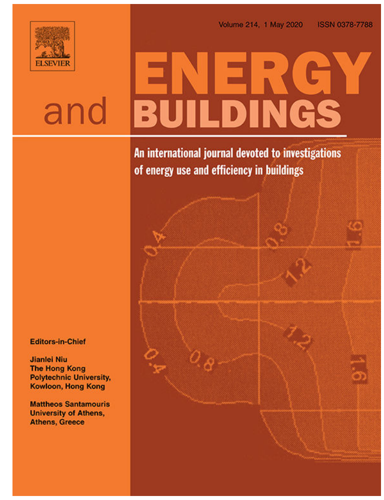
Journal Pre-proofs



Day-Ahead Prediction of Plug-In Loads Using a Long Short-Term Memory Neural Network

Romana Markovic, Elie Azar, Masab Khalid Annaqeeb, Jérôme Frisch, Christoph Treeck van

PII: S0378-7788(20)33453-8
DOI: https://doi.org/10.1016/j.enbuild.2020.110667
Reference: ENB 110667

To appear in: *Energy & Buildings*

Received Date: 3 February 2020
Accepted Date: 13 December 2020

Please cite this article as: R. Markovic, E. Azar, M.K. Annaqeeb, J. Frisch, C. Treeck van, Day-Ahead Prediction of Plug-In Loads Using a Long Short-Term Memory Neural Network, *Energy & Buildings* (2020), doi: https://doi.org/10.1016/j.enbuild.2020.110667

Day-Ahead Prediction of Plug-In Loads Using a Long Short-Term Memory Neural Network

Romana Markovic*[a], Elie Azar[b], Masab Khalid Annaqeeb[c], Jérôme Frisch[a], Christoph van Treeck[a]

[a]*E3D - Institute of Energy Efficiency and Sustainable Building, RWTH Aachen University, Mathieustr. 30, 52074 Aachen, Germany*
[b]*Department of Industrial and Systems Engineering, Khalifa University, Masdar Institute, PO Box 54224, Abu Dhabi, United Arab Emirates*
[c]*Department of Energy and Process Engineering, Norwegian University of Science and Technology, Kolbjørn Hejes v 1B, NO-7491 Trondheim, Norway*

Abstract

The aim of this work is to develop and validate a miscellaneous electric loads (MEL) predictive model that does not require occupant-wise or building-wise model training nor model adaptation while achieving competitive accuracy. For that purpose, a long-short-term memory (LSTM) model was developed using monitored data from a research building located in Abu Dhabi, United Arab Emirates (UAE). In order to test the generalization capabilities of the proposed method, the model was evaluated using data from two additional buildings, a bank office building located in Frankfurt, Germany, and a university building in Ottawa, Canada. The results showed that the developed LSTM is applicable to the tested buildings without the need for occupant-wise or building-wise calibration, hence, addressing an important gap in the existing literature. In addition, a set of MEL predictive models from the literature, that are based on a Weibull distribution and Gaussian mixture models (GMM) are implemented and evaluated using the three identical data sets. The round-robin evaluation of existing MEL predictive models showed that the proposed LSTM model outperformed them especially when a combination of MEL and occupancy information was used as inputs. Finally, the neural network saturation was identified as the key challenge when developing an LSTM-based model for MEL prediction.

Keywords: occupant behavior, plug-in loads, LSTMs, neural networks, MEL prediction

*Corresponding author. Tel.: +49-241-80 25541;
E-Mail: markovic@e3d.rwth-aachen.de

1. Introduction

1.1. Background

Buildings account for more than 36 % of the worldwide energy consumption [1], while the buildings' share of electric consumption reaches up to 70-80 % in countries with extreme climate conditions such as UAE [2, 3]. In addition, office equipment accounts for around 30 % of the overall energy consumption in commercial buildings in United States [4], while the similar trend was observed by the recent research study in the UAE [3]. One of the significant aspects of the office equipment are miscellaneous electric loads (MELs) that represent the energy used by appliances and devices outside of a building's core functions of heating, ventilation, air conditioning (HVAC), lighting, water heating, and refrigeration [4]. As a consequence, reliable estimates of the MEL is important for adequate design decision making in the area of energy efficient buildings [5]. These estimates should be incorporated in decision making for buildings in different climate zones. In hot or cold climates, the MEL is relevant primarily due to its contribution to the internal heat gain and resulting impacts on cooling or heating loads. In mild climates, the MELs account for the larger absolute proportion of the energy consumption, and therefore, they represent a major energy saving potential. In both cases, however, the energy consumption by equipment was identified to have the second highest potential in energy savings, with relatively consistent values across different building types [6].

As Yan et al. [7] already pointed out, the equipment in homes and offices represent substantial uncertainty because both the type of appliances and the use patterns are usually not controlled by building designers or operator. Resultantly, static profiles as proposed by guidelines [8, 9] lead to the unreliable estimation of the MEL and resulting internal heat gains, which was in the focus of research by multiple studies [10, 11]. For instance, Lin and Azar [10] analyzed the occurrence of energy saving actions, such as turning down laptops when they are not in use or lighting energy savings at the workplace and in residential buildings. Masoso et al. [11] analyzed the power consumption during occupied and non-occupied hours in the hot climate in Botswana.

Motivated by the MEL uncertainty and saving potential, a number of studies quantified the amplitude of the MEL energy consumption [12–21] and researched the suitable approaches for predicting MELs [5, 22–25, 27, 28], as well as the impact of the occupant behavior (OB) on resulting energy consumption [6, 10, 11, 29–35].

Mahdavi et al. [5] proposed a plug-in load model using Weibull distributions that were fitted for discrete occupancy or absence duration. Gunay et al. [23] proposed a day-ahead MEL predictive modeling using GMMs. Therefore, GMMs with two principal components were fitted using historical data. O'Brien et al. [25] proposed generating synthetic plug-in loads and lighting profiles using multivariate normal random approach. Wang et al. [27] proposed modeling the MEL, lighting, and occupancy using LSTM networks, while Lasternas et al. [28] and Zhao et al. [33] proposed different machine learning classification algorithms for detecting the occupancy status from the plug-in consumption.

In summary, the existing studies identified following challenges regarding the modeling of relationships between occupants' presence and MEL profiles:

- the inter-tenant diversity need to be addressed [5, 11, 24, 25],
- significant part of MEL consumption occurred during occupant's absence [6, 11, 35],
- OB models that are developed for one building can not be used in other settings due to building-specific contextual factors and peculiarities [30],
- the occupancy was a widely used predictor in developed appliance energy consumption models [5, 23, 24, 27]. On the other side, the power consumption has also proven to be a significant variable for estimating the occupancy [27, 28, 33]. However, the causality of these two variables was rarely investigated.

The recent research on OB modeling identified neural networks as a suitable approach for modeling energy consumption-related human actions in buildings and addressing some of the above-listed gaps. These included the occupant-centric HVAC control [36], window openings [37, 38], lighting [26] and MELs [27]. In the scope of a recent study, Wang et al. [27] used LSTMs to predict internal gains at a building thermal-zone scale. The main focus of the paper was on the data fusion process and the relevant sensor information that could be beneficial for the MEL prediction.

In summary, the extensive existing research on MEL modeling identified the need for optimizing and addressing MELs in buildings, a set of unsolved research problems and the promising modeling approaches. However, there has been little work on applicability of the MELs models to different buildings, and additionally, on defining the boundaries between the model's applicability to different buildings and models' applicability for diverging modeling objective.

1.2. Contribution

This paper proposes a method for the day ahead prediction of the MEL in office and academic research buildings. Based on the MEL and occupancy patterns over the past days, an LSTM neural network is trained to predict MEL for the next 24 hours. For that purpose, the model was developed using the data from a single building located in Abu Dhabi, while the empirical evaluation confirmed model's applicability for significantly different buildings located in Frankfurt and Ottawa.

The main contribution of this work is the development of a MEL predictive model, that achieves competitive accuracy without any occupant-wise or building-wise model calibration. An additional contribution of this paper is the presented round-robin evaluation of the existing MELs models using relatively large data sample from geographically and culturally different locations. In order to achieve these goals, the following research questions are addressed in the scope of this study:

- could the inclusion of occupancy patterns over past days improve the predictive accuracy?
- could the proposed model be applied to different buildings without any model calibration or adaptation?
- could the use of suitable LSTM architecture improve the MEL prediction accuracy when compared to existing models?

The additional contribution of this work is a conducted round robin study of the established models for MEL modeling using three different data sets. Here, the models were implemented and evaluated using relatively large data samples. In total, the data sample used for validation was significantly larger than the data available to the researchers for testing these established models in the scope of original corresponding studies. In summary, this presents a comprehensive validation of the existing models and their performance on the independent and significantly different data sample.

The rest of this paper is organized as follows: a summary of the used abbreviations is available in Table 1. The research boundaries and contribution are presented in section 2. An overview of the used data sets is available in section 3. Model development and the experimental settings are presented in section 4. The resulting predictive performance is presented and analyzed in section 5. Eventually, the results are discussed and summarized in sections 6-7.

Table 1: List of abbreviations.

