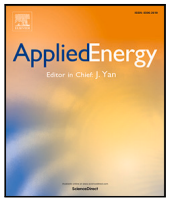
Applied Energy 367 (2024) 123379

Contents lists available at ScienceDirect

Applied Energy

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



Non-intrusive thermal load disaggregation and forecasting for effective HVAC systems

Naoya Kaneko [a], Kazuki Okazawa [a], Dafang Zhao [a,*], Hiroki Nishikawa [a], Ittetsu Taniguchi [a], Hiroyuki Murayama [b], Yoshinori Yura [b], Masakazu Okamoto [b], Francky Catthoor [c,d], Takao Onoye [a]

[a] Osaka University, 1-5 Yamadaoka, Suita, Osaka, Japan
[b] DAIKIN Industries, Ltd., 1-1 Nishi-Hitotsuya, Setsu, Osaka, Japan
[c] imec, Kapeldeef 75, 3001 Heverlee, Belgium
[d] ESAT, KULeuven, Kasteelpark Arenberg 10, 3001 Heverlee, Belgium

ARTICLE INFO

Keywords:
Non-intrusive thermal load monitoring
Thermal load disaggregation
Neural network
Time series forecasting

ABSTRACT

Non-Intrusive Thermal Load Monitoring (NITLM) tracks the sub-loads generated by each heat source (e.g. occupants, equipment, solar radiation etc.) from the total thermal load and indirectly provides a room's thermal properties without additional sensors. Since sub-loads can improve the efficiency of HVAC systems, NITLM is a very attractive technology for building energy management. NITLM has traditionally focused on analyzing past and present sub-loads. However, by forecasting future sub-loads, HVAC systems will be able to schedule operations that take into account the thermal properties of future rooms. This work focuses on a new NITLM framework that forecasts future sub-loads based on the current and past total thermal loads. In experiments, we selected occupant loads that are closely connected to HVAC systems and performed sub-load forecasting using two types of approaches. One is a two-step approach that separately performs them in turn. This approach use separately trained model for disaggregation and forecasting, this allow us to fine-tuning the hyper-parameter for dedicate model. Moreover, the two-step approach can take into account the different properties and difficulties of each inference, resulting in smaller errors in sub-load forecasting. The other is an integrated approach. This approach combines load disaggregation and forecasting into a single estimation process, eliminating error propagation and reducing overall error in sub-load forecasting. Moreover, this approach utilizes the Adaptive Moment Estimation (Adam) algorithm for effective parameter optimization, enabling complex training and improving the accuracy of sub-load forecasting. We conducted evaluations of thermal load disaggregation and forecasting across a range of realistic building scenarios. The findings indicate that the integrated approach predicts sub-loads with a MAE that is up to 34.9% lower than that of the two-step approach. Additionally, it identifies occupants presence with an 18.5% higher F-score. This demonstrates its enhanced suitability for accurately predicting sub-loads and estimating future occupancy schedules.

1. Introduction

A surge in urban development and the pursuit of enhanced daily comfort are fueling a rise in energy consumption within buildings. Such buildings currently constitute 40% of the total energy consumption and 36% of greenhouse gas emissions in numerous EU countries [1]. From this perspective, creating efficient energy management systems in buildings is imperative for advancing energy conservation objectives and fostering a decarbonized society. Heating, Ventilation, and Air Conditioning (HVAC) systems, which constitute 45% of a building's energy consumption [2], represent a primary focus for achieving energy efficiency without compromising comfort. Consequently, the development of advanced HVAC systems holds significant importance in the realm of building energy management.

The emergence of load monitoring is a promising and pivotal breakthrough in curtailing energy consumption. A comprehensive understanding of load enables the occupants and the users of buildings to assess their energy consumption statuses and refine their practices to reduce energy usage [3]. Furthermore, monitoring sub-loads attributed to individual devices within a room unveils a nuanced breakdown of energy consumption, revealing equipment usage patterns and providing

* Corresponding author.
E-mail address: zhao.dafang@ist.osaka-u.ac.jp (D. Zhao).

https://doi.org/10.1016/j.apenergy.2024.123379
Received 18 January 2024; Received in revised form 29 March 2024; Accepted 1 May 2024
Available online 11 May 2024

valuable insight into how a room is used. This approach also has the potential to support energy management in buildings. Conventionally, sub-loads have been monitored by sensors attached to devices. Such an approach is called Intrusive Load Monitoring (ILM) [4]. Due to the nature of sensing, users need to install sensors on each device to properly measure such sub-loads; unfortunately, this requirement is time-consuming and expensive because it increases the number of devices. In addition, this approach also raises privacy issues.

Hart [5] responded to this challenge by proposing Non-Intrusive Load Monitoring (NILM). NILM measures the overall power consumption from smart meters installed in buildings and disaggregates the total power consumption into sub-loads by empirically analyzing an entire building's power consumption. NILM is categorized as a low-cost, sub-load monitoring method since it does not require the installation of sensors, unlike ILM. However, it is crucial to acknowledge that NILM exclusively monitors the sub-loads generated by electronic equipment. As a result, the insights into room usage provided by NILM are confined to electronic devices. It less effectively monitors elements that do not rely on electricity [6]. HVAC operations are intricately linked to factors independent of electricity usage, such as solar radiation and the presence and number of people. Ignoring these aspects is impractical for HVAC systems.

Non-Intrusive Thermal Load Monitoring (NITLM) was proposed by Ricardo [7] as an NILM extension that considers thermal loads. Unlike NILM, NITLM monitors the overall thermal load (instead of the power one) and decomposes it into multiple thermal sub-loads. Exploiting the thermal nature, NITLM can obtain thermal sub-loads including but not limited to solar radiation, air thermal diffusion, and human thermal radiation. The accurate analysis of human thermal radiation offers such optimal control HVAC systems as on/off and temperature setting [8,9] and determines the priority of power allocation under power usage restrictions [10].

