 doi:10.20944/preprints201904.0318.v1

Article

Evaluation of Sequence Learning Models for Large Commercial Building Load Forecasting

Cristina Nichiforov [1], Grigore Stamatescu [1,2], Iulia Stamatescu[1], Ioana Făgărăşan[1]

[1] Department of Automatic Control and Industrial Informatics, Faculty of Automatic Control and Computers, University Politehnica of Bucharest, 313 Splaiul Independentei, 06004 Bucharest, Romania; cristina.nichiforov@aii.pub.ro (C.N.); iulia.stamatescu@aii.pub.ro (I.S.); ioana@shiva.pub.ro (I.F.)

[2] Institute of Technical Informatics, Technical University of Graz, 16 Inffeldgasse, 8010 Graz, Austria;

* Correspondence: gstamatescu@tugraz.at; Tel.: +40-723-425-323

Abstract: Buildings have started to play a critical role in the stability and resilience of modern smart grids, leading to a refocusing of large scale energy management strategies from the supply side to the consumer side. When the buildings integrate local renewable energy generation in the form of renewable energy resources they become prosumers and this reflects into additional complexity into the operation of the interconnected complex energy systems. A class of methods of modelling the energy consumption patterns of the building have recently emerged as black-box input-output approaches with the ability to capture underlying consumption trends. These make use and require large quantities of quality data produces by non-deterministic processes underlying the energy consumption. We present an application of a class of neural networks, namely deep learning techniques for time series sequence modelling with the goal of accurate and reliable building energy load forecasting. The Recurrent Neural Network implementation uses Long Short-Term Memory layers in increasing density of nodes to quantify prediction accuracy. The case study is illustrated on four university buildings from temperate climates over one year of operation using a reference benchmarking dataset that allows replicable results. The obtained results are discussed in terms of accuracy metrics and computational and network architecture aspects and are considered suitable for further used in future in situ energy management at the building and neighbourhood levels.

Keywords: sequence models; recurrent neural networks; energy modelling; smart buildings.

---

1. Introduction

Complex energy systems which support global urbanisation tendencies are playing an important role in the definition, implementation and evaluation of future smart cities. Within the built environment, making use of modern technologies in the areas of sensing, computing, communication and control leads to an improvement the operations of various systems and the well-being of its inhabitants. Of particular relevance to the reliable, clean and cost-effective energy supply electrical grid, thereby assuring a to ever increasing urban needs. The main target of our work is developing improved models for load forecasting of medium and large commercial buildings which act in a determining role as consumers, prosumers or balancing entities for grid stability. Statistical learning algorithms such as classical and deep neural networks represents a prime example. Black-box model achieved thorough such techniques have proven able to accurately model underlying patterns and trends driving energy consumption can be used to forecast load profiles and improve high level control strategies. In such way significant economic, through cost savings, and environmental, through limited use of scarce energy resources, benefits can be achieved.

An often quoted figure in the scientific literature [1] places the building energy use at almost 40% of primary energy use in many developed countries with an increasing trend. In modern buildings a centralised software solution, often denoted as a Building Management System (BMS), collects all the relevant data streams originating in the building subsystems and provides means for intelligent algorithms to act based on the processed data. Energy-relevant data, generated through heterogeneous instrumentation networks is subsequently leveraged to control relevant energy parameters in daily operation. For the existing older building, new technology can be used to upgrade legacy devices such as electrical meters and integrate them into wired or wireless communication networks, in a cost effective manner. Finally the resulting preprocessed energy measurements time series serve as input for accurate models of energy prediction and control.

Statistical learning algorithms have become robust and well-adopted in the last years through concentrated efforts in both research and industry. This was stimulated by data availability and exponentially increasing computing resources at lower costs, including cloud systems. Neural networks are one example of such algorithm which offers good results in many application areas, including modelling and prediction of energy systems. This is valid for both classification tasks, as well as regression tasks where the objective is to predict an output numerical value of interest. Deep learning networks are highly intricate neural networks with many hidden layers that are able to learn patterns of increasing complexity in input data sets. Initially deployed through industry driven initiatives in the areas of multimedia processing and translation systems, other technical applications currently stand to benefit from the availability of open-source algorithms and tools. For time series and sensor data, a particular types is gaining adoption with the research community namely sequence models based on recurrent neural network that can capture long term dependencies in input examples. In the nomenclature described by [2] of machine learning (ML) taxonomy for smart buildings, our work fits within the area of using ML to estimate aspects related to either energy or devices, in particular energy profiling and demand estimation.

Within this approach, large commercial buildings provide the operators/owners with the economic incentives and return of investment related to energy efficiency projects where small percentage gains on large absolute values of energy use become more attractive. An equally large market exist for improving energy forecasting accuracy in the residential sector, which is however more fragmented and the incentives to deploy such approaches have to be present at the energy supplier or through public large scale programs.

Main contributions of the paper can be thus summarised:

- illustrating a deep learning approach to model large commercial building electrical energy usage as alternative to conventional modelling techniques;
- presenting an experimental case study using the chosen deep learning techniques enabling reliable forecasting of building energy use;
- analysis of the results in terms of accuracy metrics, both absolute and relative which provide a way for replicable results towards other related research.

Additional contributions which extend the previous conference paper [3] are summarised. We have provided, as main goal for the extended version, new experimental results for recurrent neural network modelling of large commercial building energy consumption. These are further analysed also taking into account several performance metrics and computational aspects. More technical clarifications regarding the methods and data processing and modelling pipeline are also included. Also, significant revisions and extensions have been carried out to the related work section for a more timely and focused state-of-the-art to frame the work as well as to other parts of the paper to improve readability and allow the replication of the results by interested researchers on the neural network architectures presented in an energy management system.

The rest of the paper is structured as follows. Section II discusses timely related work for energy consumption modelling of direct relevance to the previously stated contribution areas. Section III presents the theoretical background of Recurrent Neural Networks (RNN) implemented with layers of Long Short-Term Memory (LSTM) units and their application for this task. A case study is described in Section IV by applying the deep learning techniques on a reference benchmarking dataset for four large commercial buildings. The main findings are also discussed in detail, including computational

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2019 doi:10.20944/preprints201904.0318.v1

aspects pertaining to the architectures of the learning algorithms that were implemented. Section V concludes the paper with regard to applicability of the derived black-box models for in situ electrical load forecasting.

2. Related Work

We first state three key factors identified as drivers of the application of statistical learning techniques and algorithms to energy system modelling and control:

- better availability of good quality datasets and computational resources that enable extensive testing and validation of the proposed methods;
- commercial and open-source algorithm libraries and software with suitable documentation and examples for a wider audience;
- increased collaboration between algorithm, computing and control experts and domain specialists in the energy and civil engineering fields; this has influenced the design of new deep learning architectures customised mostly for particular applications.

[4] describe an approach and case study for prediction of household energy consumption using feed-forward back propagation neural networks. The authors discuss their outcomes based on data collected from four residential buildings in South Korea, including the preprocessing and tuning of the algorithms. The evaluation is based on models built on raw data, normalised data and data with statistical moments with the conclusion that the accuracy metrics are best in the latter case. Learning customer behaviours for effective load forecasting is discussed by [5] who implement Sparse Continuous Conditional Random Fields (sCCRF) to effectively identify different customer behaviours through learning. A hierarchical clustering process is subsequently used to aggregate customers according to the identified behaviours.