BAS	building automation system
BPTT	back propagation through time
HVAC	heating, ventilation and air conditioning
GPU	graphics processing unit
LSTM	long-short-term memory
MRE	mean relative error
MEL	miscellaneous electric loads
MSE	mean squared error
N-MSE	normalized mean squared error
N-RMSE	normalized mean squared error
OB	occupant behavior
OCC	occupancy
ReLU	rectified linear units
R-MSE	root mean squared error
RNN	recurrent neural network
UAE	United Arab Emirates

2. Research boundaries: generalization versus domain adaptation and transfer learning for energy consumption related occupant behavior

Model generalization, domain adaptation as well as transfer learning are relevant on the path towards the more real-world applicable models for predictive modeling of OB and in general of the energy consumption in buildings. These three fields are already established in context of the general machine learning, while they are still at the very early stage in case of the OB modeling in buildings. In order to set the clear boundaries of this paper, this section summarizes the difference between these three topics in context of the energy consumption in buildings.

Formally, *generalization* can be defined as the ability to perform well on previously unobserved inputs [39]. The generalization capabilities of the proposed models are commonly explored in the scope of the round robin OB studies. The majority of existing round-robin evaluated OB studies addresses window opening behavior. As earlier summarized by Markovic et al. [37], the model originally proposed by Haldi and Robinson [40] and by Rijal et al. [41] were evaluated by a number of following studies on alternative buildings [42–45]. To the best of authors' knowledge, the round robin studies on modeling alternative types of OB have been sparse and less comprehensive, when compared to the window opening studies.

Domain adaptation refers to the situation where the model learned on one domain (i.e., distribution P_1) is exploited to improve generalization in another setting (i.e. distribution P_2) [39]. The domain adaptation is commonly applied in order to reduce costs of acquiring labeled data [48] and it is beneficial in the cases where the factors that explain variations in P_1 may be relevant to the variations that need to be captured for learning [39] . The research on domain adaptation for the OB and in general for energy efficient buildings is still in the early stage and, the number of studies on domain adaptation for OB modeling is still limited [49], [50]. Arief-Ang et al. [49] applies semi-supervised domain adaptation to apply the model for occupancy count in significantly different building settings. Their results showed, that the use of domain adaptation led to the slight improvement in the occupancy count estimation. Zhang et al. [50] applied domain adaptation to model occupancy count for different rooms that are located within the same building. They relied on weight scaling for the target domain by conducting additional training iterations. The results pointed out a significant improvement in the occupancy count after the application of the domain adaptation.

Transfer learning can be defined as the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned [51]. Namely, the alternative research question formulation from the field of transfer learning would aim to perform two different tasks on distributions P_1 and P_2, where the learned context from P_1 is relevant for P_2 [39]. In context of energy consumption in buildings, an illustrative example for transfer learning could be training the model to predict the MEL profiles, and applying transfer learning to that model to predict HVAC energy consumption profiles.

A limitation of the current state of research is, that the existing models were commonly developed using data from a single building. Eventually, the generalization capabilities were quantified by evaluating the model using the data from alternative buildings. In the scope of a recent study on general energy consumption modeling in buildings, Miller [46] comprehensively elaborated on the limited generalization capabilities of the current approaches for energy modeling. The similar approach, namely model training with the data from several source domains (in this context – buildings) should be followed in case of the OB modeling, for example by the introduction of global and local training sets. The latter idea was already introducing by Peng et al. [47] for applying the same model to the different zones of the same building.

Even though the model training using data from multiple buildings is a promising future research direction, primarily in the context of meta-learning, it was not explored in the scope of this study. Namely, the clear distinction between the meta learning and learning the knowledge from multiple domains (domain=building) and the method proposed in the scope of this study is, that this study relies on the auto-regressive properties of the occupancy and MEL streams. These MEL and occupancy streams are used in high temporal resolution and the hypothesis is set, that these auto-regressive properties share some similarities over different domains. Therefore, it was aimed to explore the model applicability to different buildings, in case it was developed using the data from a single building.

Given the aim of this work to make a step forward towards the predictive models of MEL profiles in commercial buildings, this work explores the applicability of a day-ahead MEL predictive model that is trained using data from the occupants' subset from the building in Abu Dhabi, and eventually evaluated using different occupants from Abu Dhabi, as well as the data collected in Ottawa and Frankfurt. On this place, the scope of this work is on how well can a model predict the data from different target domains. Given that the target function, namely the day ahead energy consumption is assumed to be the same within the different buildings, this work is not aiming to propose the transfer learning approach.

In summary, the model's generalization refers mainly to quantifying how well the model performs in different settings, where distributions of the input and target variables are different. In this case, different settings are addressed by using the data from different buildings and locations. The domain adaptation would be of relevance in case we would like to minimize the discrepancy between the distributions of these two domains in order to improve model performance. Finally, the transfer learning would be of relevance in case the same model is to be used not only for different distributions, but also for different tasks.

In the scope of this study, the research boundaries are set to the model generalization to different domains. The tested hypothesis is, how well can the developed model generalize to different domains for modeling the identical target function, while no domain adaptation nor transfer learning was applied. Given a model hypothesis M, that is trained to tackle the task T on the domain $D1$, with M defined as a fully trained neural network. In addition, the task T is strictly defined as MEL energy consumption in office buildings. Data from different buildings are considered different domains such as

$$D_1 ... D_n, \tag{1}$$

with n ∈ (1, N) where N is a number of different buildings. These domains have diverging distributions, that are unknown prior to the model training. The research boundaries towards the transfer learning are set by considering that the MEL consumption is the same task in different office buildings so that the resulting relationship between D_1 and T is identical to the relationship between the D_n and T. In addition, no domain adaptation is applied and the model is evaluated for the ability to predict the energy consumption by being trained only on a single target domain.

3. Data sets and data preprocessing

In order to test the model performance on alternative buildings, the screening for the open source MEL data sets was conducted. Since, the objective of this work was to model MEL in office buildings by using MEL and occupancy from the previous days, the data sets that fulfill following criteria were considered:

- the energy consumption data is related to the MEL loads,
- the measured occupancy is available in terms of presence and absence,
- the data are collected in arbitrary office buildings.

In spite of the large number of open source energy consumption data sets (see for example [23, 52–59] as well as publicly available repositories such as ”Mendeley data sets for Energy and buildings”[1]), the number of data sets that target specifically MEL loads and occupancy in office buildings is very limited. For instance, Doherthy and Trenbath [53] released the data collected on the device level in an office building over three months. However, the individual plug-ins were not assigned to each workstation and the occupancy information was not available and therefore, this data set could not be used for the model evaluation in the scope of this study. Rashid et al. [55] published the building-wise total energy consumption and occupancy data from different buildings, while Kriechbaumer and Jacobsen [56] published a data set that contains building-wise end energy consumption that was collected in Germany. In summary, there were two publicly available data sets that fulfilled the requirements that they were collected in office buildings, contained occupancy information and focused on MEL power consumption. These include the data sets collected in Ottawa [23] and in Frankfurt [58, 59] and they include both MEL and occupancy profiles in office buildings.

The progress in general energy consumption modeling that was not followed by the progress in MEL modeling. This could be observed not only in case of the spike in the number of open source released data sets related to general energy consumption, but also in case of the proposed models. Namely, the earlier presented increase in availability of general energy consumption data sets was closely followed by the recent progress in the development of the suitable modeling techniques [46, 60–65]. However, the majority of studies focused on smart meter data or total energy consumption profiles. On the other side, there have been little studies that address in particular the MEL modeling and exploring the relationship between MEL profiles and the predictive power of past MEL and occupancy profiles.

In the scope of this study, monitoring data from a single building is used for model development, while the evaluation was conducted using the data from three different buildings. The buildings’ characteristics are summarized in Table 2, while the observed mean daily occupancy and MEL profiles are presented in Figure 1.

The first building, located in Abu Dhabi, UAE is a university research building. The data were collected in an open space office, typically used by the full-time-employed research staff and graduate students. The building was operated 24/7, while the typical working hours were from 9 a.m. to 6 p.m., and the workdays were Sunday to Thursday. The plug-in data were collected using smart meters in each socket on the work station. The occupancy was measured occupant-wise using occupancy sensors. Eventually, the measured data were logged in a database using a 15 minutes’ frequency.

Table 2: Overview fo the buildings’ characteristics and monitoring design.

data set	Abu Dhabi	Frankfurt	Ottawa
building type	university building	bank office building	university building
space layout	open space	single offices	single offices
monitoring period	03.04.2017 - 01.12.2017	01.01.2005 - 31.12.2006	01.11.2014 - 06.07.2015
total monitoring duration	8 months	24 months	8 months
logging frequency	15 min	10 min	60 min
in-situ observed occupants	8	3	10

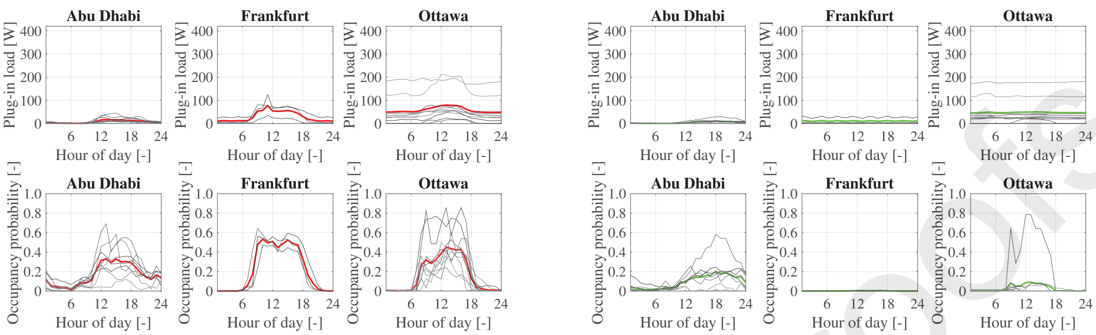
The second building was a bank office building located in Frankfurt, Germany. For further information about the monitoring design and experimental setting, the reader is referred to Kleber [58]. The MEL data were available for two single or double offices, using smart meters in each socket. Here, the MELs and binary occupancy were summed office wise, and no information about the plugged-in devices was available. The logging frequency was 10 minutes, and the data were collected over two years.