Unfortunately, existing works are hampered by a limitation where NITLM can obtain only past and present thermal sub-loads, meaning that obtaining the future thermal sub-loads of occupancy is impossible for estimating future occupancy schedules. To optimize HVAC operations in line with anticipated occupancy schedules, we have expanded the previous work [11] in such a way that the NILM concept to include not only the disaggregation of building thermal loads but also the forecasting of variations in sub-thermal loads. This advancement enables us to predict occupancy schedules by analyzing future thermal loads of occupants. This novel NITLM framework is named Non-Intrusive Thermal Load Disaggregation and Forecasting (NITLDF). However, our previous work forecasted sub-loads are for only a simple building on a simulator without providing a detailed explanation of the NITLDF mechanism. In addition, since the building conditions (e.g. weather, number of occupants, and HVAC operation schedules) undoubtedly affect the accuracy of sub-load forecasting, the effectiveness of the NITLDF approach needs to be evaluated in multiple buildings based on realistic data.

Building on the previous work [11], this work continues its focus on NITLDF and provides a more detailed description of the NITLDF approach and a performance evaluation along more realistic and multiple scenarios. The NITLDF framework consists of two types of approaches, namely two-step and integrated approach. By treating thermal load disaggregation and forecasting as separate inferences, the two-step approach can take into account the different properties and difficulties of each inference, resulting in smaller errors in sub-load forecasting. On the other hand, the integrated approach combines load disaggregation and forecasting into a single estimation process, eliminating error propagation and reducing overall error in sub-load forecasting. The following are its two main contributions.

- The previous work [11] did not describe the details of NITLDF. In contrast, this paper added comprehensive descriptions of the mechanism in our proposals.
- NITLDF was performed on five actual buildings, each of which was evaluated in multiple, more realistic scenarios based on actually observed HVAC operating schedules and weather data. This strategy allows for a more general evaluation of the NITLDF approach.

The rest of this paper is organized as follows. Section 2 describes recent work on load disaggregation and load forecasting. Section 3 overviews our NITLDF approach. Section 4 describes our experimental setup and results and compares several NITLDF approaches. Section 5 concludes this paper.

2. Related works

2.1. Load disaggregation

A number of studies have monitored indoor elements at low cost to improve the efficiency of energy management at the building and room levels. NILM breaks down the power consumption of each appliance from the total power consumption and determines whether the appliance is on or off using combinatorial optimization and edge extraction [5]. Many NILM methods have been proposed, such as Hidden Markov Models (HMMs) [12–14] and Graph Signal Processing (GSP) [15,16]. In recent years, securing computing resources has been simplified, and many approaches using machine learning have appeared. Among them, the NILM algorithm that incorporates Neural Network (NN) is attracting attention. Convolutional Neural Network (CNN) [17–20], Long Short-Term Memory (LSTM) [21–24], and Gated Recurrent Unit (GRU) [22,23] are typical examples of NN. Yadav et al. [22] incorporated such Recurrent Neural Networks (RNNs) as GRU and LSTM into a learning framework called a sequence to sequence. Load disaggregation using these techniques successfully identified the status of four appliances at a dairy farm with high accuracy [22]. These NN-based load disaggregation methods capture nonlinear time-series trends and contribute to highly accurate load disaggregation. On the other hand, since NILM is used for power consumption, only devices that consume electricity can be monitored. Thermal sub-loads closely related to HVAC systems are not disaggregated and monitored.

Ricardo et al. [7] shed light on NILM for monitoring thermal loads. Ricardo uses a Building Energy Simulation (BES) model to recreate buildings and monitor thermal loads. Xiao et al. [6,25] develop a machine learning based NITLM that monitors sub-loads by disaggregating the total thermal load. By feeding time and weather data into a Random Forest (RF) model, they achieved highly accurate extraction of four sub-loads. Okazawa et al. [26] also proposed a NITLM model using RNN. Their model used GRU and LSTM and resolved the load of occupants with higher accuracy than random forest. All of these load disaggregations [6,25,26] target current and past loads and do not focus on inference to estimate future sub-loads. Estimating future sub-loads has finally begun to receive attention in our previous work [11], where sub-loads are forecasted by an RNN-based inference approach. However, since the evaluation was conducted in a single building, experiments in more diverse and realistic buildings are needed to more generally evaluate the approach's effectiveness. The previous work [11], which is still under development, is extended by this work.

2.2. Load forecasting

Load forecasting supports the operation scheduling of HVAC systems, and various algorithms have been proposed. Such linear models as linear regression (LR) [27] and support vector machines (SVMs) [28,29] initially garnered attention. These models, which can learn with a relatively small amount of data, have achieved thermal load forecasting by combining such weather data as outside temperature [27,28]

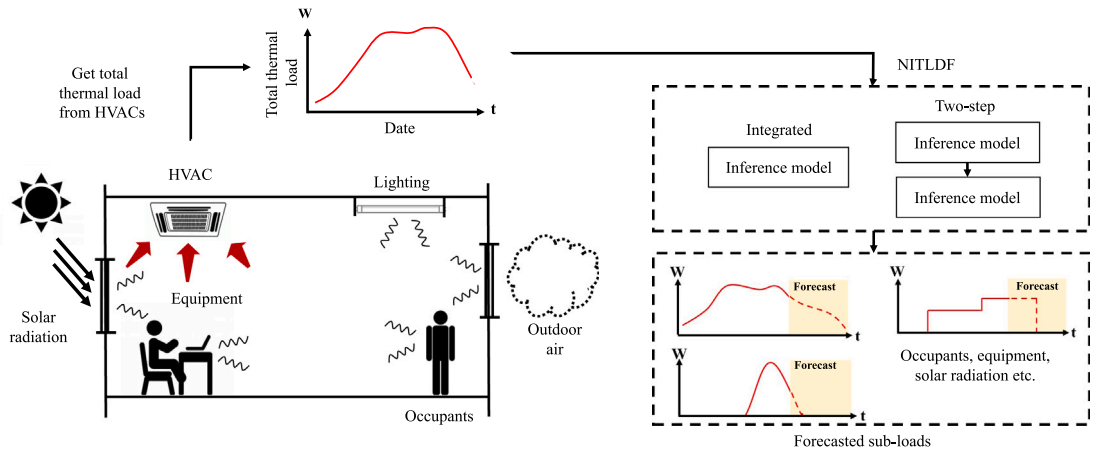


Fig. 1. Thermal load and overview of proposed framework.

