Solar Energy 207 (2020) 626–639

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Short-term energy use prediction of solar-assisted water heating system: Application case of combined attention-based LSTM and time-series decomposition

Amirreza Heidari*, Dolaana Khovalyg

Thermal Engineering for the Built Environment Laboratory (TEBEL), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ARTICLE INFO

Keywords:
Solar water heating
Energy use forecast
Machine learning
Attention mechanism
LSTM neural network
Time series decomposition

ABSTRACT

With improved insulation of building envelopes and the use of low-temperature space heating systems, the share of energy use for domestic hot water (DHW) production in buildings has increased significantly, and nearly become the most energy-expensive service in modern buildings. Early prediction of the energy use for DHW is required for many advanced applications such as smart control, demand-side management, and optimal operation of electric or heat storage. However, predicting energy use of the solar-assisted water heating system is more challenging than typical DHW systems, as it is strongly affected by two stochastic phenomena, demand pattern and solar radiation.

Given the increasing use of solar-assisted water heating systems, this paper aims to evaluate the potential to predict energy use in such systems using a novel machine learning approach. In this novel model, a Long-Short Term Memory (LSTM) neural network is enhanced by (1) implementing the attention mechanism, a recent development in deep learning inspired by human vision to pay selective attention to the input data, and (2) decomposition of input data into sub-layers. The performance of simple LSTM neural network, Attention-based LSTM neural network (ALSTM) and Attention-based LSTM using decomposed data (ALSTM-D) are compared to a Feed-Forward neural network as a baseline model. Results show that LSTM, ALSTM and ALSTM-D models have a Mean Absolute Error (MAE) of 25%, 28% and 41% lower than Feed-Forward model, respectively. These results indicate the superior performance of the proposed ALSTM-D model over conventional models for solar-assisted DHW systems.

1. Introduction

As the energy efficiency of space conditioining in buildings is improving over time, the share of energy use for hot water production is increasing significantly (Kazmi et al., 2018). In 2013, it was estimated at 16% of total heating energy use in EU28 households, while in new energy-efficient buildings this share is reported to be around 40–50% (Marszal-Pomianowska et al., 2019). Hence, the reduction of energy use for domestic hot water (DHW) production will be the key issue in improving the energy efficiency of future buildings. In recent years, the research on building energy efficiency has mainly focused on space heating and cooling needs, whereas energy use for DHW production has been overlooked (Marszal-Pomianowska et al., 2019). Considering the increasing share of energy required for hot water production, it is necessary to advance the study of DHW systems. Typically, the profiles of energy use for DHW production show many fluctuations (Booysen et al., 2019), as it is affected by stochastic parameters such as inlet water temperature or demand profile. In the case of renewable source-assisted hot water systems, energy use is even more fluctuating as this is strongly affected by the availability of renewable energy sources. Accordingly, the prediction of energy use for DHW production is challenging, especially for systems coupled with renewables. However, the prediction of their energy use is necessary for the optimal operation of the power system considering the significant share of energy use for DHW production in modern buildings (Muzaffar and Afshari, 2019). Furthermore, to include these systems for demand-side management and power shifting, the first step is to predict their energy use (Kato and Suzuoki, 2014).

According to the performed literature review, there is limited literature on the prediction of DHW demand in buildings. Among these studies, only a few researchers have directly predicted the energy use for DHW production, and the others have only predicted the volumetric

* Corresponding author.
E-mail address: amirreza.heidari@epfl.ch (A. Heidari).

https://doi.org/10.1016/j.solener.2020.07.008
Received 14 June 2020; Received in revised form 2 July 2020; Accepted 4 July 2020
Available online 13 July 2020

Nomenclature

$H_{in\ (\%)}$	Relative humidity of indoor air
LSTM	Long-Short Term Memory
MAE (kW)	Mean Absolute Error
MSE (kW)	Mean Squared Error
RMSE (kW)	Root Mean Squared Error
T_{in} (°C)	Indoor air temperature
T_{out} (°C)	Outdoor air temperature

demand of hot water, which can be the input for energy use prediction. Mathioulakis et al., (2018) developed a Feed-Forward neural network model to predict the energy use of an air-source heat pump water heater, using the outdoor air temperature and the inlet water temperature as the input features. It was shown that neural networks provide a simple and effective tool for modeling heat pump performance in various operating conditions. Dong et al., (2018) predicted the energy use of an electric water heater using three different machine learning models, including Support Vector Machine, Gaussian Naive Bayes, and Random Forest. It was indicated that the Support Vector Machine provides the best accuracy, followed by Gaussian Naive Bayes and Random Forest. Delorme-Costil and Bezian, (2017) used artificial neural networks to predict the energy use of domestic hot water for the next 8 min and the next 64 min. Three different neural network models were used, varying in their architecture and input. The models consider the number of days, ambient temperature and the lagged energy use data (in a horizon of last 8 min and last 64 min for different scenarios) as the input parameters. It was found that the model, which only uses historical data as the input feature, provides the best accuracy. Fischer et al., (2016) developed a stochastic bottom-up model to generate hourly profiles of space heating and hot water energy demand of households in Germany. Behavioral and energy balance models were coupled with stochastic methods to generate realistic load profiles. Dmytro et al., (2019) evaluated 8 different regression models for the prediction of hourly and daily hot water use in a hotel in Oslo, Norway. Evaluating the potential variables for the case study hotel indicated that the most influencing variable on the energy use for DHW production is the number of guests. Gelažanskas and Gamage, (2015) evaluated the potential of Artificial neural networks to predict the volumetric hot water demand of individual electric water heaters. Data from about a hundred dwellings were analyzed using the autocorrelation technique. After the most appropriate lags were chosen, different topologies of neural networks were tested and compared. Maltais and Gosselin, (2019) developed a hybrid Machine Learning model, integrating Recurrent neural network and Random Forest, to predict the hourly volumetric demand of hot water use in a 40 unit residential building in Quebec City, Canada. Marszal-Pomianowska et al., (2019) proposed a calculation methodology to estimate the mean hourly and the daily profiles of domestic hot water demand from hourly total heating demand. This methodology is based on a few assumptions, such as constant space heating demand during warm seasons or independence of space heating demand and indoor air temperature. The methodology was applied to a dataset of hourly space heating demand of 38 single-family houses in Denmark. De Santiago et al., (2017) used the methodology of Jordan and Vajen, (2001) to develop hot water use profiles using the data of 4 buildings monitored in Switzerland. Evaluation of the dataset shows that Swiss households follow similar patterns as other countries, such as having very little correlation with weather conditions. Table 1 presents a summary of literature on DHW demand prediction.

Solar energy, as the other stochastic driver of energy-assisted systems energy use, has been the focus of many researchers in recent years. Kalani et al., (2017) developed and compared 3 different machine learning models to investigate the efficiency and outlet temperature of a

Table 1
Summary of literature review on the prediction of energy use for DHW production.

Study	Focus	Building type	Model	Inputs	Output
Mathioulakis et al., 2018)	Energy use for DHW production	Residential	Feed-Forward neural network	Outdoor air temperature in evaporator inlet, water temperature in condenser inlet	Condenser heating rate, COP, energy use, high and low pressure of the working medium, condenser outlet temperature
Dong et al., 2018)	Energy use for DHW production	Residential	Support Vector Machine, Gaussian Naive Bayes, and Random Forest	Ambient temperature, setpoint temperature, initial tank temperature of each node, hot water draw	Current temperature and temperature change at each node, energy use
Delorme-Costil and Bezian, 2017)	Energy use for DHW production	Residential	Feed-Forward neural network	Day number, ambient temperature, lagged energy use	Energy use
Fischer et al., 2016)	Energy use for DHW production	Residential	Stochastic	Occupant type, building type, weather data	Representative profile of energy use
Dmytro et al., 2019)	Energy use for DHW production	Hotel	Stochastic regression models	Number of guests in the current day, number of guests the last day	Energy use
Gelažanskas and Gamage, 2015)	Volumetric demand of DHW	Residential	Neural Network Nonlinear Autoregressive (NAR), Nonlinear Autoregressive Exogenous (NARX)	Day number, the average consumption of the target hour, lagged hot water use data	Volumetric demand
Maltais and Gosselin, 2019)	Volumetric demand of DHW	Residential	Recurrent neural network + Random Forest	Year number, month number, day number, hour, lagged volumetric demand	Volumetric demand
de Santiago et al., 2017)	Volumetric demand of DHW	Residential	Probabilistic approach	Hot water draw	Profile of volumetric demand of DHW

photovoltaic thermal nanofluid-based collector. Particle Swarm Optimization has been used to optimize the parameters of machine learning models. Diez et al., (2019) used an artificial neural network to model the operation of a flat plate solar collector using three different working fluids. The model uses solar irradiance, ambient temperature, inlet temperature, and working fluid flow rate to predict the outlet temperature. Results show that the developed model is accurate, flexible, and generalizable for modeling other flat plate solar collectors. Feng et al., (2020) evaluated the performance of hybrid particle swarm optimization and extreme learning machine compared to 5 other conventional machine learning models for the prediction of solar irradiance. Due to the lack of accurate equations for performance modeling of heat pipe solar collectors, Shafieian et al., (2020) developed machine learning models using experimental data for more accurate modeling of these collectors. Correa-Jullian et al., (2020) implemented three deep learning models to predict the performance of an SHW system under different meteorological conditions. The training data are produced using TRNSYS simulations. Dong et al., (2020) developed a convolutional neural network model to forecast solar irradiance based on meteorological data. Genetic algorithm and Particle Swarm optimization have been used to optimize the hyperparameters. Behera and Nayak, (2020) developed a decomposition-based neural network for short-term power use prediction of PV panels. In their model, a signal decomposition and reconstruction phase is performed to de-noise the input time series. This de-noising significantly improved model performance. Alizamir et al., (2020) evaluated six different machine learning models to forecast solar radiation using climatic parameters as input. Gradient boosting Tree indicated a better performance. To evaluate the potential of a new site for PV panels, Jung et al., (2020) developed a generic LSTM neural network model which was trained by the data of 164 different PV sites. This model can then be used to evaluate the potential of other locations. Considering the high initial cost of Pyranometers for solar energy potential, they are not widely used to provide the data of different locations. Therefore, Kosovic et al., (2020) proposed a neural network model which estimates the solar radiation potential using only meteoroligcal data, which are widely available in different regions.