In [6] the authors focus on two deep learning techniques for building energy consumption namely Conditional Restricted Boltzmann Machines (CRBM) and Factored Conditional Restricted Boltzmann Machines (FCRBM). The results are illustrated in comparison to artificial neural networks and support vector machines (SVM) as well as recurrent neural networks. It is shown that for many of the test scenarios: aggregated and submetering data as well as different time horizons and resolution, the investigated methods offer better performance in terms of RMSE on the prediction horizon. Time series change point detection along with preprocessing approaches are introduced in [7]. SEP algorithm is evaluated on smart home data, both real and artificial. This uses separation distance as a divergence measure to detect change points in high-dimensional time series. Building occupancy influence on energy consumption through indirect sensing is elaborated upon by [8] where the system is able to infer occupancy counts using "proxy" measures of temperature and indoor CO2 concentrations.

The authors of [9] present an end-to-end deep learning approach for load forecasting of commercial buildings by combining stacked autoencoders (SAE) with extreme learning machines (ELM). SAE extracts the relevant features from the input dataset while ELM works as the predictor. The advantages of this particular method are justified in terms of achieving directly the output weights of the networks without iterative backpropagation. Two RNN models based on LSTM units for building energy consumption are evaluated in [10]. An important finding of this study is that RNN methods tend to perform better than others when applied to aggregated energy data and long term dependencies are more difficult to identify. Also the authors use the resulting model to perform missing value imputation on the original time series. [11] discuss the application of auto-encoders and generative networks as deep learning alternative to conventional feature engineering in learning models for electrical energy load forecasting. A method based on Support Vector Regression (SVR) is presented by [12].

While defining the context of the current work, we mention also previous implementation which analysed conventional system identification using Auto-Regressive Integrated Moving Average (ARIMA) models versus classical Artificial Neural Networks (ANN) [13]. Multiple ANN versions

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2019 doi:10.20944/preprints201904.0318.v1

have been tested [14] in terms of number of hidden layers and number of neurons per layer. The deep learning approach offers better results for our test scenarios. The final aim is to integrate the resulting validated models in a building energy decision support system, in a similar fashion to [15].

3. Electrical Load Forecasting using Sequence Model Recurrent Neural Networks

Recurrent neural networks (RNN) are becoming a very important modelling tool in applications where the input data consists of time-related sequences of "events". The inherent loop presents in an RNN carry long-term dependencies toward the final outcome. By simultaneously processing multiple elements in a sequence of inputs they are directly applicable for the evaluation of non-linear time series problems, such as electrical energy load forecasting. Given the differences to conventional artificial neural networks (ANN), RNNs use back-propagation through time (BPTT) [16] or real-time recurrent learning (RTRL) [17] algorithms to compute the gradient descent after each iteration for solving the optimisation problem. As summary, RNNs are a type of artificial neural network that include additional weights to the network to create cycles into the network graph in order to control the internal state of the network.

3.1. RNN Implementation with LSTM

Recurrent neural network can be implemented using typical with sigmoid, tanh and rectifier linear unit activation functions or by more advanced units which help better manage the information flow throughout the network. Of the latter category, most popular currently with researchers are Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) units. The main purpose of such techniques is to mitigate the effect of vanishing or exploding gradients during the training with long sequences of input data which creates significant numerical computation problems during the optimisation steps. The LSTM [18] structure contributes to enhance gradient flow over long sequences during training by means of self-connected "gates" in the hidden units. In a LSTM network the flow of information through the network is handled by reading, writing and removing information from the memory [19,20].

Figure 1 illustrates the flow of a time series x of length n, ($n \in \mathbb{N}$) through an LSTM layer. In this diagram, h stands for output, also hidden state, and c stands for cell state. The first LSTM hidden unit takes the initial state of the network, (c_0, h_0) and also the first time step of the sequence x_1 and after that computes the first output h_1 and the updated cell state c_1. This process repeats every time step.

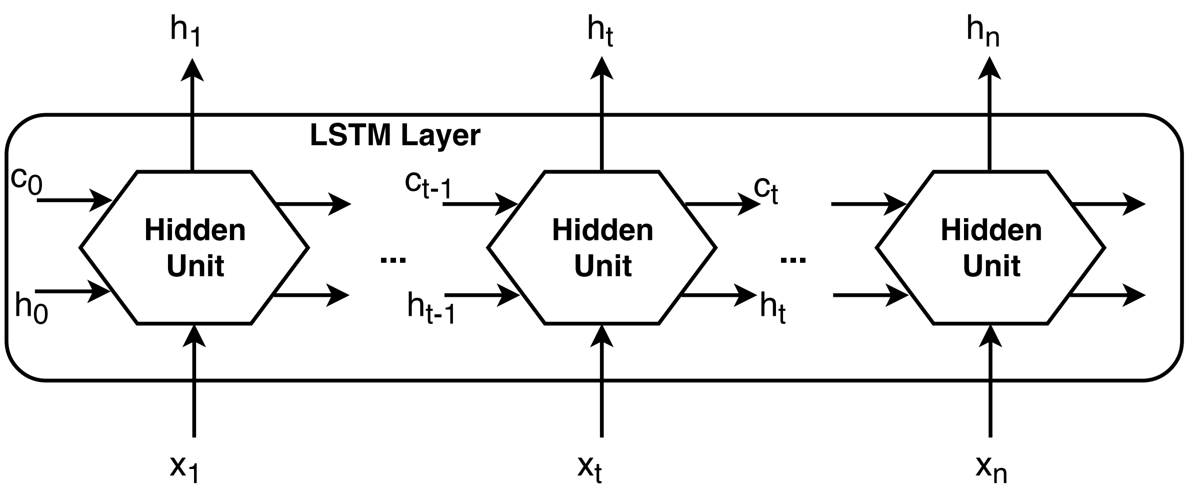


Figure 1. LSTM layer diagram

The state of the LSTM layer consists of the output state (h) and the cell state (c). On the one hand, the output state at time step t contains the output of the LSTM layer for this time step, and on the other hand the cell state contains information learned from the previous time steps. At each iteration, the layer writes information to the cell state or erases information from it, where the layer controls these updates using "gates". The "gates" represent components that control the cell state and the output state of the layer. There are four such components: the input gate (i) which controls the level of cell state update, the layer input (g) which adds information to cell state, the forget gate (f) which controls the level of cell state reset and the output gate (o) which controls the level of cell state added to output state [20].

The diagram in Figure 2 illustrates the data flow at a specific time step t inside of a hidden unit.