and humidity [28]. However, their use struggles to cope with nonlinear and complex data. Although nonlinear data can be dealt with by incorporating sensitivity analysis [30], evaluations under various scenarios are required, and issues remain. On the other hand, NN-based forecasting methods are attracting attention due to their excellent analysis of complex and nonlinear data. Since NN can uncover rules that are difficult to clarify with linear models, it achieves thermal load forecasting with high accuracy, [31–34]. Among the NN schemes, RNN is one of the most suitable models for forecasting thermal load because it incorporates the time series of data. In fact, Stefan et al. [35] achieved highly accurate forecasting by combining LSTM, one of the most famous RNN models, with feature selection using correlation analysis. However, these works [31–34] have only focused on forecasting the total thermal load without disaggregating it into sub-loads, some of which, including forecasted occupant load, can provide information that assists HVAC scheduling. Therefore, for HVAC scheduling, sub-load forecasting that has both load disaggregation and load forecasting aspects is also desirable.

3. NITLDF and approaches

In this section, we introduce NITLDF, a novel NITLM framework, and explain two NITLDF approaches. One is a two-step approach, which has a pair of inference steps that combine a technique of the conventional NITLM and a forecasting model that predicts the thermal load for each sub-load. The other is a predicting approach: that is, a one-shot, integrated forecasting model. Both approaches use RNN for inferences.

3.1. NITLDF framework

Fig. 1 shows the flow of sub-load monitoring and forecasting. For simplicity, we assume five heat sources in a single room: lighting, solar radiation, outdoor air, equipment, and occupants. We cannot measure each thermal sub-load. We also assume that HVAC cools a room with a corresponding cooling load, whose measurement is based on the total thermal loads from the heat sources. Thus, the load from HVAC is assumed to be always equivalent to the loads from the five heat sources.

Traditionally, NITLM obtains a measured thermal load from HVAC and disaggregates it to investigate a room's thermal properties. Thermal load forecasting provides insight into thermal storage and avoids intensive HVAC operation. However, NITLM cannot take into account the future thermal properties of a room, and thermal load forecasting cannot reveal the causes of future thermal load peaks. NITLDF solves these problems by combining NITLM and thermal load forecasting into a more effective HVAC system operation. Two approaches are devised for NITLDF. An integrated approach simultaneously infers thermal load disaggregation and forecasting. The two-step approach infers them separately with multiple models. Each approach is given a thermal load, temperature, humidity, and the time and calendar information for a temporal period. The weather and time information are related to the thermal load and support the analysis. In fact, many studies dealing with thermal load forecasting rely on weather data, such as temperature [35–37]. In addition, several works on NITLM have reported that time [6,26] and calendar [11] information are effective in addition to weather data. Our proposed NITLDF represents a unique approach by integrating thermal load disaggregation (NITLM) and forecasting in a unified framework. This integration allows for leveraging synergies between NILM and forecasting, potentially enhancing the accuracy and efficiency of both tasks. We believe this novel contribution fills a gap in the existing literature, where these tasks are often treated separately.

3.2. A two-step forecasting approach

In the previous section, we introduced the NITLDF task, which advances the NILM concept by extending it to the disaggregation of building thermal loads, not just decomposing energy usage but also forecasting changes in thermal loads. Moreover, the work's principal aim is to forecast future occupancy schedules by projecting the thermal loads of anticipated future occupants. This requires breaking down the disaggregation process and targeting the forecasting at specific aspects, in line with the primary objective of this work. In other words, we acknowledge the traditional approach's merit in dividing the NITLDF task into two distinct sub-tasks: NILM for disaggregation and forecasting. Load disaggregation and forecasting are different types of inferences, and the optimal parameters for each inference model may be different. The two-step approach treats these two inferences separately, trains the model separately, and optimizes their parameters. The two inference models may forecast sub-loads with small errors if the difficulty and the properties of these inferences are different.

This approach is further subdivided into two approaches based on the order of the inferences. One is the disaggregation-to-forecasting (D2F) approach (Fig. 2(a)). The operation pipeline of the D2F approach is outlined as follows:

1. Utilize temperature, humidity, time, calendar, and thermal model data as inputs for the disaggregation model.
2. The disaggregation model divides the total thermal load into various sub-loads.

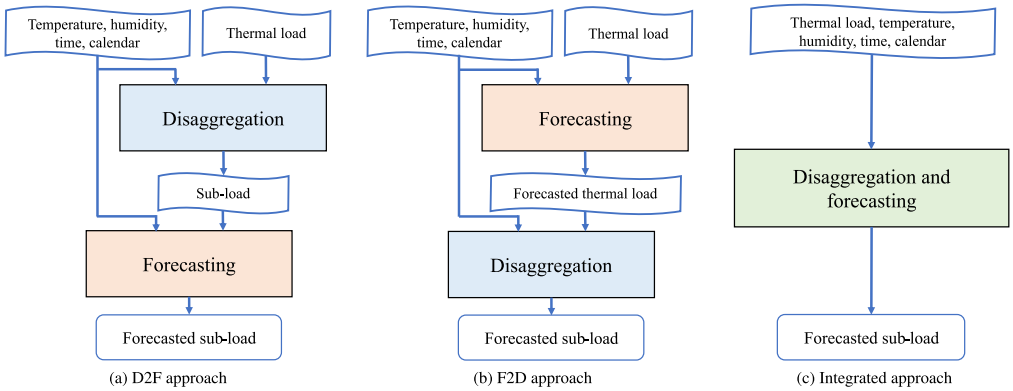


Fig. 2. Proposed approaches.