The third building in question is an academic building located in Ottawa, Canada. This was a publicly available data set, collected and released by Gunay et al. [23]. According to the original paper associated with the data set [23], the data was collected in ten private offices between November 2014 and July 2015. The MEL data was collected using meters in each socket, while the occupancy consisted of preprocessed data collected by motion sensors.

The daily occupancy and MEL consumption profiles strongly diverged between the three buildings in question (Figure 1). The open space office in Abu Dhabi was commonly occupied outside of working hours or on the weekends,

[1] https://www.journals.elsevier.com/energy-and-buildings/mendeley-data

while single/two person offices in the bank building in Frankfurt were rarely occupied outside of typical working hours. In addition, it can be observed that the measured MELs vary in magnitude and values distribution in each building.



(a) Workday profiles.

(b) Weekend profiles.

Figure 1: Measured MEL and occupancy profiles in the three buildings located in Abu Dhabi, Frankfurt and Ottawa during workdays (a) and weekends (b). The occupant-wise profiles were presented as black-colored lines, while the mean value over the corresponding data set was presented as thick red- or green-colored lines.

Prior to the model development, the data preprocessing of all three data sets was conducted. The variables from the data set collected in Abu Dhabi included occupancy state, measured MEL consumption as well as the temporal information such as the date and the time of the day, while the data were saved in .csv files. The occupancy state for each workstation was saved as a string that had values "absent" or "occupied", which were assigned the binary values in the scope of preprocessing. The MELs were logged separately for each socket, and these values were summed workstation wise. The days with the missing and "Not a Number" entries were removed from further analysis, while the presence of outliers was not observed. The data were logged in 15-minutes steps, and these were subsampled to hourly time-steps. The workplace was occupied, in case the occupancy was logged at least during one time-step during that hour, while the MEL power was integrated over each hour. Finally, the data were scaled and split into training, validation and test sets.

The second data that was collected in Frankfurt was saved in a .csv file and the used variables included occupancy and energy consumption. The occupancy was logged as a binary variable, while the MEL power was available office-wise. The entries with missing values were removed from further analysis, while the outliers in case of the power consumption were defined as the values larger than 1.000 W per two-person office. The data per each 10-minutes were subsampled to hourly data following the same steps in the case of the data set from Abu Dhabi. The open source data set from Ottawa was downloaded in the .csv format as provided by the authors as the supplementary material [23]. The data was examined for missing entries and outliers following the same steps as it was the case with data sets from Abu Dhabi and Frankfurt and eventually saved in the same format as the alternative two sets.

In order to address the observed diversity in MEL consumption, a single model that would require no tuning or recalibration for application in buildings whose MEL consumption profiles strongly diverge should be given sufficient generalization capabilities. After accounting for interruptions in the data monitoring or data loss, the available monitoring data summed up in 860, 2.170 and 1.640 occupant-monitoring days from Abu Dhabi, Frankfurt and Ottawa, respectively.

4. Model development

The starting hypothesis was that the information about the plug-in consumption is stored in past MEL profiles. This pattern could be recurring over several days or on a weekly basis. As a result, it was aimed to predict MEL consumption over the following 24 hours, based on the MEL consumption over the past days. The duration of the input sequence (previously referred to as "past days") was treated as a learned hyperparameter based on the experimental results. Here, the investigated input sequence ranged between 1 and 7 days. In order to minimize the number of time-steps addressed as model input and output, the sub hourly logged monitoring data were discretized in hourly steps, and a suitable modeling approach was investigated.

As a result, the model output consisted of 24 time-steps, while the model input varied between 24 and 168 time-steps. Neural networks with LSTM cells were identified as a suitable modeling architecture, due to their ability to bridge long discrete time-series (i.e., more than 1.000 time steps) by enforcing a constant error [66], which results in a more stable training, when compared to standard recurrent neural networks (RNNs).

The second hypothesis was that the interactions between the occupants' presence and MEL might carry the information about the MEL profile over the following 24 hours. Consequently, the occupancy over the input time window was added as an additional input variable. In order to experimentally quantify the impact of the inclusion of the occupancy on the predictive accuracy, the performance for the case where occupancy was used as input was compared to the model that used only the MEL loads as the input variable. Lastly, the model was designed to have sufficient learning capacity by including multiple hidden layers with between 10 and 100 neurons each.

In order to quantify the model's generalization capabilities, the data from above mentioned three locations is used and an overview of the defined data split on training, validation and test sets are presented in Figure 2. The models were trained and validated using the data from the occupants 1-5 from the Abu Dhabi data set. In total, the training set consisted of 660 monitoring days collected on these 5 occupants. 400 days were used for model training, while the optimal model formulation, namely model validation was conducted using the resulting 260 days of previously not considered monitoring data.

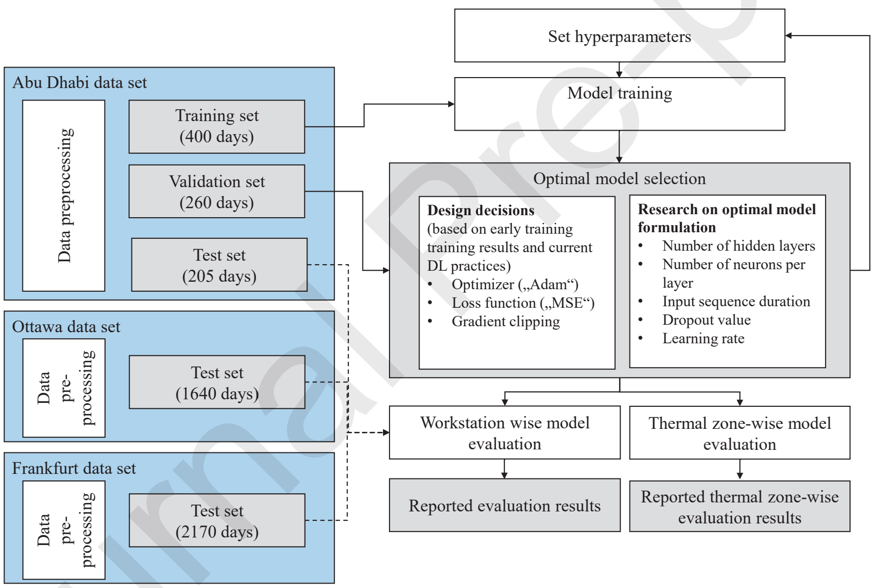


Figure 2: Methodology overview.

Eventually, the optimal model, based on the model validation results was identified and evaluated. For that purpose, the test set consisted of approximately 200 monitoring days from the Abu Dhabi data set, that were not used in the training and validation. Additionally, the model was evaluated using approximately 2170 monitoring days from the Frankfurt data set, as well as using 1640 monitoring days from the data set collected in Ottawa. In the scope of model test using independent data sets, no additional training or model adaptation was conducted.

Compared to the HVAC consumption, MEL profiles typically result in steeper change in energy consumption and in larger data imbalance. Due to these complexities, the suitable method should address the imbalanced properties of the modeled target function, as well as be suitable for time-series modeling. As a result, it was opted for LSTMs over alternative machine learning methods for HVAC or total energy modeling, that include but are not restricted to support vector machines or alternative architectures of deep neural networks. Additionally, the motivation for the use of LSTMs is their ability to capture the contextual information in relatively long temporal sequences.

LSTMs and in general gated RNNs are based on the idea of creating paths through time that have derivatives that neither vanish nor explode [39]. Due to their stability during model training, LSTMs have proved successful in modeling dependencies over 1000 time-steps [66], as well the sequences of a dynamic length. By adding the internal recurrence, the overall architecture of an LSTM cell consists of following four gates as visualized in Figure 3 (upper left corner) [39]:

- **Forget gate** (red colored lines). The forget gate is responsible for including or forgetting the information provided at the current time-step. In case the currently observed information is of relevance, the forget gate is assigned the value 1, while the forget gate has the value 0 in case the information does not contribute to the prediction[2].
- **Internal state** (green colored lines). The internal state gate stores the information over many time-steps that is relevant for the prediction. Given the newly observed information is also relevant, it will be kept by this gate over the following time-steps. An illustrative example for the use of internal gate could be, that the network learns that the feature occupancy is relevant for the MEL prediction.
- **External input gate** (blue colored lines) is responsible for updating the values of the relevant observed information. For instance, if the internal state detected feature "occupancy" as relevant, and it's value changed on the current time-step, the change in the value will be updated through external input state.
- **Output gate** (violet colored lines) gives the relevant information from the "internal memory" that consisted of the combination of the latter three states that is relevant for the current prediction.