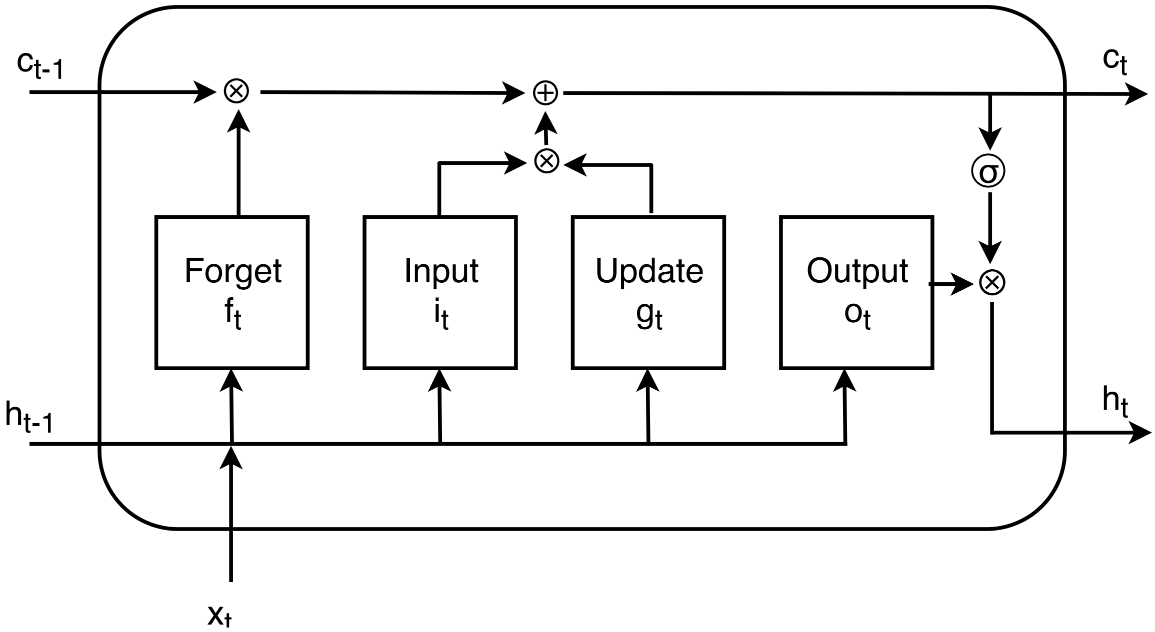


Figure 2. LSTM hidden unit diagram

A LSTM layer has the following learnable parameters: the input weights (W), the recurrent weights (R) and the bias (b). W, R and b represents matrices that consists of concatenations of the input weights, recurrent weights, and the bias of each component. The structure of each matrix is the following:

$$W = \begin{pmatrix} W_i \\ W_f \\ W_g \\ W_o \end{pmatrix}, R = \begin{pmatrix} R_i \\ R_f \\ R_g \\ R_o \end{pmatrix}, b = \begin{pmatrix} b_i \\ b_f \\ b_g \\ b_o \end{pmatrix},$$

where i, f, g and o represent the input gate, forget gate, layer input and output gate, respectively.

The state of the cell memory at time step t is updated recursively using the following formula [18]:

$$c_t = f_t \otimes c_{t-1} \oplus i_t \otimes g_t, \tag{1}$$

where $\otimes$ stands for Hadamard product and the formulas for every component at time step t are:

$$\begin{aligned} f_t &= \sigma(W_f x_t + R_f h_{t-1} + b_f), \\ i_t &= \sigma(W_i x_t + R_i h_{t-1} + b_i), \\ g_t &= \tanh(W_g x_t + R_g h_{t-1} + b_g), \end{aligned} \tag{2}$$

σ stands for the sigmoid function and tanh for hyperbolic tangent function.

The output state at time step t is given by the output gate (o) which implements a read function combined with the cell state (c). The process is described by the following formula:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2019 doi:10.20944/preprints201904.0318.v1

$$h_t = o_t \otimes tanh(c_t), \tag{3}$$

where

$$o_t = \sigma(W_o x_t + R_o h_{t-1} + b_o). \tag{4}$$

A simplified version of LSTM has been introduced by [21] named Gated Recurrent Unit (GRU). In this case the RNN cell uses a sole update gate and merges the cell and output states into a single value that is propagated across the network. A recent application of GRU vs. LSTM in electrical energy load forecasting is provided by [22].

3.2. Benchmarking data sets

We leverage and acknowledge the data sets within the Building Data Genome at the the Building and Urban Data Science (BUDS) Group at the National University of Singapore - `http://www.budslab.org`. This includes active power load traces and are part of a data collection of several hundreds of non-residential buildings, mainly academic venue, proposed for performance analysis and algorithm benchmarking to a common baseline [14,23].

The current study investigates a RNN LSTM modelling approach through using various neural network configurations and the assessment of performance between all forecasting LSTM models.

Key joint characteristics of the benchmarking data sets is the sampling period of the measurement of 60 minutes over a 1-year period. We select four educational buildings with an approximate surface area of 9.000 square meters. The chosen buildings for the study are from university campuses in Chicago (USA), two reference buildings, New York (USA) and Zurich (Europe). After the pre-processing of the noise and missing data in the initial data set using the median filter technique two time series data sets were obtained with approximately 8.670 data points each. The median filter [24] is expressed as:

$$y(n) = med[x(n-k), x(n-k+1), ..., x(n), ..., x(n+k-1), x(n+k)] \tag{5}$$

where $y(n)$ is the filtered output of the $x(n)$ input sequence. The sole parameter used for tuning the filter is represented by the filter length $n = 2k + 1$. In our case, a 15th order median filter is used by the preprocessing implementation.

The key criteria for the selection were: medium to large building, mixed usage pattern - office, laboratory space, some classrooms and non-extreme temperate climate with four distinct seasons [14]. A guiding choice in selecting the four candidates from the 508 entities in the dataset has been also similarity to a local university building from our campus where a data collection study is currently ongoing.

4. Experimental Evaluation for Building Energy Time-Series Forecasting

We present first the preprocessed time series for the buildings that make up our study and consist of hourly active power measurement from the electrical meters. Figure 3 presents the input data for the buildings in Zurich and Chicago, while Figure 4 presents the input data for the buildings in New York and the second Chicago building. All are from academic campuses and in terms of absolute electrical energy load, New York uses the most energy, followed by Zurich and Chicago in a similar range, and with Chicago 2 having the least energy needs.

The classical LSTM algorithm was implemented for experimental assessment and forecasting. The base network architecture consists of one sequence input layer, one hidden LSTM layer of varying unit numbers, one fully connected layer and one regression output layer for the resulting forecasted output value. Each network has a different configuration represented by the number of the hidden units from the LSTM layer. Based on this, the following network structures were implemented, in total 25 networks were trained, validated and evaluated: C-0, C-1, C-2, C-3, C-4, Z-0, Z-1, Z-2, Z-3,