3. Employ the disaggregated sub-load, together with weather and time information, as inputs for an independent forecasting model to predict future sub-loads.

The other is the forecasting-to-disaggregation (F2D) approach (Fig. 2(b)), which forecasts the thermal loads and disaggregates them into sub-loads. This approach forecast the sub-load in an opposite way of D2F approach. The operation pipeline of the D2F approach is outlined as follows:

1. Utilize temperature, humidity, time, calendar, and thermal model data as inputs for the forecasting model.
2. The forecasting model utilizes this information to predict the future total thermal load.
3. The disaggregation model then separates the predicted total thermal load into distinct sub-loads.

Since the forecasting target changes depending on the inference order, the error may greatly fluctuate. The D2F approach forecasts the disaggregated sub-load; the F2D approach forecasts the total thermal load. Differences in error between them reflect the characteristics of the sub-loads and the room's total thermal load.

3.3. An integrated forecasting approach

In the above two-step approach, each disaggregation and forecasting step has a dedicated inference model; in the integrated approach, only one inference model is used. Although load disaggregation and load forecasting are different inferences, an integrated approach, which simultaneously performs both types of inference, is unique to NITLDF and includes both load disaggregation and forecasting in one estimation. Not surprisingly, it is not addressed in the numerous load disaggregation and forecasting works in Section 2. Fig. 2(c) shows the integrated approach. Similar to the two-step approach, the inference model is given the total thermal load, the temperature, humidity, time, and calendar information. However, disaggregation and forecasting are performed simultaneously within one inference process, and future sub-loads are output from the model. In the training process, parameters are optimized based on the Adaptive Moment Estimation (Adam) algorithm [38]. The integrated approach requires simultaneous analysis of the changes in the time direction and the proportion of sub-loads in the total load at a certain time, and so the training difficulty level surpasses the two-step approach. The Adam algorithm enables such complex training by performing effective parameter optimization that incorporates momentum and RMSProp. Furthermore, due to its structure, since inference is performed only once, no error propagation occurs. This idea suggests that the integrated approach can forecast sub-loads with less error than the two-step approach.

Table 1
Detailed information on building models.

Building	A	Y	N	O	R
Location	Osaka		Tokyo		
Stories	6	6	9	9	9
Target floor	2	3	8	3	4
Square [m^2]	200	430	320	600	120
Type			Office		

4. Experiments

In this section, we compared the two-step and integrated approaches (described above) in five buildings. The former approach represents both D2F and F2D. To evaluate the performance of these two approaches, we select two well-known machine learning structures, namely, LSTM and GRU, as the inference model. In this paper, we focus on the occupant load that is closely involved with the HVAC system. We compared these approaches under different machine learning structures based on the error between the forecasted values and the ground truth and the accuracy of determining the presence or the absence of occupants in the room using F-scores.

4.1. Setup

This section describes the details of the dataset, the parameter optimization, and the experimental environment. Fig. 3 shows an overview of our experiment. Since there are no publically available datasets for thermal loads, we generated one with EnergyPlus [39], a physics simulator. The weather, heat source and HVAC schedules, and the building's model are given to EnergyPlus to calculate the thermal loads.

Fig. 4 shows the room layout of the building model used in the experiment. These are actual buildings in Osaka and Tokyo. Indoor units are represented by symbols from A1 to H3. The same letter is used for indoor units that are connected to the same outdoor unit (e.g., A1-A10 are connected to the same outdoor unit). In EnergyPlus, since only one indoor unit can be set in a room, we performed a simulation by dividing one room into multiple zones using virtual walls that have the same properties as air.

Table 1 shows detailed information about the buildings. Buildings A and Y are located in Osaka; the others are in Tokyo. Therefore, these building models are given weather data that are comprised of the observed values for each region in their thermal load calculation. Although each building has multiple floors, the experiment targets just one floor. Furthermore, the area of each floor is different.

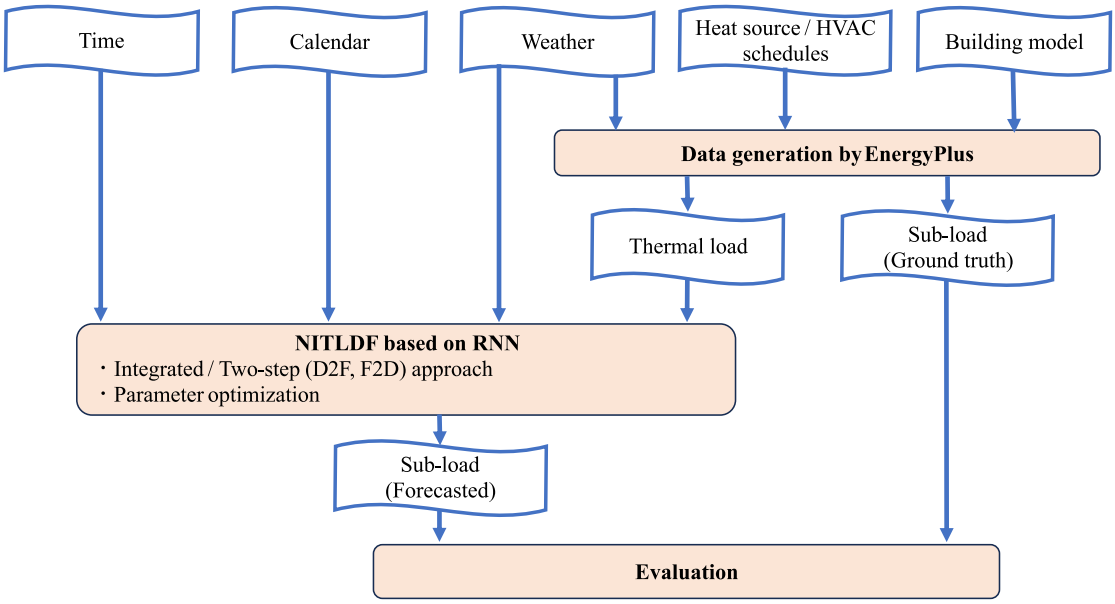


Fig. 3. Overview of data generation and NITLDF.