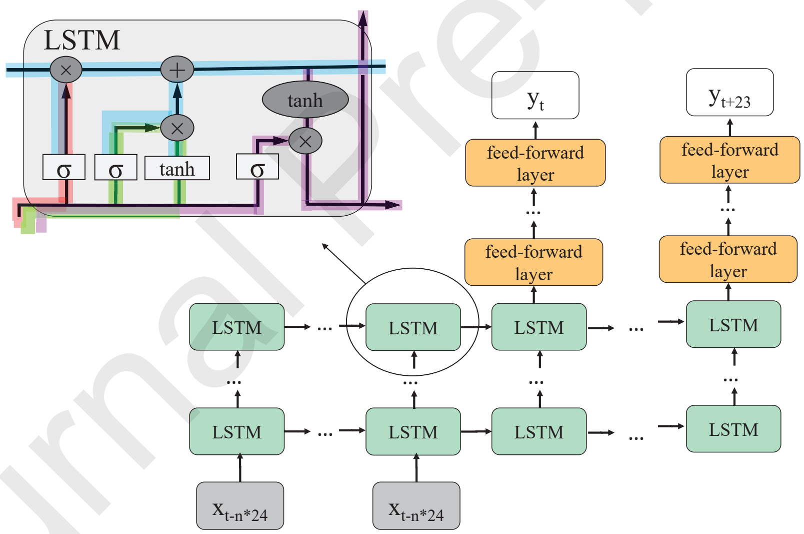


Figure 3: The architecture of the trained MEL predictive model. The visualization of the LSTM cells was adopted from Olah [67].

The architecture of the proposed neural network is presented in Figure 3. The model development included the search for the optimal model architecture that includes sufficient learning capacity. The explored models' hypotheses include the architectures with a single LSTM layer, as well as stacking multiple LSTM and feed-forward layers. The application of multi-layer architectures was motivated by the experimental results on RNNs and LSTMs applications for alternative modeling objectives [68]. Additionally, the inclusion of the feed-forward layers was explored due to the lower complexity of the feed-forward units in comparison to the LSTM units. The explored models consisted of

[2]The use of forget gate could be illustratively presented using an example from the natural language processing. Assume we aim to predict consumer's attitude towards re-buying a product based on the submitted review after the initial purchase formulated as *"I bought the product in blue color and I look forward to using it again"*. The part of the sentence "*in blue color and* "would be processed by forget gate, since it does not contribute to prediction about re-buying it.

the input layer, LSTM layers, followed by feed-forward layers and the output layer. The model input consisted of the hourly occupancy and MEL over the previous days. Here, the optimal input sequence duration was investigated in the range between 1 and 7 days, which resulted in 2 input variables with 24-168 input steps.

The LSTM layers were used as the input to the feed-forward layers, while the model output consisted of the predicted MEL over the following 24 hours. The LSTMs used "*tanh*" as activation function, while the rectified linear units (ReLU) were used as activation in the feed-forward layers.

According to the modern deep learning practices, the random or greedy search are usually the optimal hyperparameter searching approaches [69]. Given the large number of hyperparameters, the balance between accuracy, complexity, and computing needs can indeed become challenging to achieve with a grid search. However, in the scope of this study, the optimal problem hypothesis could be searched with a relatively small hyperparameter space (by testing around 10.000 model hypotheses). Therefore, grid search was deemed an appropriate method in the case of sthis study and it resulted in satisfactory results at acceptable time and computing costs. An optimal number of hidden LSTM layers was investigated in the range between 1 and 3, while the optimal number of feed-forward layers was investigated between 0 and 3. A number of neurons per hidden layer were analyzed separately for LSTM and feed-forward layers.

The narrow networks with multiple layers perform as well as the networks with more units per each layer and they lead to lower model complexity and lower number of learned weights [70]. Additionally, as pointed out by [71], zero is defined as optimal value for most of the weights in the deeper architectures. As a consequence, the optimal number of neurons was searched in range between 10 and 100 neurons. As a consequence, the investigated number of neurons per layer was {10, 50, 100}.

The duration of the input sequence was handled as an additional hyperparameter, and the optimal value was searched in the range between 1 and 7 days. The used optimizer was "Adam", while the mean squared error (MSE) was used as a loss function. The gradients were clipped to 0.1 for stability reasons. Eventually, the impact of the dropout was analyzed, in order to optimize the networks' complexity. For that purpose, the experiments were conducted where the dropout rate was 0.5 and 0.

4.1. Implementation of existing MEL models

In order to compare the performance of the proposed model with alternative approaches, a set of existing MEL predictive models was implemented. These existing models were then evaluated using the same test sets which consisted of the data from the three buildings in question. The model "A" was a Weibull distribution, originally proposed by Mahdavi et al. [5]. Here the starting hypothesis was, that the MEL consumption could be formulated as the function of presence or absence duration and the total installed plug-in power. The coefficients for the Weibull distribution were set as defined in the original study and the plug-in loads were sampled using the inverse transformation method. The parameter "total installed power" was replaced by the maximal measured plug-in power in the training set. Namely, there were 6 installed and monitored sockets on each workstation. The installed power was known for five sockets, while there was no information regarding the devices plugged in the sixth socket, that could include, but were not restricted to the water heater, phone chargers or laptop charger. Similarly, no information about the total installed power was available for the data collected in Frankfurt and Ottawa, the maximal measured power was used instead of measured power also in case of these data sets.

In contrast to the LSTM-based proposed model, the predictions were made single step ahead, instead of for the next 24 hours, which corresponds to scenario "A" as presented by the original study. Due to the probabilistic nature of the proposed method [5], the test was repeated 100 times for reliability reasons. Eventually, the reported test performance was computed as mean values over 100 repeated model tests.

The model "B" was a GMM, as proposed by Gunay et al. [23]. Here, a mixed Gaussian distribution with two principal components was fitted to predict the MEL, using presence and absence duration as model inputs. The three GMMs were fitted for each class of absence/presence duration, as proposed by the original study. In order to make a fair comparison with the alternative approaches, the model was fitted using the training set as defined in section 2. Similarly to the model "A", the prediction was made for the following hour, since the information regarding the occupancy was unknown over the 24 hours predictive horizon. The model training and test was repeated 100 times, as proposed by the original study.

4.2. Thermal zone-wise MEL prediction

In the next set of experiments, the performance of the thermal zone-wise MEL predictions was compared to the occupant-wise approach presented in previous sections. The core idea was, that the cumulative MEL consumption in the whole thermal zone would represent the sufficient granularity for the estimation of the energy consumption or thermal zone-wise heat gains. Eventually, it was aimed to compare the accuracy of the MEL predictions at the zone-level with the prediction accuracy at occupant-level granularity. For that purpose, the thermal zone-wise MEL was summed and a predictive model was trained. The thermal zone-wise MEL were collected on the 6 workstations (occupied by 8 occupants due to incoming/outgoing grad students), two occasionally occupied desks and a meeting table in the center of the room on which a plug-in was also available. In addition, the occupancy count was available, and it was assumed that the maximal occupants' number was 10 (6 workstations, 2 desks and 2 occupants at the meeting table). This model's performance was tested using only the Abu Dhabi test set. The reason for that is, that only the office from the Abu Dhabi data set had an open space layout, while the other offices were single- or double occupied.

For the model training and test, the hyperparameter set was identical to the optimal model formulation that was identified for the occupant-wise prediction. In total, three model variations for the thermal zone-wise modeling were implemented and evaluated. The first case was the LSTM with past MEL and occupancy as inputs, while the hyperparameter was defined based on experiments from the previous section. The second model was the LSTM where the only input was the past MEL consumption with the architecture defined based on earlier conducted hyperparameter tuning. Eventually, the third model (in further text referred to as model "C") was the LSTM as proposed by Wang et al. [27], with MEL as a single input and the architecture as proposed by the corresponding study.

5. Results and model evaluation

5.1. General observations during model training

Based on the training and evaluation results analysis, the network saturation was observed as one of the key challenges. The network saturation could be defined as the inability to propagate gradients well [72]. In the case of network saturation, the proposed neural network would predict similar output, even in case they were shown significantly different values of the input variables. This insensitivity towards different model input values still leads to satisfying validation results, and it could therefore be hardly detected by analysing the performance. In summary, the saturated networks would lead to similar and acceptable accuracy in case of different model hypotheses, however, they would not outperform a rule-based or over-fitted model. Based on the current state of the research, no formal metrics was adapted as a measurement of the network saturation. Therefore, in the scope of this study, network saturation was considered as the variance lower than 10 % between each of the predicted 24-steps time-series.

In spite of the normalized input features, the use of initialization as proposed by Glorot and Bengio [72], and the use of suitable activation function [73], the network saturation occurred in more than 70 % of the investigated trained models. As an example, validation results for one hyperparameter combination where the network saturation occurred are presented in Figure 4. Such a model output, however, does not result in generic and adaptive models' capabilities. Consequently, the proposed model of high complexity does not result in better performance, when compared to static MEL profiles. As a result, the hyperparameters that led to the saturated networks were labeled as non-optimal model combinations.

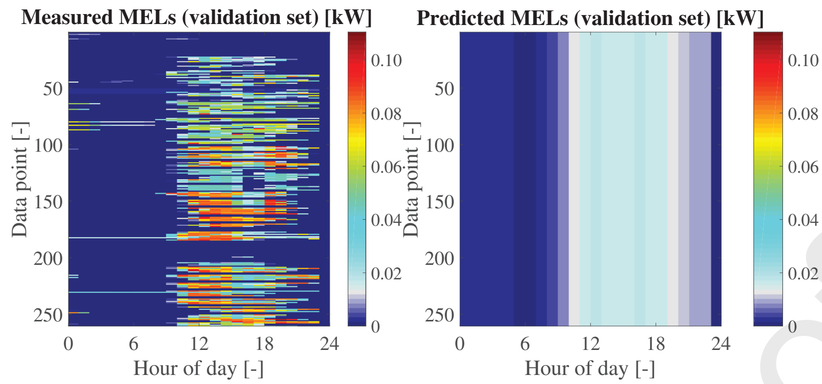


Figure 4: Visualization of the predicted MEL profiles in case of saturated neural network. Here, each row represents a daily MEL profile, The color bars refer to the MEL power in kW.