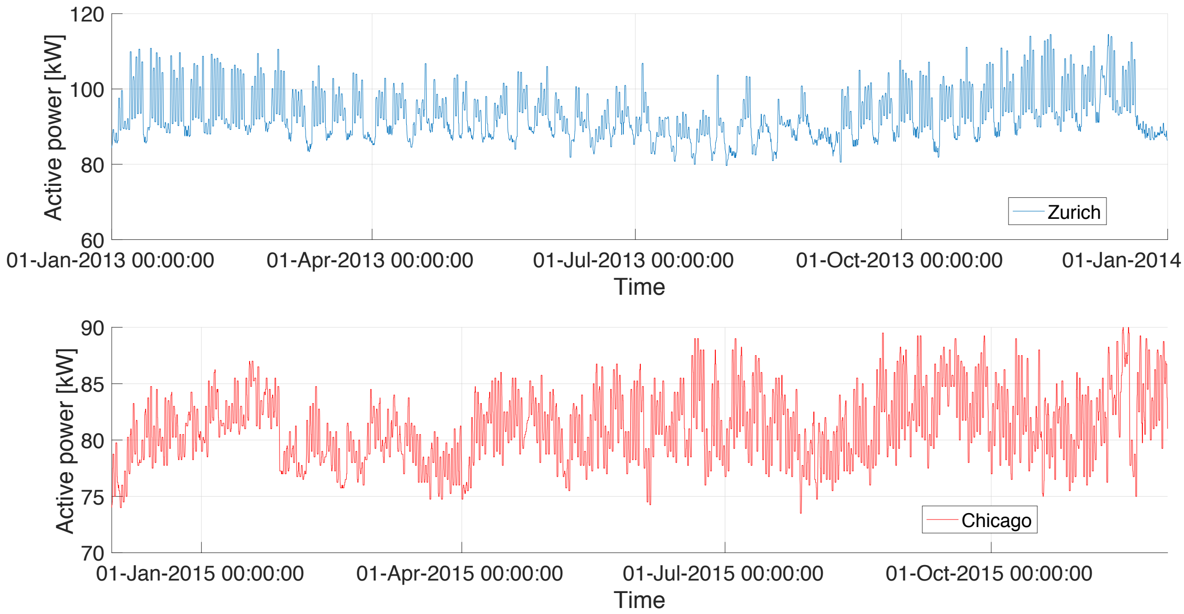


Figure 3. Data sets used for LSTM estimation and testing: Zurich and Chicago buildings

Z-4, C2-0, C2-1, C2-2, C2-3, C2-4, NY-0, NY-1, NY-2, NY-3, NY-4. In our case the identifier before the dash sign reflects the analysed building: *C* stands for the Chicago building, *Z* for the Zurich one, *C*2 for the second building from Chicago and *NY* for the New York building. The number after the building identifier marks the complexity of the network in terms of hidden units of LSTM that were implemented in the hidden layer, using a linear increase. This ranges from 5 hidden units for id 0, 25 hidden units for id 1, 50 hidden units for id 2, 100 hidden units for id 3 and finally 125 hidden units for id 4.

The optimisation method of choice for training the neural networks was through the Adaptive Moment Estimation (ADAM) algorithm [25]. This is an often used general optimisation method for first order gradient based optimisation of stochastic objective functions with momentum. The algorithm computes learning rates that can adapt in a automatic manner to the loss function which is optimised, for each parameter from estimates of first and second moments of the gradients. It maintains an element-wise moving average of the parameter gradients and their squared values, respectively . In the literature is demonstrated that the algorithm is efficient in terms of computational time, has little memory requirements and is suitable for large data problems.

One of the key optimisation parameters for carrying out the neural network training is the learning rate. This allows implementing a trade-off between the speed of the processing and its precision, in the sense that a large learning rate can in many situation miss the optimal value of the objective metric. In our case the learning rate was established through an empirical adjustment process. The initial value was set at 0.1 followed by subsequent decreases with a factor of 0.2 every 200 iterations. From observing the performance over multiple initial training runs, a second parameters, the number of training iterations, was set at 200.

Figure 5, 6, 7 and 8 presents the prediction response by the LSTM neural network with 50 hidden units in the LSTM layer versus real data for Chicago building and Zurich building, respectively. The plots demonstrate that the forecasting performances of the LSTM models for the testing data sets is very good.

To evaluate the prediction models, three performance metrics were used: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE). In addition we

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2019 doi:10.20944/preprints201904.0318.v1

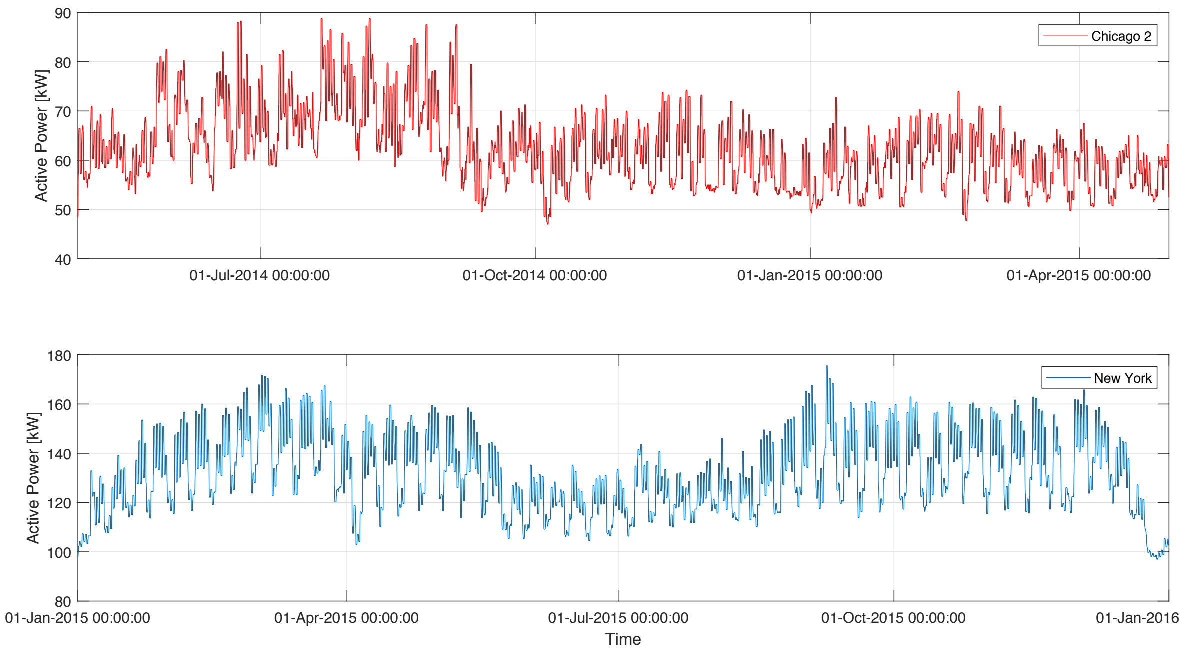


Figure 4. Data sets used for LSTM estimation and testing: Chicago 2 and New York buildings

have included the Coefficient of Variation of the Root Mean Square Error CV(RMSE) based on the evaluation discussed in [11]. The metrics are computed according to the following equations:

$$\begin{aligned} MSE &= \sum_{1}^{n} \frac{(Y_t - Yp_t)^2}{n} \\ RMSE &= \sqrt{\sum_{1}^{n} \frac{(Y_t - Yp_t)^2}{n}} \\ MAPE &= \frac{1}{n} \sum_{1}^{n} \left| \frac{Y_t - Yp_t}{Y_t} \right| 100 \\ CV(RMSE) &= \frac{\sqrt{\sum_{1}^{n} \frac{(Y_t - Yp_t)^2}{n}}}{\frac{\sum_{1}^{n} Yp_t}{n}} 100 \end{aligned} \tag{6}$$

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

A summary of the experimental results is listed in Tables 1-4 show These include: the error metrics MSE, RMSE, CV (RMSE) and MAPE as well as training/computation time for the previously defined RNN LSTM networks.