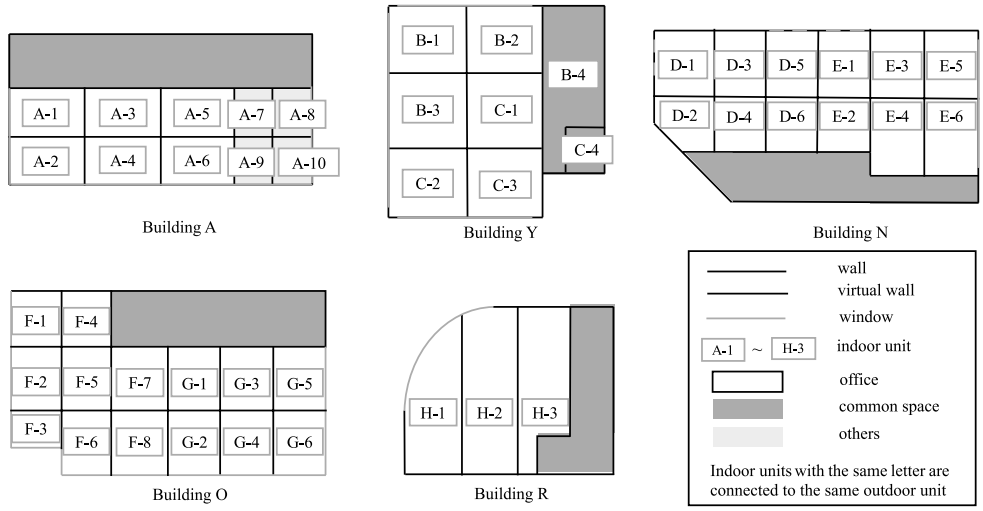


Fig. 4. Room layout of building model given to EnergyPlus.

Table 2 shows building A's basic schedule for such indoor heat sources as occupants, lighting, and equipment. The maximum number of occupants in it is 0.1 person/m^2, and the maximum thermal load from lighting and equipment is 12 W/m^2. Each heat source is scheduled at a percentage of its maximum value for each hour. The settings for the heat source in EnergyPlus are based the office section of japan energy conservation standards for non-residential buildings [40]. It should be noted that, to simplify the model, we assigned the same heat source value to both lighting and equipment. However, in reality, these values could vary significantly due to the diversity of luminaries, device typologies, and equipment types. Each schedule has four patterns, depending on HVAC on/off, weekdays, and holidays. In addition, the schedules of the occupants in the room are given a maximum of 20% noise based on random numbers that follow a uniform distribution. We actually measured the HVAC schedule in each building to generate a dataset from June 1 to September 30. The data resolution is one hour. More detail information of heat sources schedules in building Y, N, O, R are listed in Appendix.

The calculated thermal load is given to each NITLDF approach along with the weather and calendar information. In the training process, each inference model is given 672 points (four weeks) as training data, and 168 points (one week) as validation data. The parameters of the RNN, which makes the inferences in each approach, are optimized in the training phase using the Treestructured Parzen Estimator (TPE) algorithm [41], which is a Bayesian optimization method. In the experiment, we optimized the number of neurons, the batch size, the max epochs, and the input length through 60 trials. In the evaluation process, 168 points (one week) were used as test data. We evaluated the accuracy by comparing the occupant sub-loads forecasted by the two-step or integrated approaches with the sub-load calculated by EnergyPlus. The data intervals used for training, validation, and testing can slide in 168-point increments, and the sub-load is forecasted until

Table 2
Heat source schedules in building A.

Heat source		Value	
		Weekday	Holiday
Occupants	Maximum	0.1 person/m^2	
	Schedule (HVAC ON)	00:00–06:00 (0%) 06:00–08:00 (16%) 08:00–12:00 (80%) 12:00–13:00 (48%) 13:00–17:00 (80%) 17:00–19:00 (40%) 19:00–20:00 (24%) 20:00–21:00 (16%) 21:00–24:00 (0%)	00:00–24:00 (16%)
	Schedule (HVAC OFF)	00:00–24:00 (0%)	00:00–24:00 (0%)
Lighting	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–06:00 (0%) 06:00–08:00 (50%) 08:00–12:00 (100%) 12:00–13:00 (50%) 13:00–19:00 (100%) 19:00–21:00 (80%) 21:00–24:00 (0%)	00:00–24:00 (50%)
	Schedule (HVAC OFF)	00:00–24:00 (0%)	00:00–24:00 (0%)
Equipment	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–06:00 (25%) 06:00–08:00 (25%) 08:00–12:00 (100%) 12:00–13:00 (80%) 13:00–19:00 (100%) 19:00–21:00 (50%) 21:00–24:00 (25%)	00:00–24:00 (25%)
	Schedule (HVAC OFF)	00:00–24:00 (25%)	00:00–24:00 (25%)

September 30. The experiments were performed using TensorFlow 2.5.0 [42] on a GeForce RTX 3070 with 8 GB of VRAM and CUDA 11.2 [43].