Additionally, the predictions that were seen as statistically unlikely outliers were removed from the evaluation. These were defined as MEL power over more than 2000 W per occupant. They occurred in 0.8 % of data points from data set collected in Frankfurt, and 0.4 % of data points from the measurements in building in Ottawa, in both cases segmented over data points in a single week.

5.2. *Optimal model formulation*

Based on the model's validation performance, the optimal model architecture and hyperparameters were selected. The performance on the validation set showed no improvement, in case of a longer input sequence duration. Namely, the optimal MRE on the validation set was in the same range for a varied number of input days (Figure 5). In addition, the mean value over all training combinations, excluding cases where the network saturation of gradient explosion occurred was in a similar range for varied input duration. This implies that the information-rich part of the input sequence was a learned parameter.

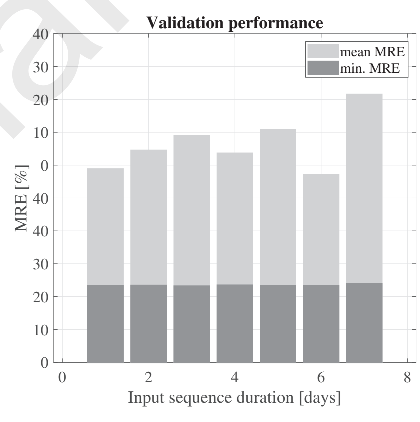


Figure 5: Mean and minimal MRE on the validation set for varied input sequence duration. The results were computed as a mean value over all hyperparameter combinations were past MEL and occupancy were used as model inputs.

Consequently, training the model with longer than necessary input vector would not result in lower performance on the validation set. As a result, the model input consisted of the MEL and occupancy during the previous 24 hours. The neural network with 3 LSTM layers and a single feed-forward layer was identified as the suitable architecture. The LSTM and feed-forward layers had 50 and 100 hidden units, respectively. The activation function used for the recurrent layers was *tanh*, while the ReLU was applied to the feed-forward layers. Learning rate was set to 0.01, and the gradients were clipped to 0.1. No dropout was included.

5.3. Performance evaluation

The test results are summarized in Table 3, showing that the normalized MRE ranged between 3 % and 13 % for the cases where the past occupancy was used as an input for the LSTM network. In addition, the normalized RMSE was in the range between 4 % and 23 %.

The results for the cases where the model inputs consisted only of plug-ins was compared to the case where inputs included plug-in loads and occupancy. The inclusion of the occupancy as additional model inputs reduced prediction MRE for approximately 3 %. The error was also reduced in terms of MSE and RMSE for 0.7 and 7 W, respectively. However, the inclusion of occupancy resulted in same MRE in case of the Frankfurt data set that ranged between 0.030 and 0.031. In case of the data set collected in Ottawa, the inclusion of occupancy lead to similar results, where difference between the both cases was lower than 1 % point.

The time required for hourly MEL prediction (final models' forward pass) was approximately 0.002 seconds clock time for between 100 and 1000 occupant-days. These experiments were conducted in CPU mode on a single laptop with Intel(R) Core(TM) i7-4710MQ CPU (2.50GHz) processor.

5.3.1. Comparison to existing models

Eventually, the performance was compared to the results obtained by the alternative models. The absolute results are expressed in *Watts*, while the relative results were obtained by dividing the predicted MEL power with the maximal measured MEL power on each data set used for testing. Here, the maximal measured hourly MEL power were 364,8 W in Abu Dhabi data set, 900 W in case of the data set collected in Frankfurt and 367,3 W in the data set collected in Ottawa. The comprehensive results are summarized in Table 3, while a sample weekly course of measured and predicted values is presented in Figure 6. The non-calibrated model earlier developed by Mahdavi et al. [5], resulted in stable performance in case of all three data sets, but higher error metrics in comparison to the developed model. In case of model "A", the resulting MRE on three data sets ranged between 24,7 % and 27,7 %, while the N-RMSE ranged between 32,5 % and 36,9 % Compared to model "A", the LSTM-based model resulted in higher accuracy, since it led to between 8-23 % points lower MRE and additionally to 13-30 % points lower N-RMSE.

Table 3: Performance evaluation on the used data sets.

	MRE [%]	MSE [W]	N-MSE [%]	RMSE [W]	N-RMSE [%]
Abu Dhabi					
plug-in + occupancy	7,9	0,3	0,4	19	20,3
plug-in	10,2	0,2	0,2	12	13,6
Model ”A” (Mahdavi et al. [5])	24,9	1	1,1	33	33,3
Model ”B”(calibrated) (Gunay et al. [23])	8,5	0,4	0,4	20	19,6
Frankfurt					
plug-in + occupancy	3,1	3	0,3	54	6,0
plug-in	2,9	2	0,2	44	4,9
Model ”A” (Mahdavi et al. [5])	26,6	107	11,9	324	36,0
Model ”B”(calibrated) (Gunay et al. [23])	29,0	7	2,2	140	15,6
Ottawa					
plug-in + occupancy	13,4	7	1,8	82	22,4
plug-in	13,0	6	1,6	77	20,9
Model ”A” (Mahdavi et al. [5])	24,7	15	4,0	119	32,5
Model ”B”(calibrated) (Gunay et al. [23])	14,4	7	1,8	81	22,1
Thermal zone-wise prediction (Abu Dhabi)					
plug-in + occupancy	13,3	286	10,6	535	19,8
plug-in	13,6	307	11,3	554	20,5
Model ”C” (Wang et al. [27])	14,2	439	16,2	663	24,5

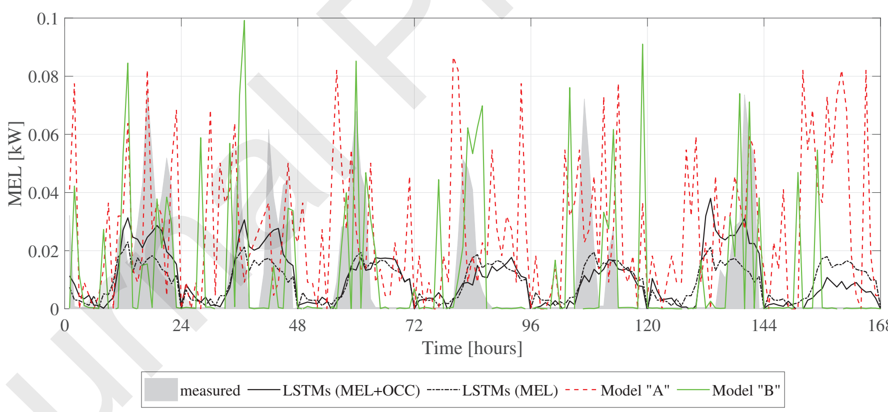


Figure 6: Comparison of the measured and predicted MEL profiles using proposed method, model ”A” and ”B” on the Abu Dhabi data set.

The performance of the day-ahead proposed LSTM method was compared to the hour-ahead results from the model “B”. Compared to model “B”, the proposed method led to lower error in terms of all used metrics on data sets from Abu Dhabi and Frankfurt. The test on the data set collected in Ottawa, the performance of the proposed method was in the same range, when compared to the model proposed by Gunay et al. [23].

In summary, the findings help answer the research questions that were stated in section 1. Starting with accuracy, the results of Table 3 show that the proposed LSTM architecture generally improves the MEL prediction when compared to existing models.

Similarly to the inputs used for models “A” and “B”, the proposed LSTM-based model used the information about the relationship between the occupancy/absence duration and the MEL loads. In addition, the developed model aimed to identify the contextual information between the target MEL and the MEL and occupancy profiles on the previous days. In summary, the increase in the learning capacity and referring to the occupancy and MEL profiles in the auto-regressive manner could be identified as the major design decisions that lead to the improvement in the model's accuracy.

This accuracy is further improved by the inclusion of occupancy patterns from the past days. Finally, and more importantly, the results confirm that the proposed model can be applied to different buildings without any model calibration or adaptation, which is a unique contribution of this work compared to previous efforts in the literature.

5.3.2. Limitations

The previous section showed that the proposed method led to the significant improvement in terms of absolute and relative performance when compared to the alternative approaches on all three data sets. Such conclusion was based on the values observed for performance metrics such as MRE, MSE, N-MSE, RMSE, and N-RMSE, which are commonly in the literature for benchmarking purposes.

The aim of this section is to go beyond these metrics and see if the model is effective at predicting MEL for specific occupants on specific days. For visibility reasons, one week of data (chosen randomly) was graphically presented in high resolution. Prior to presenting the results, the performance metrics for the chosen weekly data were compared to the performance of the whole data sets (Figure 7).

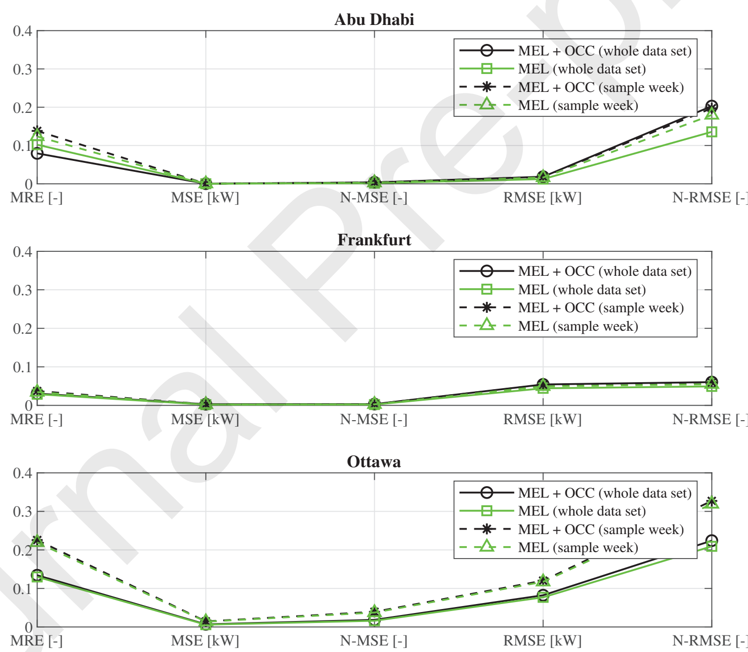


Figure 7: Evaluation metrics comparison of the representative week presented in the Figure 8, and the evaluation performance evaluated on the whole evaluation set.