Table 1. Summary of accuracy metrics for RNN LSTM model forecasting: Chicago

	C-0	C-1	C-2	C-3	C-4
Time(s)	75	93	143	247	330
MSE	0.6295	0.6132	**0.5553**	0.7486	0.9555
RMSE	0.7934	0.7831	**0.7452**	0.8652	0.9775
CV(RMSE)(%)	0.98	0.97	**0.92**	1.07	1.2
MAPE(%)	0.5623	0.5091	**0.4945**	0.5535	0.8177

The main outcome of the learning models, as reflected by the aggregate performance metrics from Table 1, 2, 3 and 4 pinpoint the best network architecture for all four testing scenarios to be the one with 50 LSTM units in the hidden layer. Also, Figure 9 presents the evolution of the MAPE error and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2019 doi:10.20944/preprints201904.0318.v1

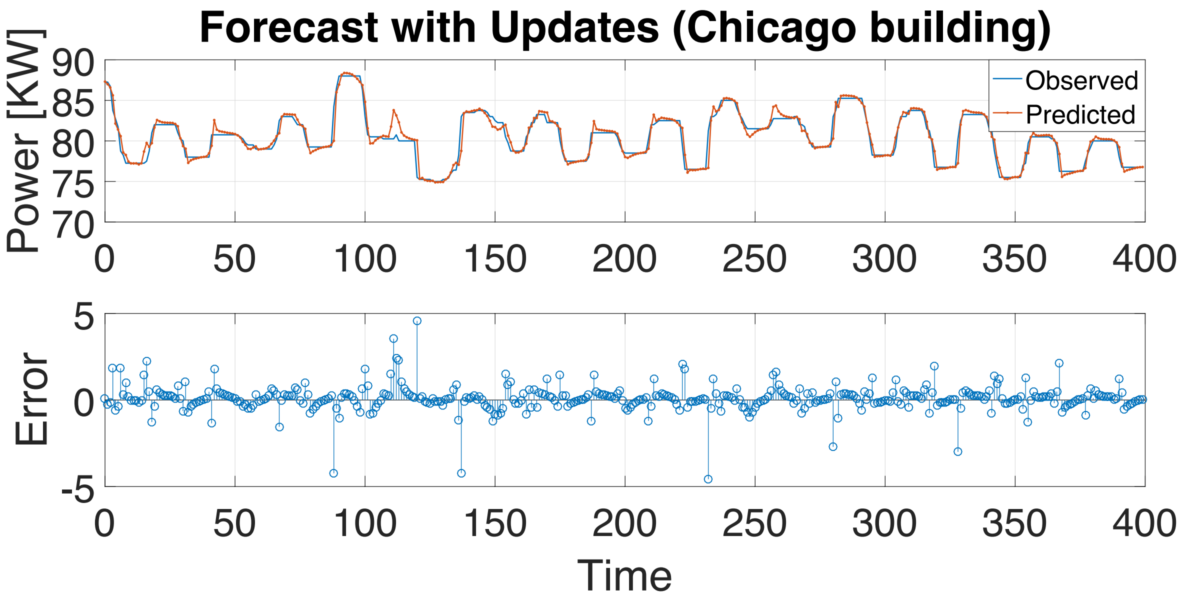


Figure 5. Building Load Forecasting with 50 unit RNN LSTM: Chicago

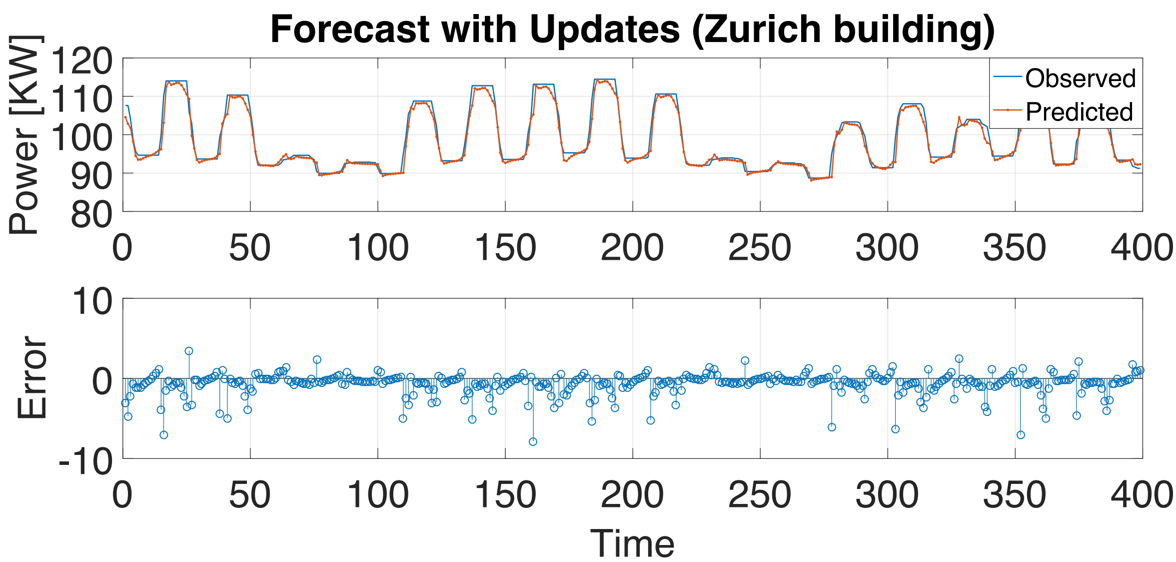


Figure 6. Building Load Forecasting with 50 unit RNN LSTM: Zurich

Table 2. Summary of accuracy metrics for RNN LSTM model forecasting: Zurich

	Z-0	Z-1	Z-2	Z-3	Z-4
Time(s)	76	94	150	275	355
MSE	2.3846	2.1732	**2.0506**	2.2506	2.034
RMSE	1.5442	1.4742	**1.432**	1.5002	1.5107
CV(RMSE)(%)	1.66	1.59	**1.58**	1.61	1.63
MAPE(%)	0.9197	0.9008	**0.828**	0.9958	3.4684

computation time for each building over each defined network. It can be seen that the computation time increases quite linearly with the number of the neurons in the LSTM layers which can be helpful for deploying more tests. We can affirm that until we achieve the best value for MAPE error, 50 neurons in the LSTM layer, this is a good compromise, but after this point the computation time increases too much without any better performances. The reference computer includes a 2.6 GHz 7th generation Intel i5 CPU, 8GB of RAM and a solid state disk, with Windows 10 as operating system. This is the baseline for the reported computation/training time for all test cases. Algorithms have been written