4.2. Metrics

Each NITLDF approach was evaluated with three error metrics and one accuracy metric. Each NITLDF approach was evaluated on three error metrics and one accuracy metric. MAE (1), RMSE (2), both of which are typical indicators of time-series analysis [44], and MRE (3) which is a relative indicator, were used as error indicators. These metrics, which have been previously used by NITLM works [11,25,26], are calculated as follows. n is the amount of evaluation data. y_i is the forecasted value, and $\hat{y}_i$ is the ground truth. The simulated load data is used as the ground truth:

$$MAE = \frac{1}{n}\sum_{k=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

$$MRE = 100 \cdot \frac{\sum_{k=1}^{n}|y_i - \hat{y}_i|}{\sum_{k=1}^{n}|y_i|} \tag{3}$$

An F-score is used as an accuracy indicator. It is one of the most popular indicators for measuring the accuracy of binary classification and is often used in NILM works [22,45]. F-scores are calculated by Eq. (6) using *Recall* and *Precision* given by Eqs. (4) and (5). TP represents the proportion of the estimated positives that are actually positive, and FP represents the proportion of those that are actually negative. FN represents the proportion of the estimated negative results that were actually positive:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F\text{-}score = \frac{2 \cdot Recall \cdot Precision}{Recall + precision} \tag{6}$$

In these experiments, if an occupant load exists, it is defined by the presence of occupants and labeled as positive. On the other hand, when the occupant load is 0, it is defined as the absence of occupants and is labeled as negative.

4.3. Results

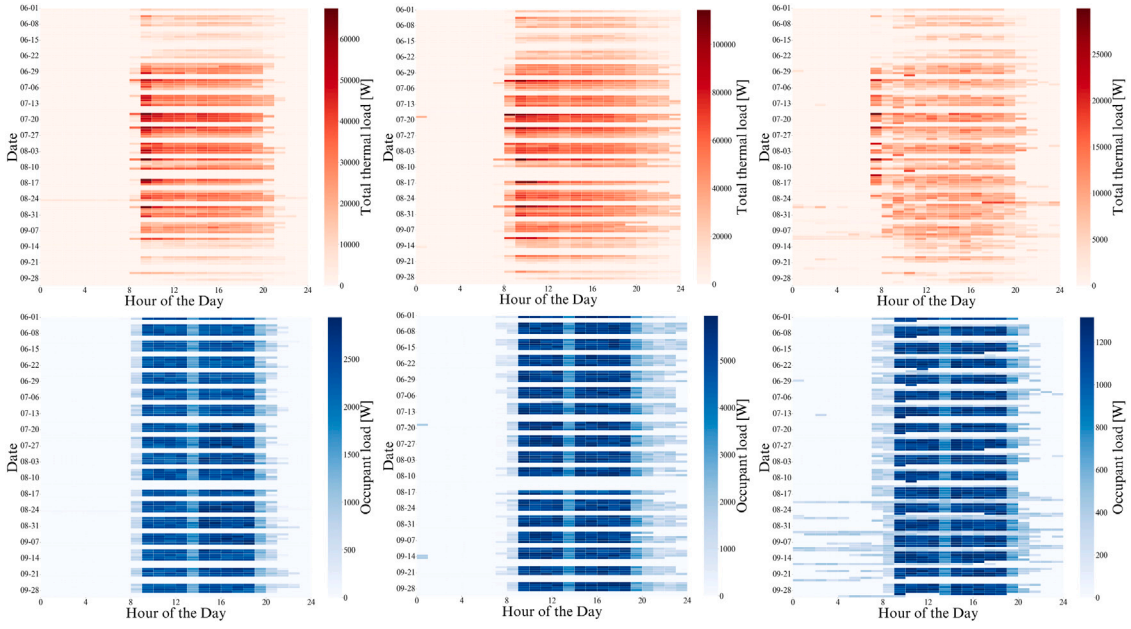
The NITLDF results for the occupant loads are shown in Table 3 and present the MAE, RMSE, MRE, and F-scores for the D2F, F2D, and integrated approaches under different machine learning structures. The best value among the three approaches for each metric and building is highlighted. The findings indicate that for both MAE and MRE, the LSTM surpasses the GRU by margins of up to 21.9% and 21.8%, with average improvements of 5.0% and 4.1%, respectively. However, in terms of RMSE and F score, the superiority of the LSTM model over the GRU model is narrower, peaking at 15.3% and 10.0%, respectively. Furthermore, the average performance improvements do not show significant differences. Therefore, this experiment suggests that LSTM exhibits marginally superior forecasting capabilities compared to GRU.

Focusing on the approaches, the integrated approach is the best among all the metrics and outperformed the two-step approaches, including the D2F and F2D approaches. The difference is particularly remarkable for building N, where the integrated approach shows a maximum of a 34.9% lower MAE and an 18.5% higher F-score than the two-step approach in LSTM structures. Therefore, in general, an integrated model is clearly more suitable for forecasting sub-loads to detect the number of occupants or their absence. However, for some metrics of buildings A and Y, the difference in error between D2F and the integrated approach was small, and for building R, the superiority of D2F and the integrated approach was reversed. This result suggests that the integrated approach is unsuitable for some buildings.

To investigate the apparent superiority reversal in building R, Fig. 5 shows the total thermal load and occupant loads of buildings N, O, and R. Simulated by EnergyPlus, the total thermal load is given to the inference models as input data, and the occupant loads are the ground truth. For each building, the daily and hourly changes in the total thermal load are given in the red heat map. On the other hand, the occupant loads are given by a similar blue heat map. The vertical axis shows the dates from June 1 to September 30, and the horizontal axis shows the times. The redder the color is, the greater the total thermal load, and the bluer the color is, the greater the occupant loads. These results show the operating hours of each building's HVAC systems and the hours when occupants are present. These three buildings are all located in Tokyo, so the outdoor temperature and humidity given for the sub-load forecasting are identical. Naturally, identical calendar information is also given. Therefore, we believe that the total thermal load greatly impacted the forecasting results of these three buildings. Looking at Fig. 5(c), building R's total thermal load often occurred all night long in late August, and the occupant load also occurred all night long. No such trend change is seen before late August or in buildings N and O. This change in the load trend may have complicated forecasting the sub-loads for building R. In addition, this may have particularly affected the integrated approach, where training the inference model is more difficult than the two-step approach, perhaps reversing the superiority of D2F and the integrated approach in building R. If the load pattern switches suddenly, as in building R, the integrated approach may lose its advantage over the two-step approach due to the difficulty of the training model. This suggests a certain trade-off between increasing the

Table 3
Performance of NITLDF approaches.