This analysis showed that the accuracy for the presented sample of weekly data had less than 10 % points deviation from the overall test performance and could therefore, be used to depict the general relationship between the predicted and measured MEL profiles.

The comparison between actual and predicted MEL for the three occupants in the different locations is shown in Figure 8. The actual MEL levels of the occupants' are shown in the solid-line curves, while their predicted values are shown in dotted lines. Overall, there is a clear discrepancy between the predicted and actual levels, especially for the two studied occupants from the Ottawa and Frankfurt datasets.

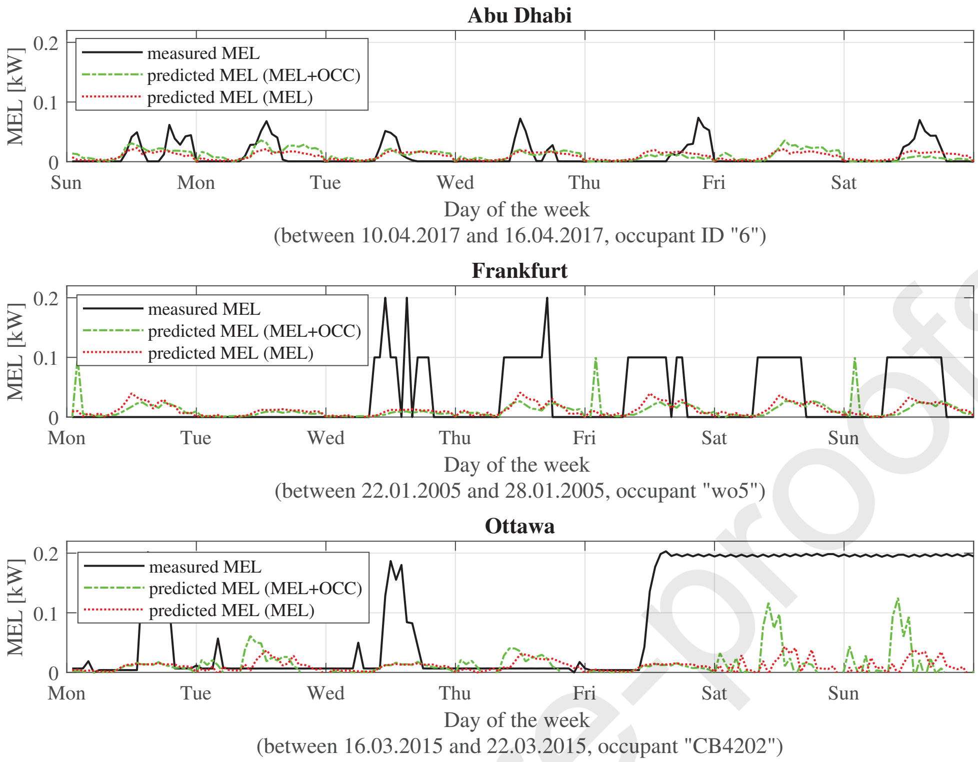


Figure 8: A weekly course of measured and predicted MEL consumption for each data set. The green colored lines referred to the predictions made using past occupancy and MEL as inputs, while the red-colored lines referred to the predictions based solely on the past MELs.

More specifically, the measured (i.e. actual) MEL levels of the occupants seemed random at times, such as the absence of MEL on Monday and Tuesday for the occupant in Frankfurt and the high consumption levels over the weekend for the occupant in Ottawa. Such randomness in OB was not properly handled and predicted by the models.

In summary, while the proposed model provided an improved modeling accuracy for total MEL consumption when compared to the state-of-the-art models (See Table 3), it showed limitations for the analysis done at the workstation level. This limitation is worth investigating further as part of future work, especially for applications that target individuals such as energy feedback mechanisms.

In particular, the predicted MEL consumption has only a limited parity with the measured profiles in temporal domain, which could be seen as a major burden for the usability of the proposed model for the practical applications. As a result, the duration, beginning and the end of each sequence where plug-in energy consumption occurred was estimated with low accuracy. Therefore, the practical applicability of the proposed model is mainly seen for estimating the internal thermal gains in order to optimize the day-ahead HVAC operation, since the thermal gains for this setting could be represented with performance improvement, when compared to alternative established models or fixed schedules. However, the current model formulation is not accurate enough for the real-time model predictive control, energy feedback mechanisms or the information inclusion in the further interfaces with user feedback functions.

5.3.3. Performance for the thermal zone-wise MEL prediction

Finally, the performance of the thermal zone-wise MEL predictive models was evaluated and the results are summarized in the bottom rows of Table 3. As detailed in the methodology section, the main difference in the thermal zone-wise MEL modeling approach is that the predictions are made based on the aggregated MEL of individual occupants, as opposed to the sum of their individual MEL predictions. The results, showed, that multiple stacked LSTM layers led to a lower error in terms of all used evaluation measurements, when compared to model "C". In addition, the use of occupancy as an additional input led to performance improvement when compared to both model formulations (occupant-wise as well as thermal zone-wise) where the MEL was used as the single model input. However, the overall accuracy for the zone-level modeling was lower, when compared to the occupant-wise model formulation. Namely, MRE for the occupant-wise evaluation on the Abu Dhabi data set was approximately 5 % lower, when compared to the zone-level model formulation. Additionally, thermal zone-wise modeling led to 286 W MSE and 535 W RMSE, while these metrics were in the range between 0.3 and 19 W in case of the occupant-wise modeling.

In summary, the findings confirm that the occupant-wise MEL modeling approach showed better accuracies than the thermal zone-wise approach, confirming the value of desk-level monitoring and analysis for prediction purposes.

6. Discussion and future work

In the scope of this study, the LSTM neural network was developed to predict MEL energy consumption. The results showed that stacking multiple LSTM layers led to performance improvement. Based on the validation performance, the architecture with three LSTM layers and a single feed-forward layer was identified as the optimal model formulation. The analysis of the training performance showed that network saturation was the key modeling challenge in developing LSTM for MEL prediction.

The optimal input sequence duration was 24 hourly steps. The relatively short optimal input sequence indicates, that the LSTM units could be replaced by the RNN units which represents a potential alternative modeling approach. Hence, the replacement of LSTMs with RNNs may not result in a more efficient problem formulation, since both LSTM training and RNN training using back propagation through time (BPTT) result in quadratic time complexity per time-step [74].

The network saturation was observed as one of the major challenges during the model training. Even though the network saturation is more common phenomena in case of the very deep network architectures, it could also be observed in case of the network sizes that are in the same range as the architectures explored in this paper. Namely, in the existing theoretical and experimental work, Glorot and Bengio [72] presented the observations on network saturation on the example of a feedforward architecture with four hidden layers. Since this study focus on architectures with up to 7 hidden layers and up to 168 propagated time-steps, the network saturation is very probable phenomenon to observe in this study. In order to solve the saturation issues the impact of different optimization algorithms and learning rates on the saturation was analyzed in the scope of initial experiments, yet to clear solution could be identified. As a consequence, the network saturation in case of the time-series related to energy consumption modeling needs to be further researched. Given the objectives of the presented work, such an extensive effort is considered beyond the scope of the current paper and should be included in future expansions of the work.

An additional issue in case of the model evaluation using the data set from Frankfurt was the different data resolution. The energy consumption data were logged in 100 W steps, which resulted in lower resolution, when compared to 1 W granularity of the other two data sets. As a consequence, the minor power loads were not registered, which is a particular issue in case of modeling at a high spatial resolution. However, it was still opted to include the data set from Frankfurt in the model evaluation, since this represents a real-world problem that is to be expected in case of monitored MEL consumption across different buildings. Therefore, the presented evaluation points out both the potential and the limitations of the proposed method. Consequently, this can be beneficial for comparing the performance of the presented model on alternative low resolution data sets.

Eventually, the performance of the occupant-wise predictive performance and the modeling per thermal zone were compared. The results showed that the occupant-wise model evaluation led to approximately 5 % points lower MRE when compared to zone-level prediction. Based on these results, occupant-wise modeling was identified as the optimal modeling approach. In the proposed setting for the MEL prediction on the workstation-level granularity, no occupant wise-tuning was required. Namely, the single model was developed to have sufficient generalization capabilities, and it was trained using a subset of occupants from one building, while it was evaluated using the previously unseen occupants from multiple other buildings. Eventually, the pre-trained model was applied in the form of an agent-based model to each occupant to act individually.

As presented in section 5.3.2, one of the major limitation of presented method is caused by the low parity of measured and predicted energy consumption in temporal domain. In order to bridge the gap towards the better model applicability, the better overlap between the phases of measured and predicted MEL energy consumption is required.