Despite the increasing share of hot water energy demand in buildings, very few studies have focused on the real-time prediction of hot water energy use. Today there is an increasing interest in the use of hybrid energy systems in which hybrid systems enable cooperation between various types of sources and installations (Dudkiewicz and Fidorów-Kaprawy, 2017; Heidari et al., 2019). It has led to the implementation of renewable energy-assisted hot water systems, which are one of the most cost-effective energy systems in many climates (Fadl and Eames, 2019). However, due to the transient nature of renewable energy sources, these systems are harder to predict and control. Accordingly, the energy use prediction in renewable energy-assisted hot water systems will be helpful for better management of future hot water systems. The present paper aims to develop a novel model for more accurate prediction of energy use by a solar-assisted heat pump water heating system. The following steps are performed to achieve this aim:

1. **Developing an LSTM neural network model:** Due to the highly stochastic nature of this problem, a long sequence of input data is required to achieve acceptable accuracy. Therefore, LSTM neural network is chosen as the basis of the model, due to its capability of dealing with long term dependency in time series data.
2. **Implementing the attention mechanism on the LSTM model:** The Attention mechanism enables the LSTM model to pay selective attention to the input time-series data, and focus on the data which carry more information. It is a recent advancement in deep neural networks (Yu et al., 2019), which is used more in speech recognition and machine translation applications (Cho et al., 2014; Lipton et al., 2015; Sutskever et al., 2014). However, its application in energy prediction has been largely unexplored (Rahman et al., 2018).
3. **Decomposing the input time series data:** Decomposition of input time series into sub-layers allows the model to learn each sub-layer separately. The decomposition of input data using signal transformation methods has shown high performance in recent studies on time series prediction such as prediction of PV panels power production (Li et al., 2020). This study uses Naive decomposition method to break down the time-series data to further improve the LSTM model using the attention mechanism.

Even though the accuracy of deep learning models has been improved in recent years, most improvements have been achieved by introducing more complex topologies. Nonetheless, it is possible to further improve the LSTM model by independently focusing on the sub-layers of the input time series, and also paying selective attention to the input data. This paper advances the current knowledge by introducing a hybrid model that combines the attention mechanism and time series decomposition with LSTM neural networks to obtain a higher accuracy for the prediction of long-term time-series data. The superiority of this model over conventional models is highlighted and potential further

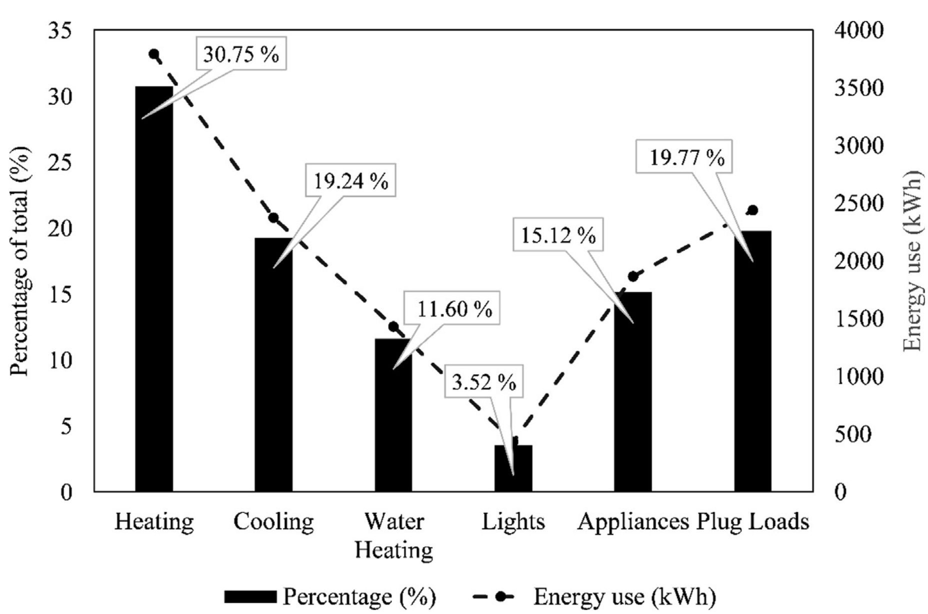


Fig. 1. Annual energy demand by different end uses in case study building (Data from Lipton et al., 2015).

developments are suggested.

The remainder of the paper is organized as follows/into 3 sections: section II describes methodology, including a description of case study building, an overview of neural network models, attention mechanism and time series decomposition. section III discusses the results, and section IV concludes the paper.

2. Methodology

2.1. Case study description

This study uses the data of a 250.8m^2residential building in Gaithersburg, Maryland. This building has been monitored by the US Department of Energy to demonstrate the feasibility of achieving net Zero Energy consumption during a year. The building was not occupied, rather a virtual family of four was simulated to determine energy use throughout a year. A specific methodology, based on recommendations of the U.S. Department of Energy, has been used to replicate occupancy profiles, hot water usage, lighting usage, electric plug loads, cooking loads, appliances loads, and sensible and latent loads. Detailed information about the occupancy simulation can be found in (Omar et al., 2013). This building is monitored for research purposes, providing a high resolution (minutely and hourly) data set of the building operation over two years (Healy et al., 2018). Many research papers used the data set to address various aspects of this building. Studies that have focused on the water heating system in this building include the evaluation of efficient water heating strategies for an all-electric building (Balke et al., 2016), year-long performance analysis of the heat pump water heater (Ullah and Healy, 2016), and evaluation of heat losses of the hot water distribution system (Ullah and Healy, 2015). Different from the previous papers, the current study focuses on developing an energy use prediction model using real data from this case study building.

Actual energy use by different end uses in the case study building is represented in Fig. 1. Hot water energy use accounts for 11% of the actual energy use in this building. Hot water in this building was provided by a solar assisted heat pump. This hot water system was monitored for a period of one year. Fig. 2 represents the layout of this system, together with the location of temperature and flow sensors installed. Specifications of the water heating system are listed in Table 2. Average daily hot water consumption for showers, baths, and

Table 2
Specifications of solar-assisted water heating system (Data from Lipton et al., 2015).

Parameter	Value
Solar System	
Collector type	single-glazed flat plate
Number of solar panels in operation	2
Collector dimensions	1.1 m × 2 m
Collector orientation	South
Collector tilt angle	18.4°
Heat transfer fluid	50% (by volume) propylene glycol in water solution
Auxiliary Heat Pump	
Coefficient Of Performance (COP) (Thermal energy delivered to the electrical energy use)	2.6
Storage volume	189 L
Setpoint temperature	49 °C
Standby losses	0.2 °C/h
Electric element backup	4.5 Kw