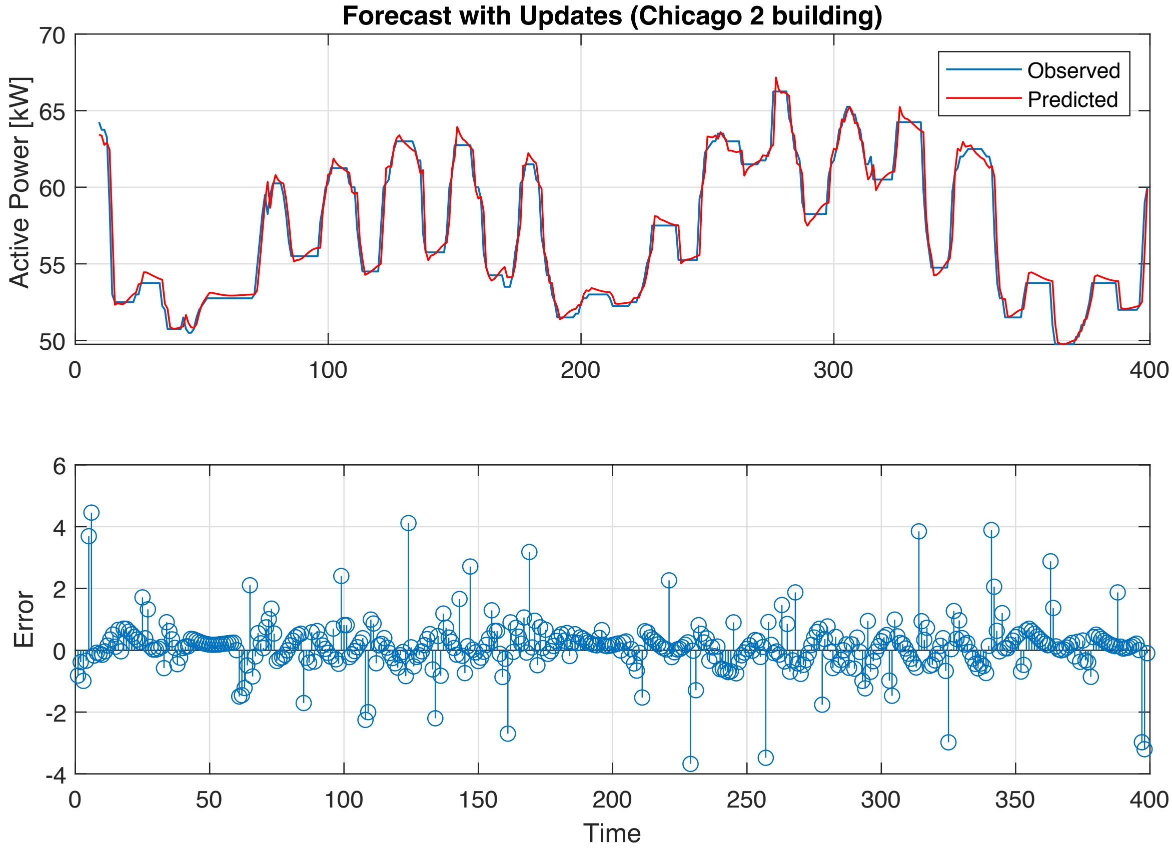


Figure 7. Building Load Forecasting with 50 unit RNN LSTM: Chicago 2

Table 3. Summary of accuracy metrics for RNN LSTM model forecasting: Chicago 2

	C2-0	C2-1	C2-2	C2-3	C2-4
Time(s)	72	91	137	280	347
MSE	0.8013	0.7801	**0.7163**	0.7626	0.8413
RMSE	0.8952	0.8832	**0.8464**	0.8733	0.9172
CV(RMSE)(%)	1.63	1.6	**1.53**	1.57	1.65
MAPE(%)	1.0005	0.9439	**0.8982**	0.9736	1.019

Table 4. Summary of accuracy metrics for RNN LSTM model forecasting: New York

	NY-0	NY-1	NY-2	NY-3	NY-4
Time(s)	79	102	145	294	346
MSE	5.4433	5.4012	**4.7778**	5.1203	5.5522
RMSE	2.3309	2.3241	**2.1858**	2.2628	2.3563
CV(RMSE)(%)	1.75	1.73	**1.7**	1.72	1.78
MAPE(%)	1.0409	1.004	**0.9602**	0.9813	1.1073

and run under MATLAB, version R2018a, which provides a robust high-level technical programming environment. We leverage built-in functions from the machine and deep learning toolboxes as well as dedicated scripts for data ingestion and preprocessing.

Table 5 provides a summary of the statistical indicators for the comparable relative accuracy metrics: CV-RMSE and MAPE over the tested scenarios, four building with five networks each. The reported statistical indicators are: minimum, maximum, mean μ, standard deviation σ, skewness and kurtosis.

The performance evolution during training for the Zurich and New York buildings is graphically depicted in Figures 10 and 11. The graphic represents the gradual decrease in the RMSE metric over

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2019 doi:10.20944/preprints201904.0318.v1