This could be achived by further improving the formulation of the LSTMs based models, or potentially by relying on probabilistic graphical models that include the information about the sequence durations and state changes. For that purpose, the Semi Markov Models may be a promising alternative modeling approach.

Potential applications of the developed MEL predictive model include the incorporation in smart HVAC control to predict the internal heat gains caused by MEL. For instance, the model could be used to predict the MELs profiles over the following day and provide a better estimate of the cooling load requirements. For that purpose, the cooling load estimation could be obtained using simulation, using a simulation test-bed, or using some alternative, data-driven approach. In practice, the burden to include MEL internal gains in the HVAC is that the MEL is typically not defined as one of the macros in building automation system (BAS) according to the existing room automation guidelines such as the German guideline VDI 3813:2015 [75]. Therefore, they were often not considered as one of the inputs in the HVAC control. On the other side, improvements in the building physics [27], and the recent increase in installed computational power, such the expanding use of graphics processor units (GPUs), raise the impact of the MEL on the internal heat gains and therefore resulting HVAC energy consumption.

As underlined in the previous paragraph, the main potential application of the presented method for MEL prediction is to consider the internal thermal gains in the HVAC control. For that purpose, a zone-wise calculation of the thermal gains would present a sufficient granularity for estimating the required HVAC supply. Hence, the thermal zone-level granularity results in suboptimal problem formulation and lower accuracy. To conclude, the occupant-wise modeling does present higher granularity than required by the end purpose, but it leads to better problem formulation and higher accuracy. In summary, applying the occupant-wise model would result in significantly improved accuracy on costs of higher granularity. At the same time, the algorithmic runtime complexity remains satisfying for the prediction estimation. On this place, the O(n) time-complexity instead on O(1), as it was the case for thermal zone-wise modeling. To conclude, these results underline that the use of occupant-wise models would be the suitable way for representing the MELs in an open space office.

An alternative potential application of the developed model can be electrical consumption load scheduling. This can be of particular importance in cases where the electrical energy comes from renewable resources. In that case, a reliable prediction of a day ahead energy consumption may be beneficial for the economic distribution of electricity generated from wind and solar power.

Lastly, a scientifically important contribution of this paper is the conducted round-robin evaluation of two existing MEL predictive models using a relatively large data sample. These results showed, that the accuracy of both GMMs and Weibull distribution was in the same range when tested on alternative data sets when compared to the results presented in the original studies. All compared models used the occupancy and historical information as model inputs, while the modeled output was the plug-in load consumption. On this place, two crucial points have to be considered in scope of the model comparison. The predictive horizon in case of the models “A” and “B” was one hour, while their performance was benchmarked against the developed model that provided predictive performance over 24 hours. This difference goes in favor of the existing models “A” and “B”, but it has to be considered in scope of the comprehensive elaboration on the results. However, the proposed model could outperform the existing models on all used data sets. Secondly, the literature screening pointed out that the most established models (except the model “C”) are static models. Due to the limited exiting research on time-series based MEL predictive models, no benchmarking against sequential models was conducted. Nonetheless, we hope that the predictive performance improvement achieved by relying on time-series based modeling will spark the interest of the research community for the further work in this direction.

The model was developed using the data from Abu Dhabi and eventually tested with the data from unseen occupants form Abu Dhabi, as well as with data from Frankfurt and Ottawa. The model required no further training nor alternative domain adaptation steps prior to the application for the different buildings. Consequently, the models’ generalization capabilities to agnostic commercial buildings were evaluated. These results could be used a baseline for the models’ further development and the future work on addressing a large number of different occupants and buildings with a single-time model training.

From a modeling perspective, an important but unsolved challenge in the scope of this study was the implementation of the sliding window for the model input. Namely, the current version of the proposed method uses the monitoring data as input from 00:00 until 23:59, to predict the energy consumption over the following day, strictly starting at 00:00 over 24 hours duration. However, a model formulation that uses historical data starting from an arbitrary hour of the day could potentially further improve the model’s generalization capabilities, especially for addressing the buildings with flexible operational hours. In the scope of the current study, the sliding window technique did not lead to satisfactory validation results.

In addition, future studies should address the model evaluation using alternative data sources, including, but not restricted to different building layouts or alternative building usages.

7. Conclusion

This paper presented an approach for predicting the day-ahead energy plug-in loads using LSTMs. The developed model was evaluated using data from three buildings, located in Abu Dhabi, Frankfurt, and Ottawa. The main contribution of the presented work is the development of the MEL predictive model that does not require occupant-wise or building-wise model training nor model adaptation while achieving competitive accuracy. In addition, the key findings can be summarized as follows:

- including binary occupant-station wise presence in the input slightly improved the prediction accuracy and led to more reliable of dynamic signals of MEL,
- contrary to our starting hypothesis, considering multiple past days as model input did not improve the evaluation accuracy,
- hence, the duration of the time window where the information regarding future actions was stored could be a learned parameter. Consequently, considering longer input sequences did not have a negative impact on the predictive performance,
- the model showed competitive evaluation accuracy when applied to alternative buildings without any additional model calibration,
- the achieved accuracy was improved on all three data sets, when compared to the performance of the existing approaches.

8. Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – TR 892/4-1. We thank Marcel Schweiker and Andreas Wagner of Karlsruhe Institute of Technology, Germany, for sharing the data set ”KfW Ostarkade”, which was used for model evaluation. This work is also part of the research activities of IEA EBC Annex 79.

References

[1] International Energy Agency. (2019). Energy Efficiency: Buildings. The global exchange for energy efficiency policies, data and analysis. https://www.iea.org/topics/energyefficiency/buildings/.

[2] Fayyad, M., John, J. (2017). Defining Nearly Zero Energy Buildings in the UAE – 2017 Emirates GBC Report. Emirates Green Building Council, 2017.

[3] Al Amoodi, A., Azar, E. (2018). Impact of Human Actions on Building Energy Performance: A Case Study in the United Arab Emirates (UAE). Sustainability, 10(5), 1404.

[4] Rauch, E. M. (2011). Assessing and Reducing Miscellaneous Electric Loads (MELs) in Lodging (No. PNNL-21055). Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

[5] Mahdavi, A., Tahmasebi, F., Kayalar, M. (2016). Prediction of plug loads in office buildings: Simplified and probabilistic methods. Energy and Buildings, 129, 322-329.

[6] Azar, E., Menassa, C. C. (2014). A comprehensive framework to quantify energy savings potential from improved operations of commercial building stocks. Energy Policy, 67, 459-472.

[7] Yan, D., O'Brien, W., Hong, T., Feng, X., Gunay, H. B., Tahmasebi, F., Mahdavi, A. (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. Energy and Buildings, 107, 264-278.

[8] American Society of Heating, Refrigerating and Air-Conditioning Engineers. (2016). Standard 90.1-2016 – Energy Standard for Buildings Except Low-Rise Residential Buildings.

[9] Verein Deutscher Ingenieure. (2015). Calculation of thermal loads and room temperatures (design cooling load and annual simulation) (VDI 2078:2015).

[10] Lin, M., Azar, E. (2019). Mixing work and leisure? Energy conservation actions and spillovers between building occupants at work and at home in the UAE. Energy Research and Social Science, 47, 215-223.

[11] Masoso, O. T., Grobler, L. J. (2010). The dark side of occupants' behaviour on building energy use. Energy and buildings, 42(2), 173-177.

[12] Bennet, I. E., O'Brien, W. (2017). Office building plug and light loads: Comparison of a multi-tenant office tower to conventional assumptions. Energy and Buildings, 153, 461-475.

[13] Kawamoto, K., Shimoda, Y., Mizuno, M. (2004). Energy saving potential of office equipment power management. Energy and Buildings, 36(9), 915-923.

[14] Ketchman, K. J., Parrish, K., Khanna, V., Bilec, M. M. (2018). Synergizing disparate component-level energy resources into a single whole building tool to support energy conservation action in small commercial buildings. Energy and Buildings, 176, 325-332.

[15] Lobato, C., Pless, S., Sheppy, M., Torcellini, P. (2011). Reducing plug and process loads for a large scale, low energy office building: NREL's research support facility (No. NREL/CP-5500-49002). National Renewable Energy Lab.(NREL), Golden, CO (United States)

[16] Menezes, A. C., Cripps, A., Buswell, R. A., Wright, J., Bouchlaghem, D. (2014). Estimating the energy consumption and power demand of small power equipment in office buildings. Energy and Buildings, 75, 199-209.

[17] Moorefield, L., Frazer, B., Bendt, P. (2008). Office plug load field monitoring report. Ecos Consulting.

[18] Nordman, B., Piette, M.A., Kinney, K. (1996). Measured Energy Savings and Performance of Power Managed Personal Computers and Monitors. Presented at the American Council for an Energy Efficient Economy Summer Study.

[19] Roberson, J. A., Homan, G. K., Mahajan, A., Nordman, B., Webber, C. A., Brown, R. E., ... Koomey, J. G. (2002). Energy use and power levels in new monitors and personal computers (No. LBNL-48581). Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

[20] Tetlow, R. M., van Dronkelaar, C., Beaman, C. P., Elmualim, A. A., Couling, K. (2015). Identifying behavioural predictors of small power electricity consumption in office buildings. Building and Environment, 92, 75-85.

[21] Webber, C. A., Roberson, J. A., McWhinney, M. C., Brown, R. E., Pinckard, M. J., Busch, J. F. (2006). After-hours power status of office equipment in the USA. Energy, 31(14), 2823-2838.

[22] Chae, Y. T., Horesh, R., Hwang, Y., Lee, Y. M. (2016). Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy and Buildings, 111, 184-194.