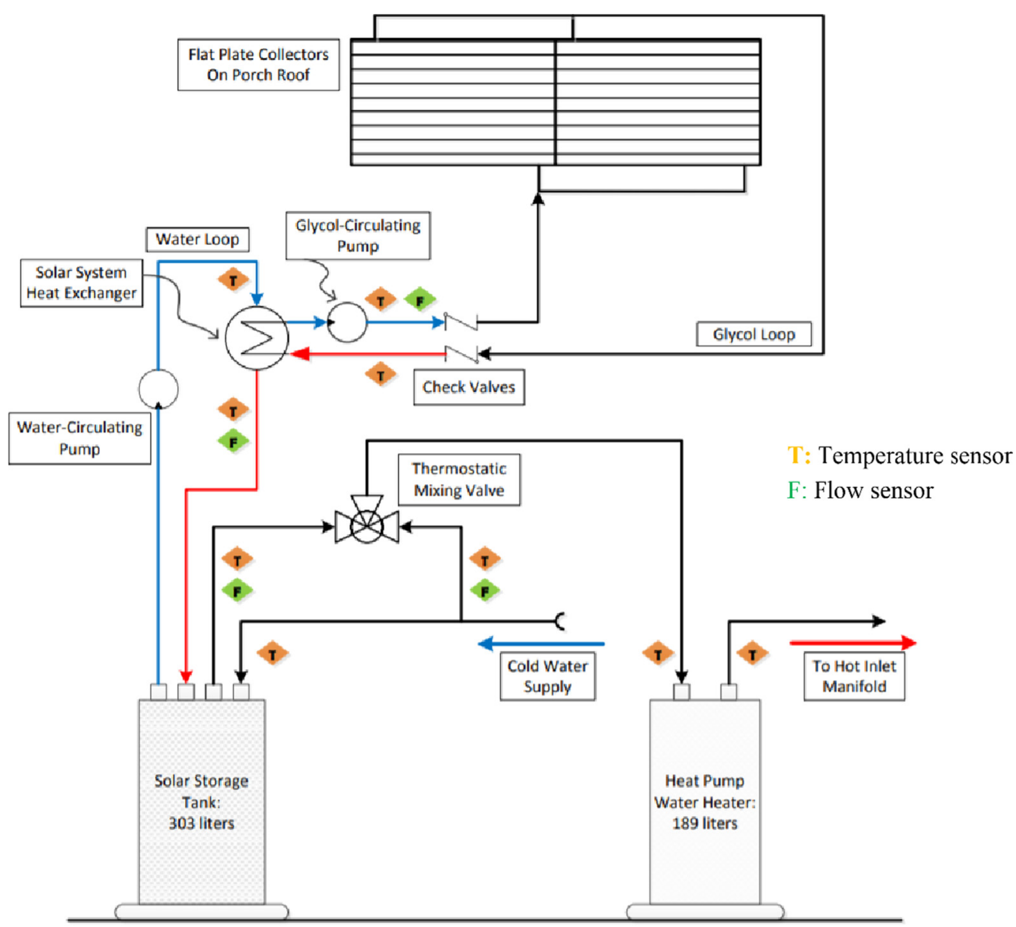


Fig. 2. Schematics of solar-assisted heat pump water heating system in case study building with measurement points (Edited from Domich et al., 2015).

sinks was 123.7 L/day, 31 L/day and 110.3 L/day, respectively. The total energy of 2894 kWh was needed for domestic hot water production during the first year, from which 54% was provided by the solar collectors and the remaining by the heat pump system. To deliver this energy, the heat pump system consumed a total of 1112 kWh including 137 kWh for back-up electric heating in the storage tank.

2.2. Prediction model development

The development of prediction model is performed in three main steps to quantify the improvement achieved by each step. First a simple LSTM neural network is developed, then attention mechansn is added to the LSTM neural network, and finally decomposition of input sequence is performed. Following each of the three steps are presented.

2.3. Step 1: Development of LSTM neural network

In a Recurrent neural network (RNN), there is a limited memory of the context of input data. This is mainly because of the influence of a given input, either decay or explodes as it cycles around the network's recurrent connections (NIST, 2019). This problem is often referred to as the Vanishing Gradient Problem. LSTM neural networks are developed to solve the problem of Vanishing Gradients. In an LSTM neural network, the summation units of RNNs are replaced by the memory blocks. The memory blocks of LSTM work based on gating the information, which enables the network to store and access information over longer periods. These gates (including input gate, forget gate and output gate) determine either the LSTM unit should maintain, update or delete the cell state information. Fig. 3 (a) represents the schematics of LSTM neural network in this study. The input to the model is a sequence of previous energy use data, and the output is either the sequence or single value of future energy use.

2.4. Step 2: Implementation of attention mechanism

Although the output of an LSTM unit is a function of all previous time steps, it might not be able to effectively capture information in long term inputs due to its limited memory (Vinayavekhin et al., 2018). Attention mechanism can further improve the performance of LSTM neural networks in dealing with long term input data (Zhou et al., 2019). Attention is a mechanism inspired by the biological vision system, which allows us to focus on specific objects in observations (Zhou et al., 2019). It has been mostly applied to machine translation (Choi et al., 2018), video analysis (Li et al., 2018), and image analysis (Song et al., 2018), while its application in energy forecasting has been overlooked so far.

The Attention mechanism allows a neural network to focus on the input data which are more important to the current output. Fig. 3 (b) shows the schematics of Attention-based LSTM neural network in this study. The Attention mechanism in this network is implemented between two layers of the LSTM network to apply selective importance to the input data. There are different approaches to implementing an attention mechanism. In this research, the Bahdanau algorithm (Bahdanau et al., 2014) is used, which is one of the most commonly used algorithms. The selective importance in this model is applied by calculating a vector C_t, known as a semantic vector, at every time step. To calculate this vector, first e_t is calculated at each time step based on the hidden state h_t of the encoding LSTM as equation (1)

$$e_{ti} = \tanh(W_a[s_{t-1}, h_i]),\ e_i \in [-1, 1] \tag{1}$$

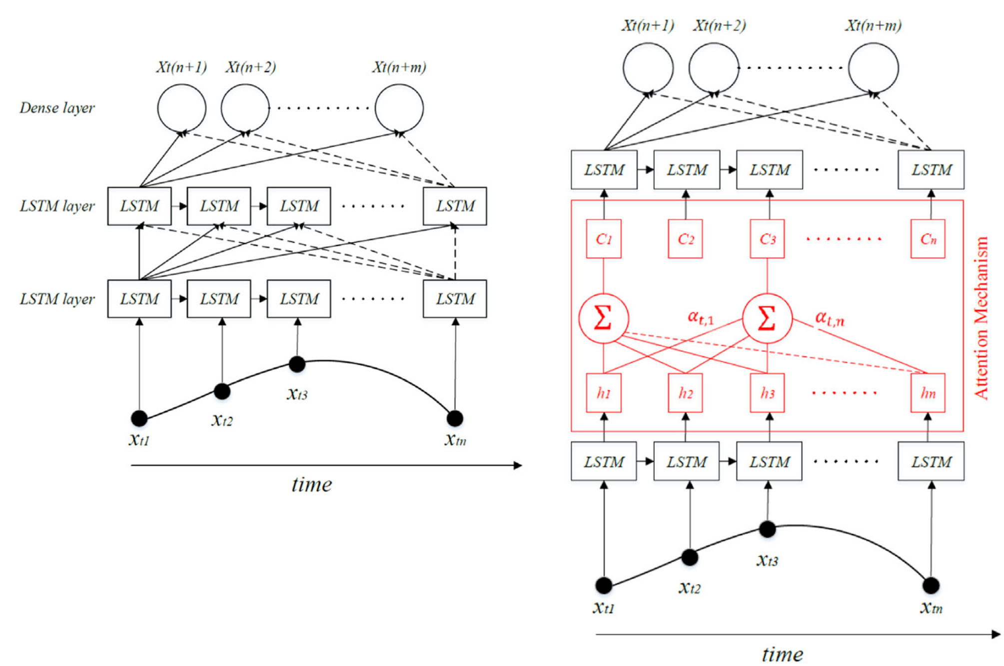
In which s_{t-1} is the hidden state of the LSTM unit at the next layer at one time step before, and W_a is a weight matrix, adjusted during the training process. In fact, e_{ti} is an alignment model score that describes the matching between an input at position i and an output at position t. Then, this score is normalized using a Softmax function as follow:

$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{k=1}^{T}\exp(e_{tk})} \tag{2}$$

This normalized score is then used to form the semantic vector C_t as follow:

$$C_t = \sum_{t=1}^{T}\alpha_{ti}h_i \tag{3}$$

This semantic vector is then used to calculate the hidden state of the



(a) (b)

Fig. 3. Layout of a simple LSTM neural network (a) and an Attention-based neural network (b).

next layer as follow:

$$s_t = \tanh(W_b[s_{t-1}, y_{t-1}, C_t]) \quad (4)$$

In which y_{t-1} is the unit output at one time step before.

2.5. Step 3: Decomposition of input time series

A given time series can be assumed as the summation of different sub-layers. Therefore, the original time series can e decomposed into sub-layers, and each sublayer can be predicted independently. It can also help to remove noise from the original data (Behera and Nayak, 2020). Different methods can be used for decomposition, and the type of sub-layers are therefore different. For example, a wavelet packet decomposition (Li et al., 2020) decomposes the original time series into sub-layers with different frequencies. On the other hand, a Naïve decomposition method decomposes the time series into trend, seasonality and residual sub-layers, in which the summation of these layers will form the original series as below:

$$\text{Original time series} = \text{Trend} + \text{Seasonality} + \text{Residual} \quad (5)$$

In this study, the Naïve decomposition method is used to decompose the original series, and then for each sub-layer, a separate LSTM model using attention mechanism is trained and used to predict that sub-layer. The results of predictions are then summed up to form the main predictions. The main predictions are then compared to the test set to calculate the accuracy of the entire model. Fig. 4 show the schematics of this approach.

3. Results and discussion

The energy use of solar-assisted heat pump water heaters is influenced by several stochastic parameters. These parameters include the profile of hot water usage, outdoor air temperature, which affects the heat pump COP, availability and intensity of solar radiation which affects the temperature rise of water flowing through the collectors and also energy use of solar system pump, and wind speed, which affect the heat loss of collectors. As shown in Fig. 5, these influencing parameters are highly varying throughout the year. Indoor air temperature and relative humidity are also considered to evaluate their potential influence on the performance of the model. The data used was from February 2015 to February 2016, providing the hourly data for one year. As could be seen, there was one week of missing data in October. Also, at the beginning of January 2016, the energy use was extremely low and did not follow the same pattern as other weeks. Although this abnormal pattern would increase the prediction error, it was not excluded from the data set as it would be interesting to evaluate the model performance during an unexpected period. The total energy use of the system can be divided into the energy use by the auxiliary heat pump, and the energy use of the solar system pump. The hourly energy use of the solar system pump has a direct relation with solar energy, as it will be working only when there is a sufficient amount of solar radiation. The heat pump energy use, on its turn, has an opposite relation with the solar radiation, and direct relation with the outdoor air temperature. This is because the higher solar radiation will result in the higher heat gain by solar collectors and, therefore, lower energy required to operate an auxiliary heat pump for pumping water at a lower flow rate. The direct relation with outdoor air temperature is because the higher outdoor air temperature increases the COP of the heat pump, and therefore reduces the required power. Wind speed is directly affecting the magnitude of heat losses of the solar system and, therefore, the heat pump energy use, which is however very small in the case of the solar collectors. A summary of the dataset is presented in Table 3.

