F. Meggers, "The building data genome project: An open, public data set from non-residential building electrical meters," *Energy Procedia*, vol. 122, pp. 439 – 444, 2017.

[18] K. Nahi, W. Rene, S. Maarouf, and G. Serman, "International journal of electrical power & energy systems," *Elsevier*, vol. 28, no. 8, pp. 525 – 530, 2006.

[19] B. Jaime and A. Shihab, "Short-term forecasting of electric loads using nonlinear autoregressive artificial neural networks with exogenous vector inputs," *Energies*, vol. 10, no. 1, 2017.