[23] Gunay, H. B., O'Brien, W., Beausoleil-Morrison, I., Gilani, S. (2016). Modeling plug-in equipment load patterns in private office spaces. Energy and Buildings, 121, 234-249.

[24] Mahdavi, A., Tahmasebi, F. (2017). Predictive models of electrical energy use in office buildings due to plug loads.Proceedings of the 15th IBPSA Conference.

[25] O'Brien, W., Abdelalim, A., Gunay, H. B. (2019). Development of an office tenant electricity use model and its application for right-sizing HVAC equipment. Journal of Building Performance Simulation, 12(1), 37-55.

[26] Park, J. Y., Dougherty, T., Fritz, H., Nagy, Z. (2019). LightLearn: An adaptive and occupant centered controller for lighting based on reinforcement learning. Building and Environment, 147, 397-414.

[27] Wang, Z., Hong, T., Piette, M. A. (2019). Data fusion in predicting internal heat gains for office buildings through a deep learning approach. Applied Energy, 240, 386-398.

[28] Lasternas, B., Zhao, J., Yun, R., Zhang, C., Wang, H., Aziz, A., Lam, K.P., Loftness, V., (2014). Behavior orinted metrics for plug load energy savings in office environment. ACEEE Summer Study on Energy Efficiency in Buildings

[29] Wagner, A., O'Brien, W., Dong, B. (Eds.). (2017). Exploring occupant behavior in buildings: methods and challenges. Springer.

[30] D'Oca, S., Gunay, H. B., Gilani, S., O'Brien, W. (2019). Critical review and illustrative examples of office occupant modelling formalisms. Building Services Engineering Research and Technology.

[31] Sun, K., Hong, T. (2017). A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures. Energy and Buildings, 146, 383-396.

[32] Sun, K., Hong, T. (2017). A simulation approach to estimate energy savings potential of occupant behavior measures. Energy and Buildings, 136, 43-62.

[33] Zhao, J., Lasternas, B., Lam, K. P., Yun, R., Loftness, V. (2014). Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. Energy and Buildings, 82, 341-355.

[34] Papadopoulos, S., Azar, E. (2016). Integrating building performance simulation in agent-based modeling using regression surrogate models: A novel human-in-the-loop energy modeling approach. Energy and Buildings, 128, 214-223.

[35] Azar, E., Menassa, C. C. (2016). Optimizing the performance of energy-intensive commercial buildings: Occupancy-focused data collection and analysis approach. Journal of Computing in Civil Engineering, 30(5), C4015002.

[36] Zhang, Z., Chong, A., Pan, Y., Zhang, C., Lu, S., Lam, K. P. (2018). A deep reinforcement learning approach to using whole building energy model for HVAC optimal control. In 2018 Building Performance Analysis Conference and SimBuild.

[37] Markovic, R., Grintal, E., Wölki, D., Frisch, J., van Treeck, C. (2018). Window opening model using deep learning methods. Building and Environment, 145, 319-329.

[38] Markovic, R., Frisch, J., van Treeck, C. (2019). Learning short-term past as predictor of window opening-related human behavior in commercial buildings. Energy and Buildings, 185, 1-11.

[39] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep learning. MIT press.

[40] Haldi, F., Robinson, D. (2009). Interactions with window openings by office occupants. Building and Environment, 44(12), 2378-2395.

[41] Rijal, H. B., Tuohy, P., Nicol, F., Humphreys, M. A., Samuel, A., Clarke, J. (2008). Development of an adaptive window-opening algorithm to predict the thermal comfort, energy use and overheating in buildings. Journal of Building Performance Simulation, 1(1), 17-30.

[42] Schweiker, M., Haldi, F., Shukuya, M., Robinson, D. (2012). Verification of stochastic models of window opening behaviour for residential buildings. Journal of Building Performance Simulation, 5(1), 55-74.

[43] Laurent, J. G. C., Samuelson, H. W., Chen, Y. (2017). The impact of window opening and other occupant behavior on simulated energy performance in residence halls. Building Simulation Vol. 10, No. 6, 963-976. Tsinghua University Press. https://doi.org/10.1007/s12273-017-0399-3

[44] Langevin, J., Wen, J., Gurian, P. L. (2015). Simulating the human-building interaction: Development and validation of an agent-based model

of office occupant behaviors. Building and Environment, 88, 27-45.

[45] Wolf, S., Wölki, D., Robinson, D., van Treeck. C. (2017). Evaluation and Re-training of two window opening models using an independent dataset. In: Healthy Buildings Europe Conference.

[46] Miller, C. (2019). More Buildings Make More Generalizable Models—Benchmarking Prediction Methods on Open Electrical Meter Data. Machine Learning and Knowledge Extraction, 1(3), 974-993.

[47] Peng, Y., Rysanek, A., Nagy, Z., Schlüter, A., (2017). Occupancy learning−based demand−driven cooling control for office spaces. Building and Environment, 122, 145-160.

[48] Gjoreski, M., Kalabakov, S., Luštrek, M., Gams, M., Gjoreski, H. (2019). Cross−dataset deep transfer learning for activity recognition. In Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers (714−718).

[49] Arief−Ang, I.B., Hamilton, M., Salim, F.D. (2018). A scalable room occupancy prediction with transferable time series decomposition of co2 sensor data. ACM Transactions on Sensor Networks (TOSN), 14(3-4), 1-28.

[50] Zhang, T., Ardakanian, O. (2019). A domain adaptation technique for fine-grained occupancy estimation in commercial buildings. In Proceedings of the International Conference on Internet of Things Design and Implementation (pp. 148-159).

[51] Torrey, L., Shavlik, J. (2010). Transfer learning. In Handbook of research on machine learning applications and trends: algorithms, methods, and techniques (pp. 242-264). IGI global.

[52] Martin, R.A., Poll, S. (2014). Energy Analysis of Multi−Function Devices in an Office Environment. ASHRAE Transactions, 120(1). Data set available under https://data.nasa.gov/w/xd8r−vngv/default?cur=olfQKm_VPG

[53] Doherty, B., Trenbath, K. (2019). Device-level plug load disaggregation in a zero energy office building and opportunities for energy savings. Energy and Buildings, 204, p.109480.

[54] Doherty, B., Trenbath, K. (2019). Dataset: Data for: Device-level plug load disaggregation in a zero energy office building and opportunities for energy savings

[55] Rashid, H., Singh, P., Singh, A. (2019). I−BLEND, a campus−scale commercial and residential buildings electrical energy dataset. Scientific data, 6, p.190015.

[56] Kriechbaumer, T., Jacobsen, H.A. (2018). BLOND, a building−level office environment dataset of typical electrical appliances. Scientific data, 5, p.180048.

[57] Hebrail, G., Berard, A. (2012). Individual household electric power consumption Data Set. UC Irvine Machine Learning Repository. Data available under https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption

[58] Kleber, M. (2006). Abschlussbericht KfW Ostarkade. Universität Karlsruhe (TH).

[59] Schweiker, M., Kleber, M., Wagner, A. (2019). Long−term monitoring data from a naturally ventilated office building. Scientific data, 6(1), 1-6.

[60] Miller, C., 2019. What's in the box?! Towards explainable machine learning applied to non-residential building smart meter classification. Energy and Buildings, 199, pp.523-536.

[61] Park, J.Y., Yang, X., Miller, C., Arjunan, P., Nagy, Z. (2019). Apples or oranges? Identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset. Applied energy, 236, 1280-1295.

[62] Park, J.Y., Wilson, E., Parker, A. Nagy, Z. (2020). The good, the bad, and the ugly: Data-driven load profile discord identification in a large building portfolio. Energy and Buildings, 215, 109892.

[63] Zhan, S., Liu, Z., Chong, A., Yan, D. (2020). Building categorization revisited: A clustering−based approach to using smart meter data for building energy benchmarking. Applied Energy, 269, 114920.

[64] Zufferey, T., Ulbig, A., Koch, S., Hug, G. (2016). Forecasting of smart meter time series based on neural networks. In International workshop on data analytics for renewable energy integration (pp. 10-21). Springer, Cham.

[65] Piscitelli, M.S., Brandi, S., Capozzoli, A. (2019). Recognition and classification of typical load profiles in buildings with non-intrusive learning approach. Applied Energy, 255, 113727.

[66] Gers, F. A., Schmidhuber, J., Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.

[67] Olah, C. (2015). Understanding LSTM networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[68] Graves, A., Mohamed, A. R., Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In 2013 IEEE international conference on acoustics, speech and signal processing (6645-6649). IEEE.

[69] Bergstra, J., Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of machine learning research, 13(Feb), 281-305.

[70] Alvarez, J.M., Salzmann, M. (2016). Learning the number of neurons in deep networks. In Advances in Neural Information Processing Systems (2270-2278).

[71] Hu, H., Peng, R., Tai, Y.W., Tang, C.K. (2016). Network trimming: A data-driven neuron pruning approach towards efficient deep architectures. arXiv preprint arXiv:1607.03250.

[72] Glorot, X., Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (249-256).

[73] Maas, A. L., Hannun, A. Y., Ng, A. Y. (2013, June). Rectifier nonlinearities improve neural network acoustic models. In Proc. ICML (Vol. 30, No. 1, p. 3).

[74] Hochreiter, S., Schmidhuber, J. (1997). LSTM can solve hard long time lag problems. In Advances in neural information processing systems, 473-479.

[75] Verein Deutscher Ingenieure. (2015). Building automation and control systems (BACS): Fundamentals for room control (VDI 3813-1:2015).