The next step in the primary analysis of data is autocorrelation analysis. Autocorrelation, also known as serial correlation, is the correlation of the data set with a lagged version of itself. Autocorrelation analysis is a useful method to identify periodic patterns in data (Al-Karawi and Mohammed, 2019). Pearson autocorrelation function is used for the energy use data. As shown in Fig. 6, there is a peak on the 168 h lag (with an autocorrelation coefficient of 0.7), which means that the pattern of energy use is the most repetitive with a one-week interval. This is mainly because the hot water use profile in residential buildings usually follows a weekly pattern due to the routines of water use (Delorme-Costil and Bezian, 2017). There is another peak on 24 h lag (with an autocorrelation coefficient of 0.55), which shows that the energy use profile also shows a similarity between the days. The first reason is that the hot water use profile in dwellings usually follows a similar pattern between the weekdays and between the weekend days, with one peak in the morning and another in the evening (Gelažanskas and Gamage, 2015). The other driver is the similarity of solar radiation patterns between the days.

In addition to the lagged data of energy use, other influencing variables shown in Fig. 5 can also be used as the input features. These variables can be independent or correlated to each other, and the strength of their relationship can be measured by the correlation coefficient (Ji et al., 2019). Evans, (1996), in his study on power use forecasting, classified the absolute value for correlation factor as very weak (0.00 – 0.19), weak (0.20 – 0.39), moderate (0.40 –0.59), strong (0.60 – 0.79) and very strong (0.80 – 0.10). Pearson correlation analysis is therefore used to measure the dependencies between the available variables. Temporal variables of an *hour* and *day numbers* are also included as the potential features. The heat map of the correlation factors is presented in Fig. 7. As shown in the figure, the energy use data is highly correlated with the indoor air temperature, followed by solar

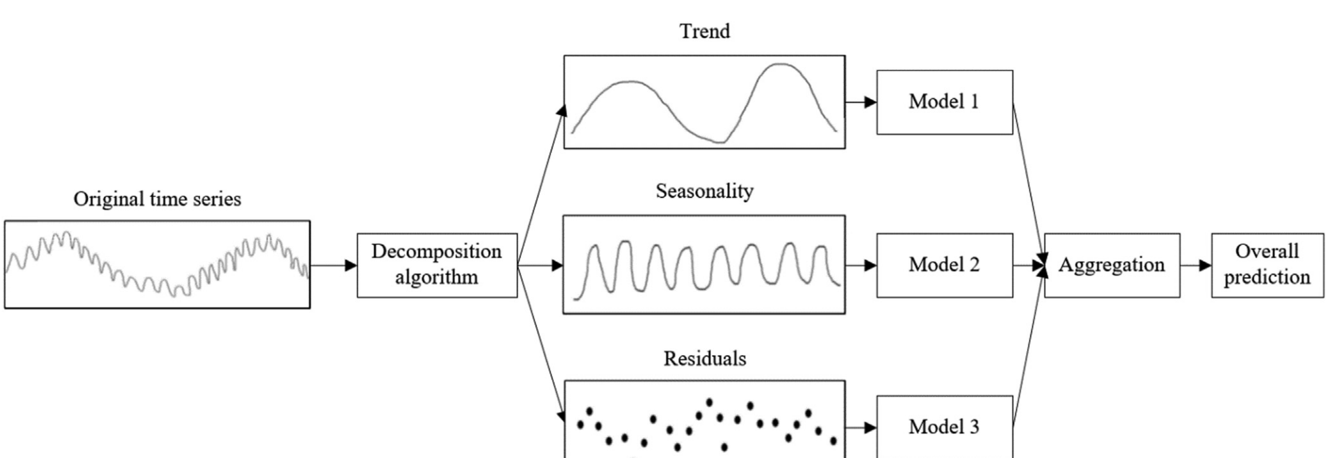


Fig. 4. Naive decomposition method for LSTM neural network using the attention mechanism.

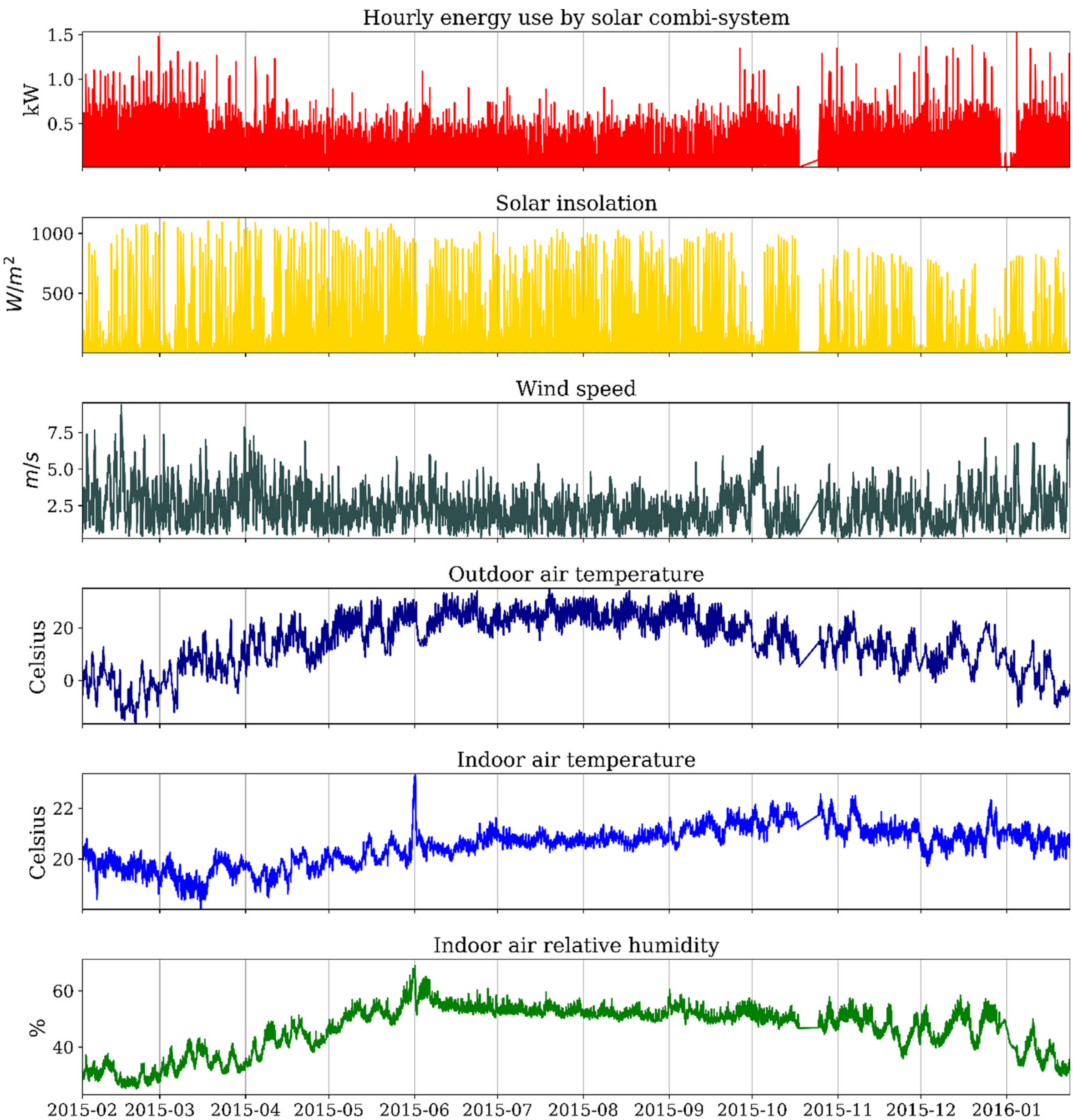


Fig. 5. Variations of different influencing parameters during one year.

Table 3
Summary of the dataset.

	Energy use (kW)	Rh_{in} (%)	T_{in} (°C)	T_{out} (°C)	Wind speed (m/s)	Insolation (W/m^2)
Count	8400	8400	8400	8400	8400	8400
Mean	0.16	0.46	20.57	14.53	2.32	179.55
Standard deviation	0.22	0.08	0.75	10.40	1.32	274.61
Minimum	0	0.25	18.02	−16.55	0.26	4.95
Percentile 25%	0.01	0.40	20.05	7.33	1.31	13.28
Percentile 50%	0.04	0.49	20.68	15.75	2.09	18.12
Percentile 75%	0.21	0.52	21.08	22.83	3.03	230.70
Maximum	1.53	0.69	23.36	34.99	9.55	1134.31

radiation, hour, and outdoor air temperature. However, all of them could be classified as having a very week correlation with energy use because the correlation factor is below 0.19. The existing correlation of energy use data with Hour and Day number shows that there are some repetitive patterns in hot water energy use. This seasonality is due to the daily routine in solar radiation as well as hot water demand. The existing routine in the daily behavior of occupants results in the correlation of temporal factors with hot water demand and, therefore, with hot water energy use.