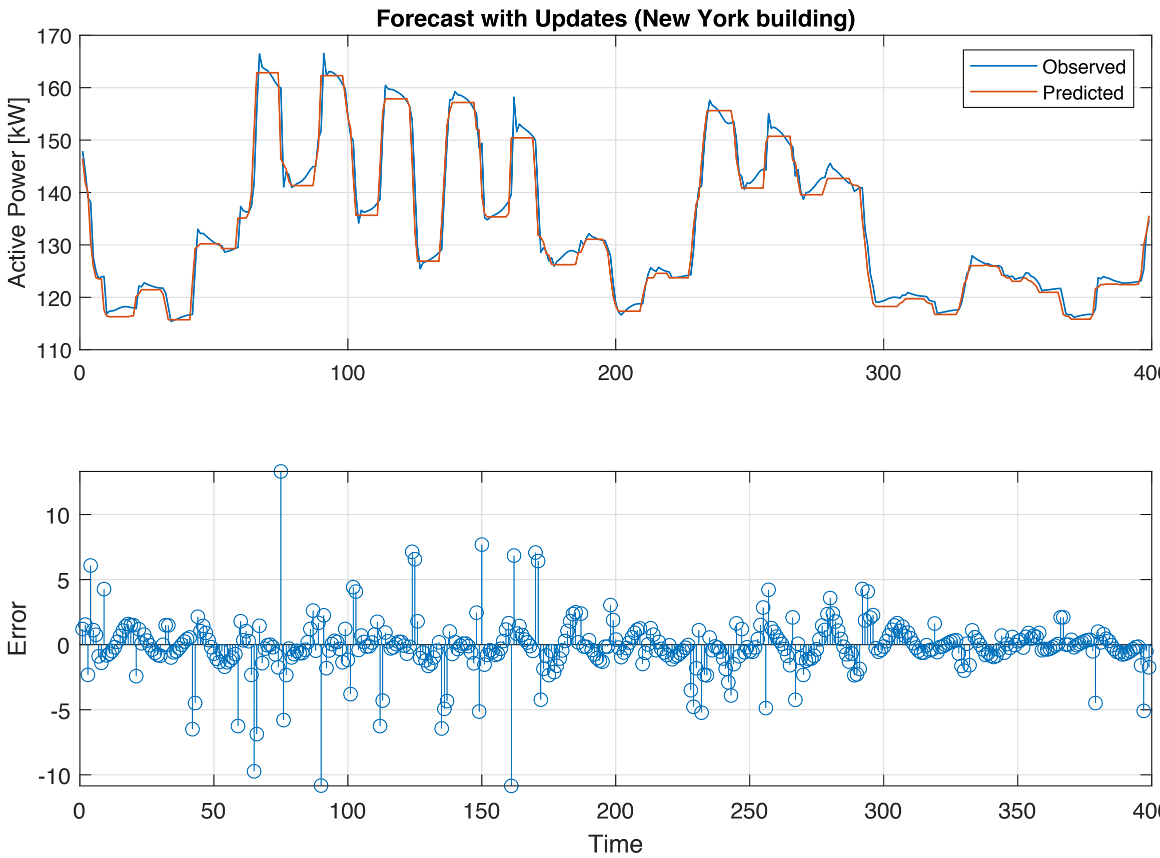


Figure 8. Building Load Forecasting with 50 unit RNN LSTM: New York

Table 5. Statistical indicators for relative peformance metrics CV-RMSE and MAPE

	Min	Max	μ	σ	s	k
CV (RMSE)	0.92	1.78	1.4935	0.2873	−1.0846	2.5502
MAPE	0.4945	3.4684	**0.9989**	0.6108	3.4661	14.8790

200 iterations, with the worst and best case scenarios. In the first case the worst performance is seen on the Z4 network which tries to overfit the data given the more complex structure and as such the RMSE presents multiple increases and decreases over the training horizon. In the positive case, Z2, we observe convergence in just under 100 training iterations as compared to the previous 120 iterations needed by the denser network. Similar networks are represented for the New York dataset. A different behaviour is observed in this case with where the best-case convergence of the RMSE is slower at the beginning with a more gentle slope of the graphic over the first steps.

5. Conclusions

We have presented an application of sequence models, implemented by means of Recurrent Neural Network with Long Short-Term Memory units, for electrical energy load forecasting in large commercial buildings. Results have shown good performance for modelling and forecasting accuracy while evaluating against typical time series based metrics such as: MSE, RMSE, CV-RMSE and MAPE. Generally, based on the number of layers that were selected, a value around 50 LSTM units in the hidden layer was found to yield the best estimates for the underlying time series. Beyond this the network has a strong tendency to overfit the input data and perform poorly on the testing samples. The results have been evaluated on year-long selected building energy traces from a reference benchmarking dataset and the MAPE relative metric was found to be between 0.5% and 1% for all

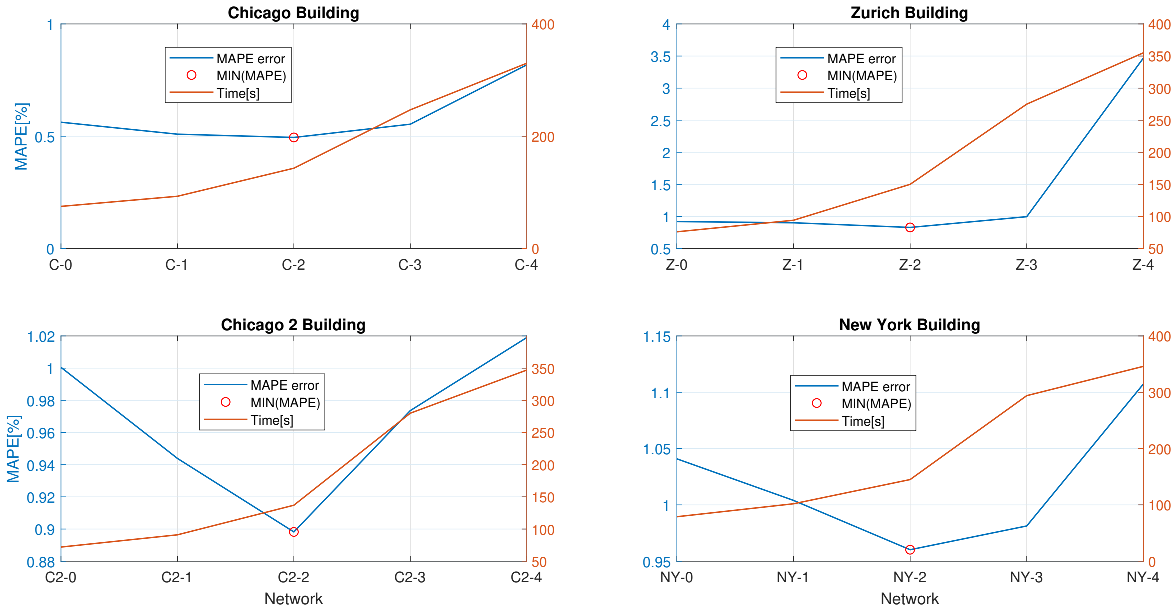
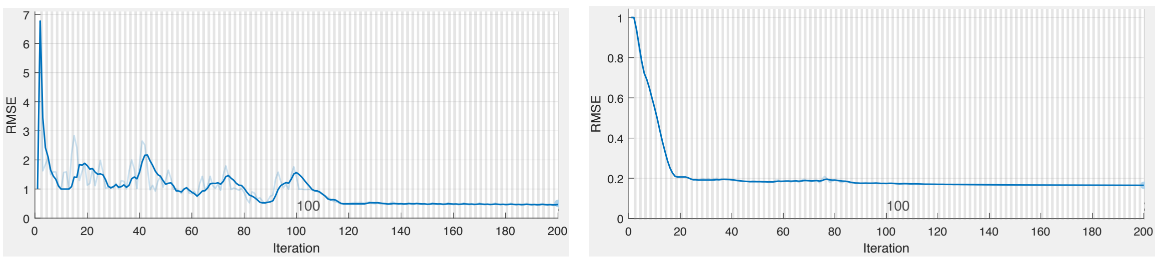


Figure 9. Computation Time vs. MAPE Error evolution for each building



(a) (b)

Figure 10. RMSE Accuracy During Training: Zurich a) Z-4 b) Z-2

the investigated buildings. The coefficient of variation of the RMSE (CV-RMSE) is stable between the various scenarios.

For future work we aim at validating the approaches presented in this article on the full BUDS dataset of 508 buildings using an appropriate which goes beyond the resources available on a single local machine or server. For this the use of a cloud based of GPU computing infrastructure is planned in order to reach feasible computing time. Similar to the approach [26] a hardware-in-the-loop architecture is envisioned to deploy the resulting black-box models on an embedded device for inference while streaming data from a real building. In such scenarios, the models would be pre-trained with only partial retraining and model update locally based on continuously streamed energy consumption values from local smart meters.

Author Contributions: Formal analysis, Iulia Stamatescu; Investigation, Cristina Nichiforov; Methodology, Cristina Nichiforov; Software, Grigore Stamatescu; Supervision, Ioana Fagarasan; Validation, Ioana Fagarasan; Writing – original draft, Grigore Stamatescu; Writing – review & editing, Iulia Stamatescu.