Building	Approach	LSTM				GRU			
		MAE [W]	RMSE [W]	MRE [%]	F-score [%]	MAE [W]	RMSE [W]	MRE [%]	F-score [%]
A	D2F	129.2	**265.1**	22.4	67.5	142.2	305.8	24.7	66.1
	F2D	151.1	313.0	26.2	69.7	151.9	313.7	26.4	71.1
	Integrated	**122.1**	279.0	**21.2**	**71.3**	**134.3**	**289.7**	**23.3**	**74.9**
Y	D2F	292.8	551.5	25.1	**80.3**	273.0	**475.7**	23.4	75.3
	F2D	332.3	619.6	28.5	75.5	320.5	550.7	27.5	75.5
	Integrated	**255.8**	**494.3**	**22.0**	77.2	**267.8**	509.0	**23.0**	**77.4**
N	D2F	106.8	281.3	15.6	66.9	120.4	296.8	17.6	60.9
	F2D	133.3	305.4	19.5	63.7	123.4	280.6	18.1	**70.1**
	Integrated	**86.8**	**255.7**	**12.7**	**75.5**	**91.0**	**248.7**	**13.3**	67.8
O	D2F	283.4	661.7	19.3	71.6	345.4	717.7	23.5	73.7
	F2D	334.7	777.1	22.8	73.3	346.9	732.7	23.6	69.7
	Integrated	**233.2**	**533.9**	**15.9**	**78.2**	**236.2**	**524.3**	**16.1**	**79.0**
R	D2F	**137.2**	**219.7**	**33.2**	**74.8**	**146.2**	**223.6**	**35.4**	74.1
	F2D	149.6	229.5	36.2	72.6	146.9	218.1	35.5	74.3
	Integrated	147.9	249.9	35.8	74.2	155.1	258.9	37.5	**77.0**



(a) Building N (b) Building O (c) Building R

Fig. 5. Total thermal load and occupants load of buildings N, O, and R.

complexity for higher-quality forecasting and the ability to respond to rapid load pattern switching.

On the other hand, the F2D approach shows large errors in almost all the buildings and indicators. Even when compared with the D2F approach, which is identical as the two-step approach, there is a difference in the error. The discrepancy between the D2F and F2D approaches is the order of the inference, and the order in which forecasting is performed after extracting the sub-loads may be more suitable for forecasting the load of the occupants.

To confirm the results of the occupant-load forecasting, the results for one week in August for building A are shown in Fig. 6. The vertical axis is the occupant load, and the horizontal axis is the date. The forecasting results by Integrated, D2F, and F2D are shown by red, green, and blue lines. The ground truth calculated by the simulator is shown by a black dotted line. Focusing on the integrated approach, it successfully captured the large decrease in the load around noon on weekdays. On the other hand, the two-step approaches of D2F and F2D successfully captured general waveforms, although it seems difficult to capture more detailed fluctuations over several hours. In Fig. 6, the D2F and F2D waveforms change almost horizontally after the load rises, suggesting that detailed waveform was not captured. These results support the superior value of the integrated approach over the two-step approach (Table 3). In addition, the thermal load increases earlier in the F2D approach than in the other approaches. Such several-hour error might have significantly increased the error of the F2D approach.

To illustrate the magnitude of the error in the two-step approach, we look at the results of the first step of each approach. Fig. 7 shows the results of the first step in the D2F approach: that is, the disaggregation

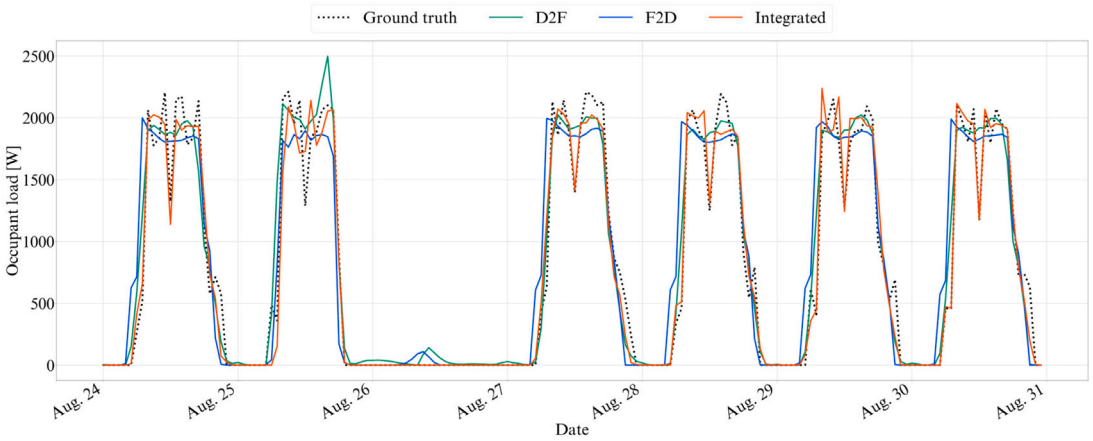


Fig. 6. NITLDF results using LSTM for building A.

step. The displayed results are the disaggregation results for building A for one week in August, a time period which is identical as in Fig. 6. The vertical axis represents the occupant load, and the horizontal axis represents the date. The disaggregated occupant load is shown in red; the ground truth calculated by the simulator is shown in black. Although the D2F approach failed to capture detailed waveforms in Fig. 7, it successfully captured the load decrease around noon in the disaggregation step, a result that suggests that the forecasting step may have caused the increase in the error in the D2F approach.