Model development is performed step by step to quantify the improvement by each step. The first step is developing an LSTM neural network by investigating different topologies. The input features in this step are the lagged data of energy use time series. The sliding window method was used for model development. In this method, a fixed size

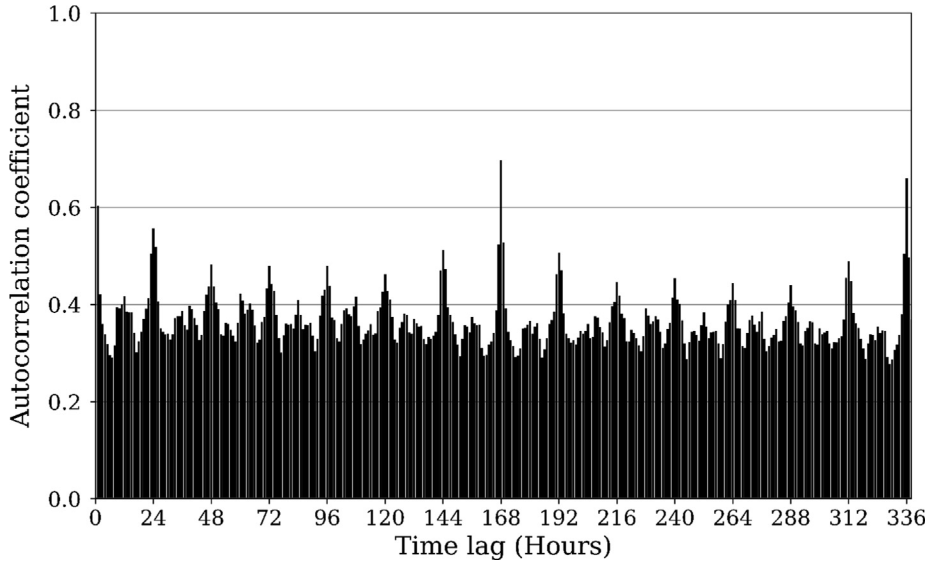


Fig. 6. Autocorrelation analysis of the energy use data.

window slides over the time series data as illustrated in Fig. 8. The values inside the window are the input vector of training, and the value next to the window is the target vector to be predicted. The target vector could be a single time step or multiple time steps in the future. Due to the highly stochastic nature of the energy use in this study, the input features should include enough data to provide sufficient information to the model. According to the autocorrelation results, the energy use data at time step (t) has the highest correlation with the energy use at the time step ($t - 168$). Therefore, in this step, the sliding window includes the last 168 h of energy use data and the target value is the next time step ($t + 1$).

Six different topologies have been evaluated for the LSTM neural network as represented in Table 4. An equal number of neurons was used in two LSTM layers for each topology for easier comparison between different topologies. The results show that including more layers and neurons does not necessarily improve the accuracy of the model. Topology 6 with 2 layers of 150 LSTM neurons shows the best performance between the analyzed topologies, therefore, this topology is considered as the base LSTM model for implementing the additional improvements.

The input features can be chosen based on the correlation factors or feature engineering. Alternatively, different combinations of features can be evaluated to quantify the model performance with different features. Therefore, the model was trained and tested with different combinations of features listed in Table 5. As can be seen from the values of MSE, RMSE, and MAE, the inclusion of additional features does not increase the accuracy of the model. It can be concluded that the required information is already included in the lagged energy use

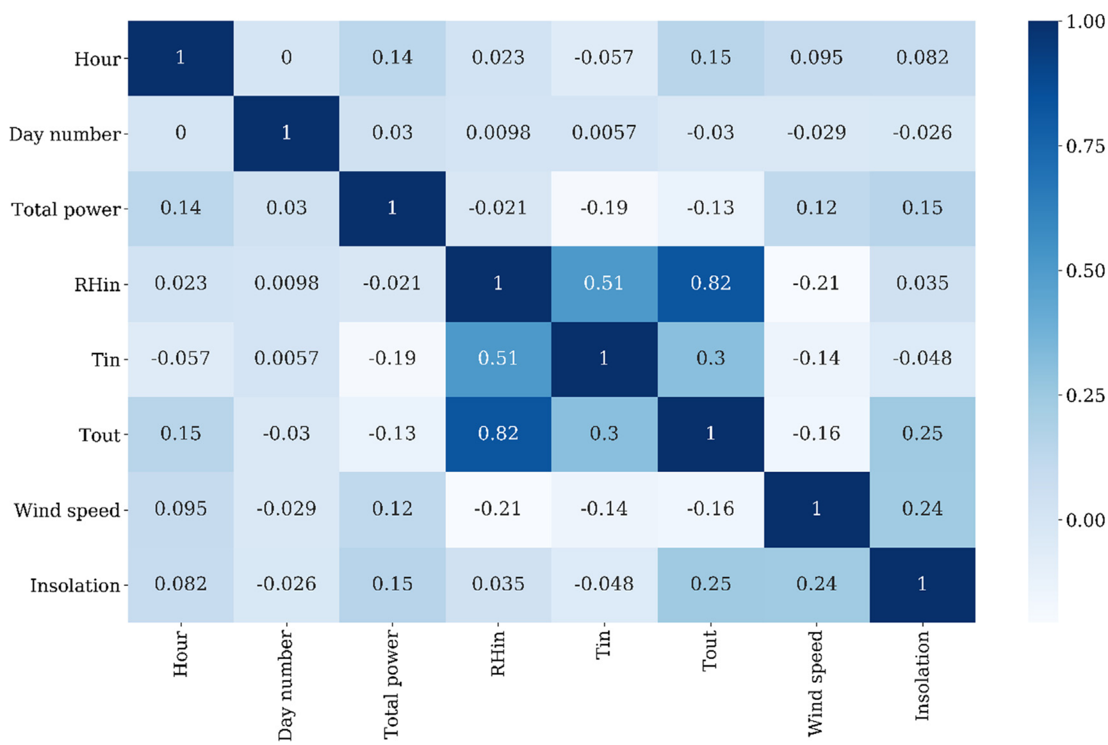


Fig. 7. Correlation analysis between the available features.

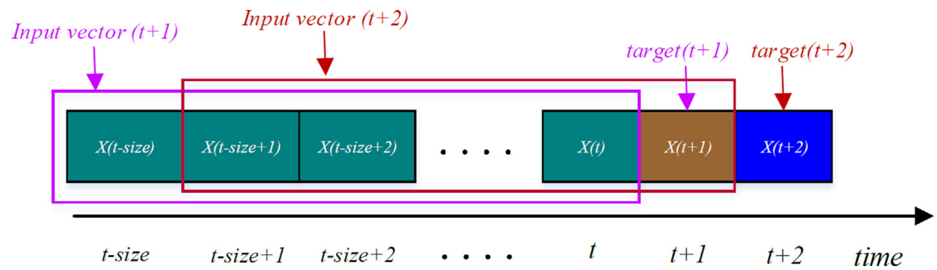


Fig. 8. Sliding windows method for prediction of the next time step(s).

data, and additional features provide unnecessary information which prevents the generalization capacity of the neural network. This finding is in line with the work by (Delorme-Costil and Bezian, 2017), in which the model using only historical data of hot water use indicated the highest accuracy in prediction of the future hot water demand.

After the evaluation of topologies and features, the model with topology 6 (2 LSTM layers of 150 neurons) was selected as the base LSTM model, which, however, shows a moderate accuracy. Therefore, this model should be further improved. The model improvement could be done by increasing the number of lagged energy use data, which provide more information to the model, and also including more layers of neurons. To this aim, the LSTM model is first improved by including the lagged energy use data of 4 preceding weeks, and also increasing the number of neurons in each layer to 800. This model is then improved in two steps. In the first step, an attention mechanism is implemented between the two LSTM layers. Considering the long sequence of input data, this attention mechanism is used to improve the model's ability to extract useful information from the input sequence. The second step is to implement the decomposition method to the attention-based LSTM model. Fig. 9 shows the decomposition of time series into sub-layers. As shown in this figure, once the trend and seasonality sub-layers are extracted they are easier to follow, and therefore easier to be predicted. The irregular pattern of residuals also separated from the main data, and therefore an independent model can better focus on this pattern. A separate attention-based LSTM model with the same topology as the last step is used for each sub-layer, and the predictions of these models are finally summed up to form the final predictions. Finally, a Feed-Forward neural network is also developed with the same input features and the same topology to be used as the benchmark model. Table 6 summarizes the performance of these models and the improvements compared to the baseline model.

Acording to the results, the performance of the LSTM model is significantly better than the baseline model, with 25% lower MSE and 25.4% lower MAE. Implementing the Attention mechanism into the LSTM model further improves the model performance, with 28.5% lower MSE and 27.86% lower MAE to the base case. Interestingly, decomposition of input time series for attention-based LSTM model significantly improved the performance, with a 65% reduction in MSE and 41% reduction in MAE to the baseline model. The performance of this hybrid model is also significantly better than the LSTM model, with 40% lower MSE and 21% lower MAE. As mentioned earlier, the power use of a solar-assisted water heating system is mainly affected by two stochastic phenomena, solar irradiance, and hot water demand. Both of these phenomena include a significant seasonality and trend. The effects of this seasonality and trend are added together in power use time series. The hybrid model shows promising results as it learns the seasonality and trend separately, and have a good memory to learn long term relationship in the input sequence. Accordingly, it is expected that this hybrid model can be used for other types of solar-assisted hot water systems while outperforming the conventional models.