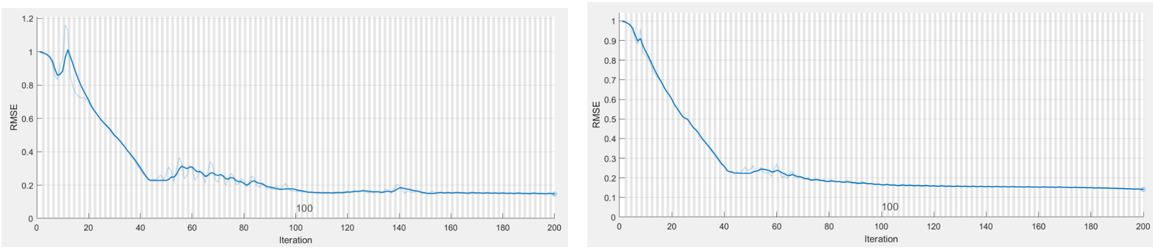
Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

 doi:10.20944/preprints201904.0318.v1



(a) (b)

Figure 11. RMSE Accuracy during Training: New York a) NY-4 b) NY-2

ADAM	Adaptive Moment Estimation
ARIMA	Auto-Regressive Integrated Moving Average
BPTT	Backpropagation Through Time
CV - RMSE	Coefficient of Variation of RMSE
LSTM	Long Short-Term Memory
DL	Deep Learning
GRU	Gated Recurrent Unit
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RTRL	Real-Time Recurrent Learning

1. Berardi, U. Building Energy Consumption in US, EU, and BRIC Countries. *Procedia Engineering* **2015**, *118*, 128 – 136. Defining the future of sustainability and resilience in design, engineering and construction, doi:https://doi.org/10.1016/j.proeng.2015.08.411.
2. Djenouri, D.; Laidi, R.; Djenouri, Y.; Balasingham, I. Machine Learning for Smart Building Applications: Review and Taxonomy. *ACM Comput. Surv.* **2019**, *52*, 24:1–24:36. doi:10.1145/3311950.
3. Nichiforov, C.; Stamatescu, G.; Stamatescu, I.; Calofir, V.; Fagarasan, I.; Iliescu, S.S. Deep Learning Techniques for Load Forecasting in Large Commercial Buildings. 2018 22nd International Conference on System Theory, Control and Computing (ICSTCC), 2018, pp. 492–497. doi:10.1109/ICSTCC.2018.8540768.
4. Fayaz, M.; Shah, H.; Aseere, A.M.; Mashwani, W.K.; Shah, A.S. A Framework for Prediction of Household Energy Consumption Using Feed Forward Back Propagation Neural Network. *Technologies* **2019**, *7*, 30.
5. Wang, X.; Zhang, M.; Ren, F. Learning Customer Behaviors for Effective Load Forecasting. *IEEE Transactions on Knowledge and Data Engineering* **2019**, *31*, 938–951. doi:10.1109/TKDE.2018.2850798.
6. Mocanu, E.; Nguyen, P.H.; Gibescu, M.; Kling, W.L. Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks* **2016**, *6*, 91 – 99. doi:https://doi.org/10.1016/j.segan.2016.02.005.
7. Aminikhanghahi, S.; Wang, T.; Cook, D.J. Real-Time Change Point Detection with Application to Smart Home Time Series Data. *IEEE Transactions on Knowledge and Data Engineering* **2019**, *31*, 1010–1023. doi:10.1109/TKDE.2018.2850347.
8. Jin, M.; Bekiaris-Liberis, N.; Weekly, K.; Spanos, C.J.; Bayen, A.M. Occupancy Detection via Environmental Sensing. *IEEE Transactions on Automation Science and Engineering* **2018**, *15*, 443–455. doi:10.1109/TASE.2016.2619720.
9. Li, C.; Ding, Z.; Zhao, D.; Yi, J.; Zhang, G. Building Energy Consumption Prediction: An Extreme Deep Learning Approach. *Energies* **2017**, *10*. doi:10.3390/en10101525.
10. Rahman, A.; Srikumar, V.; Smith, A.D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Applied Energy* **2018**, *212*, 372 – 385. doi:https://doi.org/10.1016/j.apenergy.2017.12.051.

11. Fan, C.; Sun, Y.; Zhao, Y.; Song, M.; Wang, J. Deep learning-based feature engineering methods for improved building energy prediction. *Applied Energy* **2019**, *240*, 35 – 45. doi:https://doi.org/10.1016/j.apenergy.2019.02.052.
12. Zhong, H.; Wang, J.; Jia, H.; Mu, Y.; Lv, S. Vector field-based support vector regression for building energy consumption prediction. *Applied Energy* **2019**, *242*, 403 – 414. doi:https://doi.org/10.1016/j.apenergy.2019.03.078.
13. Nichiforov, C.; Stamatescu, I.; Făgărăşan, I.; Stamatescu, G. Energy consumption forecasting using ARIMA and neural network models. 2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE), 2017, pp. 1–4. doi:10.1109/ISEEE.2017.8170657.
14. Nichiforov, C.; Stamatescu, G.; Stamatescu, I.; Fagarasan, I.; Iliescu, S.S. Intelligent Load Forecasting for Building Energy Management. 2018 14th IEEE International Conference on Control and Automation (ICCA), 2018.
15. Stamatescu, I.; Arghira, N.; Fagarasan, I.; Stamatescu, G.; Iliescu, S.S.; Calofir, V. Decision Support System for a Low Voltage Renewable Energy System. *Energies* **2017**, *10*. doi:10.3390/en10010118.
16. Werbos, P.J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE* **1990**, *78*, 1550–1560. doi:10.1109/5.58337.
17. Williams, R.J.; Zipser, D. Backpropagation; L. Erlbaum Associates Inc.: Hillsdale, NJ, USA, 1995; chapter Gradient-based Learning Algorithms for Recurrent Networks and Their Computational Complexity, pp. 433–486.
18. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. doi:10.1162/neco.1997.9.8.1735.
19. Marino, D.L.; Amarasinghe, K.; Manic, M. Building Energy Load Forecasting using Deep Neural Networks. *CoRR* **2016**, *abs/1610.09460*, [1610.09460].
20. Srivastava, S.; Lessmann, S. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Solar Energy* **2018**, *162*, 232 – 247.
21. Chung, J.; Gülçehre, Ç.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv e-prints* **2014**, *abs/1412.3555*. Presented at the Deep Learning workshop at NIPS2014.
22. Ugurlu, U.; Oksuz, I.; Tas, O. Electricity Price Forecasting Using Recurrent Neural Networks. *Energies* **2018**, *11*. doi:10.3390/en11051255.
23. Miller, C.; Meggers, F. The Building Data Genome Project: An open, public data set from non-residential building electrical meters. *Energy Procedia* **2017**, *122*, 439 – 444. doi:https://doi.org/10.1016/j.egypro.2017.07.400.
24. Broesch, J.D. Chapter 7 - Applications of DSP. In *Digital Signal Processing*; Broesch, J.D., Ed.; Instant Access, Newnes: Burlington, 2009; pp. 125 – 134. doi:https://doi.org/10.1016/B978-0-7506-8976-2.00007-9.
25. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *CoRR* **2014**, *abs/1412.6980*, [1412.6980].
26. Stamatescu, G.; Stamatescu, I.; Arghira, N.; Făgărăsan, I.; Iliescu, S.S. Embedded networked monitoring and control for renewable energy storage systems. 2014 International Conference on Development and Application Systems (DAS), 2014, pp. 1–6. doi:10.1109/DAAS.2014.6842417.