In order to represent how the energy use varies over time, and how the predictions follow the real energy use, the predicted and measured energy use data over the whole test period are presented in Fig. 10. For a better resolution, the first week of predictions is also shown in Fig. 11. The energy use in a hybrid hot water system happens in discrete time intervals; there are some instances of zero energy use and some instances of energy use, which also highly varies in amount. It makes it harder for the model to follow the variations compared to the systems in which the energy use is continuous. As could be seen, all the models usually well predict the time of power use, however, they slightly underestimate the amount of energy use. This underestimation of magnitude can be managed by including the uncertainty of predictions, depending on the application. The time and magnitude of future power use can then be used by smart buildings, energy prosumers, smart grid operation, and demand-side management.

Due to the discrete nature of power use in the solar-assisted system, it is hard to visually distinguish the best model, as in some cases the base model performs better while in some other the improved models do.

Fig. 12 shows the predicted and measured energy use during one week of abnormal energy use in the test period. While the inclusion of this abnormal week increases the overall error, it would be interesting to see the performance of different models as such abnormal patterns or unexpected shutdown of the system is very probable in practice. As shown, when there is no energy use, the models are still predicting some energy use. There are some small energy use instances in-between, which prevent the models to continue this decrement in predicted power and be adjusted to zero. In the case of these small power use instances, the models usually overestimate the power use, different from the normal patterns where the models underestimate it. In these small power use instances, the amount of power use predicted by the decomposition-based model is usually lower and closer to the actual

Table 4
Evaluation of different topologies of LSTM neural network using last week of energy use data to predict the next time step.

Topology	Number of LSTM layers	Number of Neurons in each layer	Number of Epochs	Prediction MSE (kW)	Prediction RMSE (kW)	Reduction of prediction MSE to the base case (%)	Prediction MAE (kW)	Reduction of prediction MAE to the base case (%)
1	1	50	100	0.040	0.20	base	0.112	base
2	2	50	100	0.034	0.18	15.0	0.125	−11.6
3	1	100	100	0.039	0.20	2.5	0.125	−11.6
4	2	100	100	0.038	0.19	5.0	0.116	−3.6
5	1	150	100	0.056	0.24	−40.0	0.150	–33.9
6	**2**	**150**	**100**	**0.025**	**0.16**	**37.5**	**0.100**	**10.7**

Table 5
Model performance with different combinations of features.

Scenario	Features								MSE (kW)	RMSE (kW)	MAE (kW)
	Sequence of previous power use	Hour	Day Number	T_{in}	Insolation	T_{amb}	RH_{in}	Wind speed			
Baseline	✓								**0.025**	**0.16**	**0.100**
1	✓	✓							0.029	0.17	0.111
2	✓	✓	✓						0.028	0.17	0.110
3	✓	✓	✓	✓					0.031	0.18	0.112
4	✓	✓	✓	✓	✓				0.032	0.18	0.113
5	✓	✓	✓	✓	✓	✓			0.031	0.18	0.111
6	✓	✓	✓	✓	✓	✓	✓		0.030	0.17	0.113
7	✓	✓	✓	✓	✓	✓	✓	✓	0.030	0.17	0.112
8	✓		✓						0.033	0.18	0.114
9	✓			✓					0.029	0.17	0.108
10	✓				✓				0.032	0.18	0.116
11	✓					✓			0.032	0.18	0.117
12	✓						✓		0.030	0.17	0.112
13	✓							✓	0.033	0.18	0.117

power use. The Feed-Forward model also shows a good performance in this abnormal period. These models only use the last power to use data for predictions. However, in case of implementing these models in smart buildings, their performance can be further enhanced by including the data from occupancy sensors, because the abnormal patterns are usually the result of unoccupancy.

The box plot of hourly absolute error (absolute difference of predicted and real value in each hour) of different models are presented in Fig. 13. As could be seen, the median of error by decomposed attention-based LSTM, attention-based LSTM and LSTM are similar and lower than Feed-Forward model. However, the maximum value of absolute error by decomposed attention-based LSTM is lower than all the other models (8.2% lower than Attention-based LSTM, 4.8% lower than LSTM, and 17.8% lower than Feed-Forward). As the maximum value of error is expected to happen during the abnormal week, it can be concluded that the decomposed attention-based LSTM can better adapt to the abnormal conditions and therefore is a more robust model.

Based on the results, it can be concluded that long term sequence data are needed for accurate energy use prediction in solar-assisted hot water systems. Accordingly, the decomposed LSTM model using the attention mechanism proposed in this study can be successfully used to deal with such long-term data and predict the energy use of solar-assisted water heating systems in different locations with a better performance of conventional models. The topology of the model is dependent on the dataset and therefore for other datasets, the number of layers and neurons should be tunned based on the data. The focus of this research is on the comparative study of this novel model to the conventional models in the prediction of energy use. Therefore, more complicated topologies or higher number of inputs are avoided in the current research to reduce the computational power needed. However, further improvement of the model accuracy can be achieved by including other features like occupancy, optimizing the model

Fig. 9. Decomposition of input time series.

Table 6
Performance of enhanced models compared to the baseline model.

Model	Number of LSTM layers	Number of neurons in each layer	Input features	Prediction MSE (kW)	Prediction RMSE (kW)	Reduction of prediction MSE to the base case (%)	Prediction MAE (kW)	Reduction of prediction MAE to the base case (%)
Feed-Forward	2	800	4 weeks of lagged energy use	0.028	0.17	Baseline	0.122	Baseline
Simple LSTM	2	800	4 weeks of lagged energy use	0.021	0.14	25	0.091	25.40
Attention-based LSTM	2	800	4 weeks of lagged energy use	0.020	0.14	28.5	0.088	27.86
Attention-based LSTM using decomposed data	2 (for each model)	800 (for each model)	4 weeks of lagged energy use	0.0098	0.10	65	0.072	41

architecture, or simplifying the prediction output, such as predicting the total power of the next day.

A limitation of the current study is that the data are associated with a case study building with a simulated family. Clearly, the inclusion of real occupants increases the stochastic nature of data. Due to the difficulty of collecting enough data, in this research, the available data has been used for comparative evaluation. According to the performed literature review, very few papers have focused on the application of the attention mechanism and time series decomposition in the energy field. Further studies will need to be undertaken to reveal the potential applications of this method in energy systems. We propose that further research can be undertaken in the following areas:

- Combination of the attention mechanism and uncertainty-aware models like Bayesian neural network;
- Comparative evaluation of different attention mechanisms for energy use prediction;
- Comparative study on different decomposition models for energy use prediction;

To evaluate the generalization ability of the models in this study it would be worthful to implement them on other datasets of water heating systems. The authors have developed the models on Python Notebook and can share them upon request.

4. Conclusion

Regarding the increasing share of hot water energy use in modern buildings, the early prediction of this energy use can be useful for various applications such as power use operation or demand-side management. Previous studies have shown that it is very challenging to predict the energy use of water heating systems. It is even more challenging in the case of renewable energy-assisted systems, as their operation is affected by stochastic environmental factors. This research proposed the application of a novel model combining the attention mechanism, a recent development in deep learning which is mostly used in machine translation, and decomposition of time series, with the LSTM neural network. One year data of a solar-assisted heat pump water heater is used to develop the models. The development of the model was done step by step, starting with a simple LSTM model by evaluating different topologies and features. It was found that the model with two layers of 150 neurons using only the lagged energy use data performs the best, with an MSE and MAE of 0.025 kW and 0.100 kW. To further improve the model performance, the number of lagged energy use data increased to 4 weeks, and the number of LSTM neurons in each layer also increased to 800. Then this base model was enhanced by implementing the attention mechanism and decomposition of input time series in separate steps. Finally, a Feed-Forward model was implemented with the same topology to be used as the baseline model. It was found that, concerning the Feed-Forward model (MSE = 0.028 kW, MAE = 0.122 kW), the simple LSTM model and Attention-based LSTM model show 25% and 28.5% higher MSE, and also 25.4% and 27.86% higher MAE, respectively. Finally, the ultimate model which combined the attention mechanism and decomposition of input time series significantly improve the model performance, with 41% lower MAE and 65% lower MSE with regard to the baseline model. It can be concluded that long-term input data is needed for the prediction of highly stochastic hot water energy use, and the hybrid model of this study can deal with the long term input data better than the conventional models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to

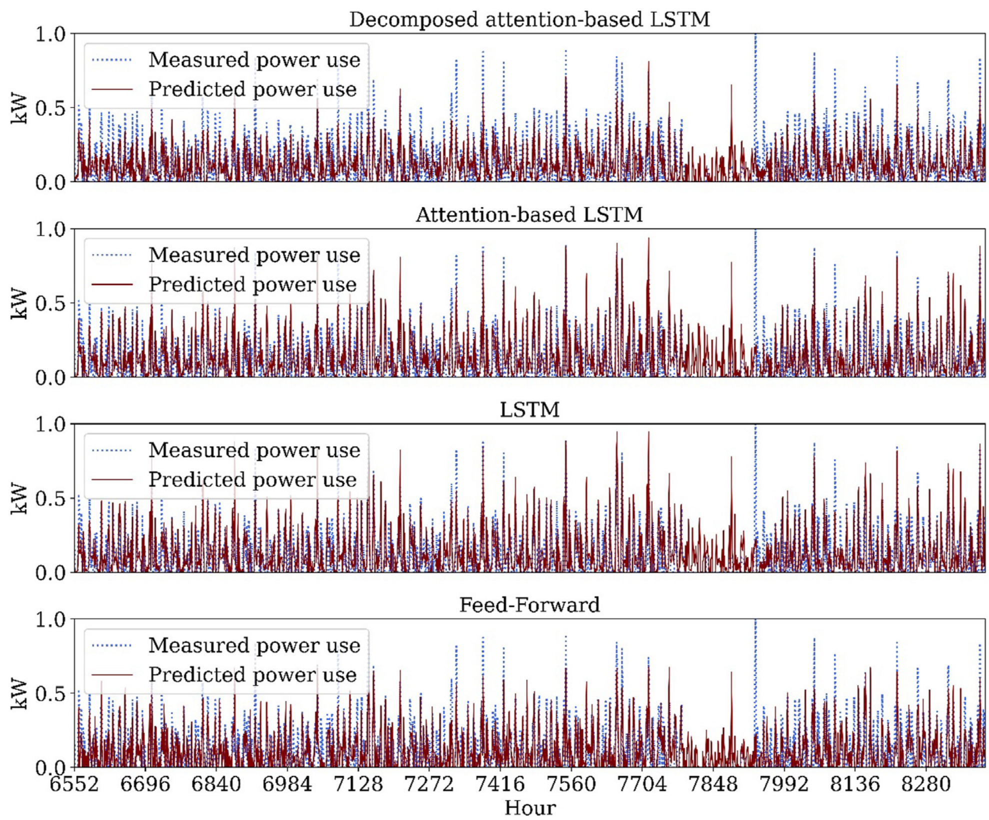


Fig. 10. Predicted energy use versus measured energy use by different models over the test period.

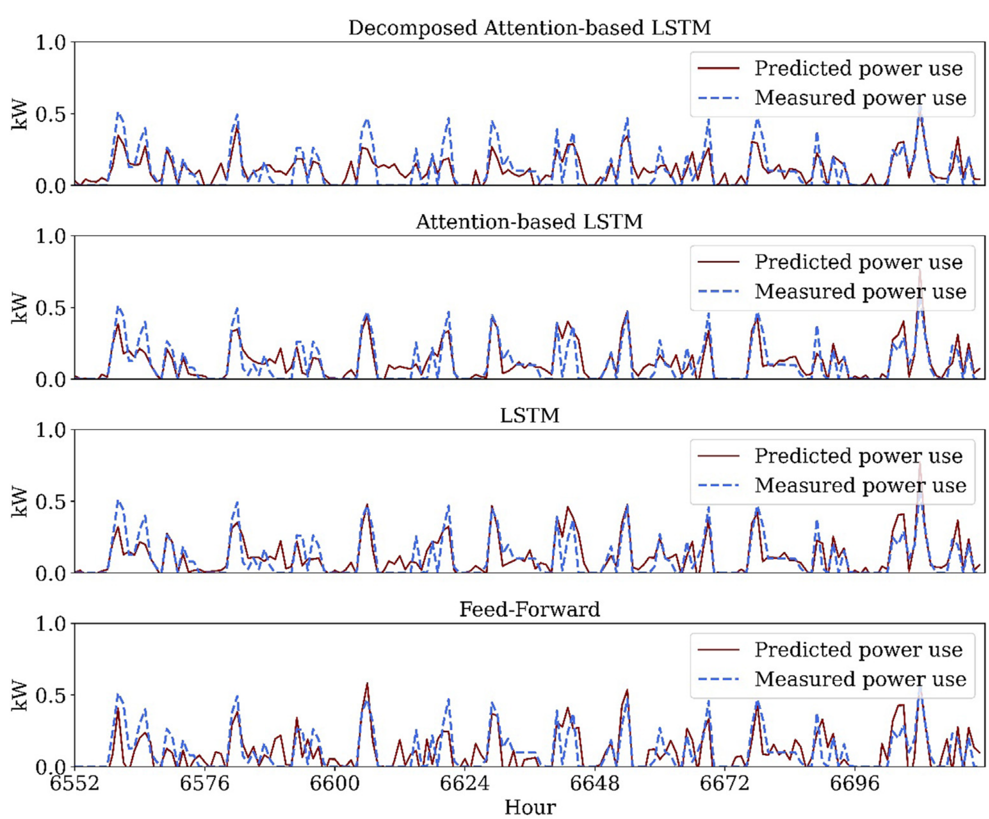


Fig. 11. Predicted energy use versus measured energy use by different models for one week in the test period.

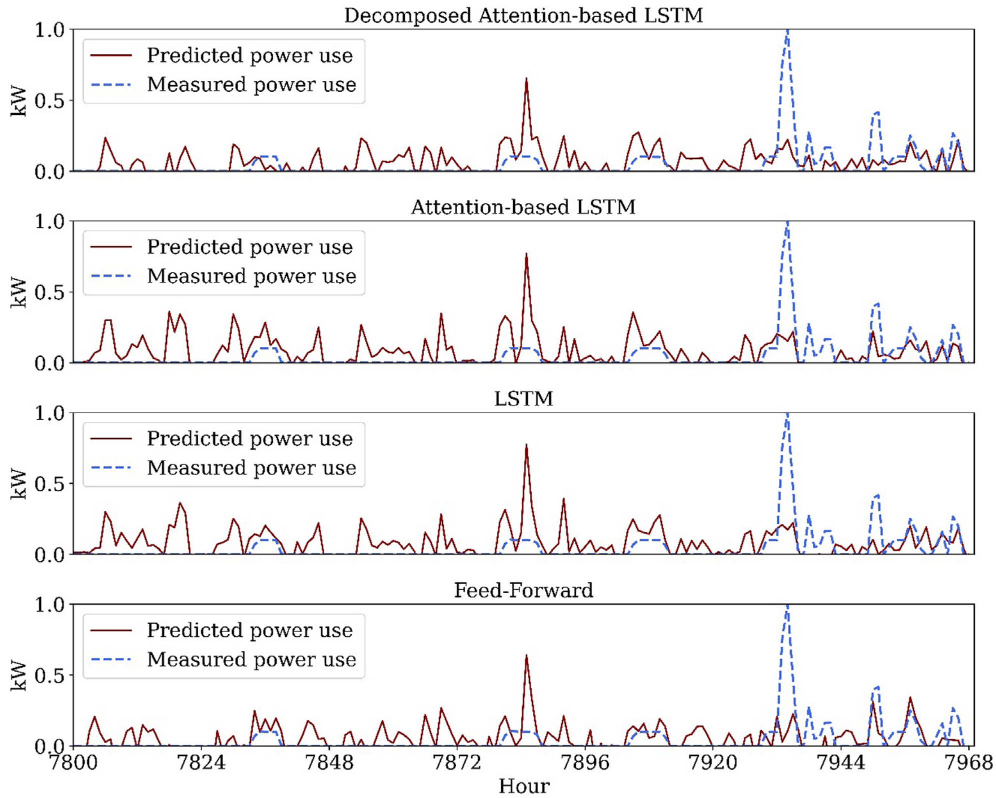


Fig. 12. Predicted energy use versus measured energy use by different models for one abnormal week in the test period.

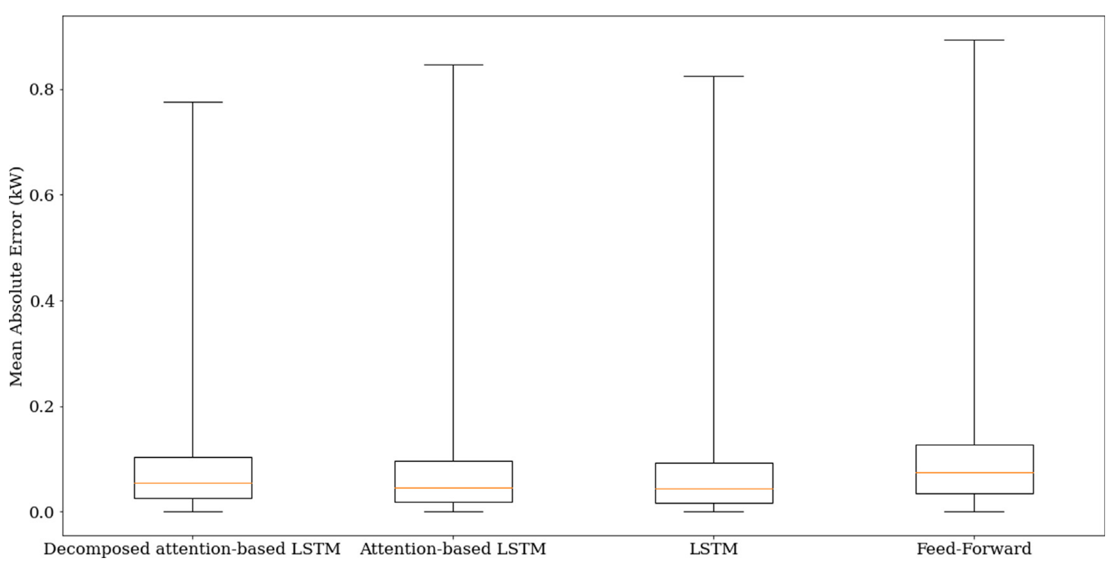


Fig. 13. Hourly variations of absolute error over the test data.

influence the work reported in this paper.

Acknowledgements

We acknowledge the National Institute of Standards and Technology, U.S. Department of Commerce, for sharing the data set of Net-Zero Energy Residential Test Facility.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.solener.2020.07.008.

References

Kazmi, H., Mehmood, F., Lodeweyckx, S., Driesen, J., 2018. Gigawatt-hour scale savings on a budget of zero: Deep reinforcement learning based optimal control of hot water systems. Energy 144, 159–168.

Marszal-Pomianowska, A., Zhang, C., Pomianowski, M., Heiselberg, P., Gram-Hanssen, K., Hansen, A.R., 2019. Simple methodology to estimate the mean hourly and the daily profiles of domestic hot water demand from hourly total heating readings. Energy Build. 184, 53–64.

Booysen, M., Engelbrecht, J., Ritchie, M., Apperley, M., Cloete, A., 2019. How much energy can optimal control of domestic water heating save? Energy Sustain. Develop. 51, 73–85.

Muzaffar, S., Afshari, A., 2019. Short-Term Load Forecasts Using LSTM Networks. Energy Procedia 158, 2922–2927.

Kato, T., Suzuoki, Y., 2014. Autonomous scheduling of heat-pump water heaters for

mitigating voltage rise caused by photovoltaic power generation systems. Appl. Therm. Eng. 71 (2), 652–657.
Mathioulakis, E., Panaras, G., Belessiotis, V., 2018. Artificial neural networks for the performance prediction of heat pump hot water heaters. Int. J. Sustain. Energ. 37 (2), 173–192.
Dong, J., Munk, J., Cui, B., Boudreaus, P.R., Kuruganti, T., 2018. Machine-Learning Model of Electric Water Heater for Electricity Consumption Prediction.
Delorme-Costil, A., Bezian, J.-J., 2017. Forecasting domestic hot water demand in residential house using artificial neural networks. IEEE, pp. 467–472.
Fischer, D., Wolf, T., Scherer, J., Wille-Haussmann, B., 2016. A stochastic bottom-up model for space heating and domestic hot water load profiles for German households. Energy Build. 124, 120–128.
Dmytro, I., Natasa, N., Lekang, S.Å., Igor, S., Wester, P.T.S., Walnum, H.T., 2019. Prediction of DHW energy use in a hotel in Norway. IOP Publishing, pp. 052018.
Gelažanskas, L., Gamage, K.A., 2015. Forecasting hot water consumption in dwellings using artificial neural networks, IEEE, pp. 410–415.
Maltais, L.-G., Gosselin, L., 2019. Predicting Domestic Hot Water Demand Using Machine Learning for Predictive Control Purposes, p. 6.
de Santiago, J., Rodriguez-Villalón, O., Sicre, B., 2017. The generation of domestic hot water load profiles in Swiss residential buildings through statistical predictions. Energy Build. 141, 341–348.
Jordan, U., Vajen, K., 2001. Realistic domestic hot-water profiles in different time scales. Report for IEA-SHC Task 26.
Kalani, H., Sardarabadi, M., Passandideh-Fard, M., 2017. Using artificial neural network models and particle swarm optimization for manner prediction of a photovoltaic thermal nanofluid based collector. Appl. Therm. Eng. 113, 1170–1177.
Diez, F., Navas-Gracia, L., Martínez-Rodríguez, A., Correa-Guimaraes, A., Chico-Santamarta, L., 2019. Modelling of a flat-plate solar collector using artificial neural networks for different working fluid (water) flow rates. Sol. Energy 188, 1320–1331.
Feng, Y., Hao, W., Li, H., Cui, N., Gong, D., Gao, L., 2020. Machine learning models to quantify and map daily global solar radiation and photovoltaic power. Renew. Sustain. Energy Rev. 118, 109393.
Shafieian, A., Parastvand, H., Khiadani, M., 2020. Comparative and performative investigation of various data-based and conventional theoretical methods for modelling heat pipe solar collectors. Sol. Energy 198, 212–223.
Correa-Jullian, C., Cardemil, J.M., Droguett, E.L., Behzad, M., 2020. Assessment of Deep Learning techniques for Prognosis of solar thermal systems. Renew. Energy 145, 2178–2191.
Dong, N., Chang, J.-F., Wu, A.-G., Gao, Z.-K., 2020. A novel convolutional neural network framework based solar irradiance prediction method. Int. J. Electr. Power Energy Syst. 114, 105411.
Behera, M.K., Nayak, N., 2020. A comparative study on short-term PV power forecasting using decomposition based optimized extreme learning machine algorithm. Eng. Sci. Technol., Int. J. 23 (1), 156–167.
Alizamir, M., Kim, S., Kisi, O., Zounemat-Kermani, M., 2020. A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the USA and Turkey regions. Energy, 117239.
Jung, Y., Jung, J., Kim, B., Han, S., 2020. Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: Case study of South Korea. J. Cleaner Prod. 250, 119476.
Kosovic, I.N., Mastelic, T., Ivankovic, D., 2020. Using Artificial Intelligence on environmental data from Internet of Things for estimating solar radiation: Comprehensive analysis. J. Cleaner Prod. 121489.
Dudkiewicz, E., Fidorów-Kaprawy, N., 2017. The energy analysis of a hybrid hot tap water preparation system based on renewable and waste sources. Energy 127, 198–208.
Heidari, A., Roshandel, R., Vakiloroaya, V., 2019. An innovative solar assisted desiccant-based evaporative cooling system for co-production of water and cooling in hot and humid climates. Energy Convers. Manage. 185, 396–409.
Fadl, M., Eames, P.C., 2019. An experimental investigation of the heat transfer and energy storage characteristics of a compact latent heat thermal energy storage system for domestic hot water applications. Energy 188, 116083.
Yu, Z., Yu, J., Cui, Y., Tao, D., Tian, Q., 2019. Deep Modular Co-Attention Networks for Visual Question Answering, pp. 6281–6290.
Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:14061078.
Lipton, Z.C., Berkowitz, J., Elkan, C., 2015. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:150600019.
Sutskever, I., Vinyals, O., Le, Q., 2014. Sequence to sequence learning with neural networks. Advances in NIPS.
Rahman, A., Srikumar, V., Smith, A.D., 2018. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl. Energy 212, 372–385.
Li, P., Zhou, K., Lu, X., Yang, S., 2020. A hybrid deep learning model for short-term PV power forecasting. Appl. Energy 259, 114216.
Omar, F., Omar, F., Bushby, S.T., 2013. Simulating Cccupancy in the NIST Net-zero Energy Residential Test Facility, US Department of Commerce, National Institute of Standards and Technology.
National Institute of Standards and Technology, 2019. Available on: https://nist.gov/.
Balke, E.C., Healy, W.M., Ullah, T., 2016. An assessment of efficient water heating options for an all-electric single family residence in a mixed-humid climate. Energy Build. 133, 371–380.
Ullah, T. and Healy, W.M., 2016. The performance of an auxiliary heat pump water heater installed in a dual-tank system in a net-Zero energy residence.
Ullah, T., Healy, W.M., 2015. The Effect of Piping Heat Losses on the Efficiency of Solar Thermal Water Heating in a Net-Zero Energy Home.
Domich, P.D., Pettit, B., Fanney, A.H., Healy, W.M., 2015. Research and Development Opportunities for the NIST Net Zero Energy. Residential Test Facility.
Vinayavekhin, P., Chaudhury, S., Munawar, A., Agravante, D.J., De Magistris, G., Kimura, D., Tachibana, R., 2018. Focusing on What is Relevant: Time-Series Learning and Understanding using Attention. IEEE, pp. 2624–2629.
Zhou, H., Zhang, Y., Yang, L., Liu, Q., Yan, K., Du, Y., 2019. Short-Term Photovoltaic Power Forecasting Based on Long Short Term Memory Neural Network and Attention Mechanism. IEEE Access 7, 78063–78074.
Choi, H., Cho, K., Bengio, Y., 2018. Fine-grained attention mechanism for neural machine translation. Neurocomputing 284, 171–176.
Li, W., Guo, D., Fang, X., 2018. Multimodal architecture for video captioning with memory networks and an attention mechanism. Pattern Recogn. Lett. 105, 23–29.
Song, K., Yao, T., Ling, Q., Mei, T., 2018. Boosting image sentiment analysis with visual attention. Neurocomputing 312, 218–228.
Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:14090473.
Al-Karawi, K.A., Mohammed, D.Y., 2019. Early reflection detection using autocorrelation to improve robustness of speaker verification in reverberant conditions. Int. J. Speech Technol. 22 (4), 1077–1084.
Ji, S.-H., Baek, U.-J., Shin, M.-G., Goo, Y.-H., Park, J.-S., Kim, M.-S., 2019. Best Feature Selection using Correlation Analysis for Prediction of Bitcoin Transaction Count. IEEE, pp. 1–6.
Evans, J.D., 1996. Straightforward statistics for the behavioral sciences. Thomson Brooks/Cole Publishing Co.
Healy, W., Chen, T.H., Dougherty, B., Fanney, A.H., Ullah, T., Payne, W.V., Ng, L., Omar, F., 2018. Net Zero Energy Residential Test Facility Instrumented Data; Year 1.