Fig. 8 shows the results of the first step in the F2D approach: that is, the forecasting step. Similar to Figs. 6 and 7, one week in August at building A is shown. The vertical axis represents the total thermal load, and the horizontal axis represents the date. The forecasted total thermal load is shown in red, and the ground truth calculated by the simulator is shown in black. Focusing on the time when the load increases in the early morning, F2D forecasts that the load will occur earlier than the ground truth, albeit just slightly. Since the total thermal load is much larger than the occupant load, the small error in Fig. 8 seems to have been a major factor in increasing the error during the F2D disaggregation step. Also in both Figs. 7 and 8, a load was incorrectly estimated on August 26, a day without an occupant load: a holiday. This error propagates and affects the final occupant-load forecasting result in Fig. 6.

This experiment forecasted the occupant loads in multiple realistic buildings. The integrated approach usually showed a better sub-load forecasting performance for most buildings compared to the two-step approaches, such as D2F and F2D. On the other hand, in buildings that experience rapid changes in load patterns, such as building R, the integrated approach's advantage is squandered. Both two-step approaches remain problematic, mainly in the forecasting step. Although D2F captured the load fluctuations in detail in the disaggregation step, it failed to capture them in the forecasting step. F2D forecasted the total thermal load in the forecasting step. Since the total thermal load greatly exceeds the occupant load, slight forecasting error created a very large error propagation. In this experiment, for a more basic and fair comparison between the integrated and two-step approaches, we unified the inference models for each approach to the RNN model, which has the highest accuracy in the current NITLM work. However, more advanced deep learning models can be applied, such as transformer and sequence to sequence models. Perhaps errors can be reduced by frequency analysis technology, which is used in many time-series forecasting techniques. A more comprehensive evaluation using these techniques is desired.

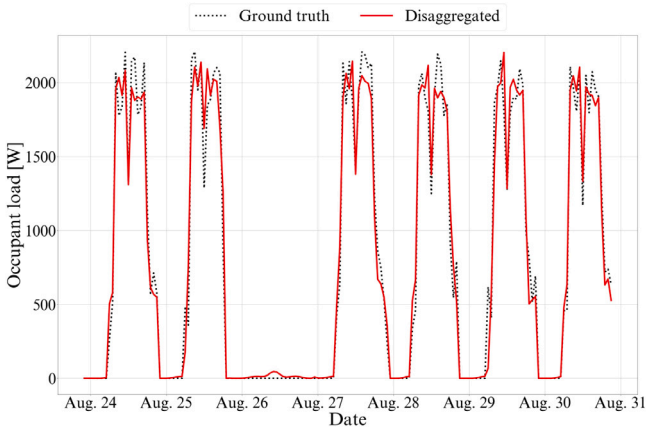


Fig. 7. Results of disaggregation step with D2F using LSTM for building A.

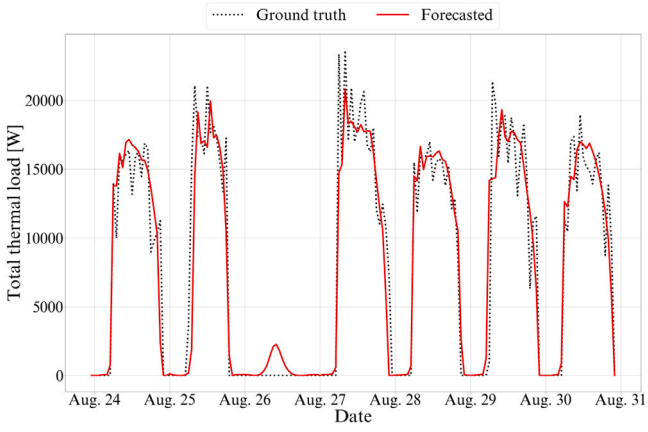


Fig. 8. Results of forecasting step with F2D using LSTM for building A.

5. Conclusions

In this study, we focused on a novel NITLM framework (NITLDF) and forecasted the occupant loads in five realistic buildings based on actual measurements. In our experiments, the occupant loads were forecasted by a two-step approach that infers thermal load disaggregation

Table A.4
Heat source schedules in building Y.

Heat source		Value	
		Weekday	Holiday
Occupants	Maximum	0.1 person/m^2	
	Schedule (HVAC ON)	00:00–07:00 (0%) 07:00–08:00 (16%) 08:00–12:00 (80%) 12:00–13:00 (48%) 13:00–18:00 (80%) 18:00–19:00 (40%) 19:00–20:00 (24%) 20:00–21:00 (16%) 21:00–24:00 (0%)	00:00–24:00 (16%)
	Schedule (HVAC OFF)	00:00–24:00 (0%)	00:00–24:00 (0%)
Lighting	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–06:00 (0%) 06:00–07:00 (50%) 07:00–12:00 (100%) 12:00–13:00 (50%) 13:00–20:00 (100%) 20:00–21:00 (80%) 21:00–23:00 (50%) 23:00–24:00 (0%)	00:00–24:00 (50%)
	Schedule (HVAC OFF)	00:00–24:00 (0%)	00:00–24:00 (0%)
Equipment	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–08:00 (25%) 08:00–12:00 (100%) 12:00–13:00 (80%) 13:00–21:00 (100%) 21:00–23:00 (50%) 23:00–24:00 (25%)	00:00–24:00 (25%)
	Schedule (HVAC OFF)	00:00–24:00 (25%)	00:00–24:00 (25%)

Table A.5
Heat source schedules in building N.

Heat source		Value	
		Weekday	Holiday
Occupants	Maximum	0.1 person/m^2	
	Schedule (HVAC ON)	00:00–07:00 (0%) 07:00–08:00 (16%) 08:00–12:00 (80%) 12:00–13:00 (48%) 13:00–18:00 (80%) 18:00–19:00 (40%) 19:00–20:00 (24%) 20:00–22:00 (16%) 22:00–24:00 (0%)	00:00–24:00 (16%)
	Schedule (HVAC OFF)	00:00–24:00 (0%)	00:00–24:00 (0%)
Lighting	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–07:00 (0%) 07:00–08:00 (50%) 08:00–12:00 (100%) 12:00–13:00 (50%) 13:00–20:00 (100%) 20:00–21:00 (80%) 21:00–22:00 (50%) 22:00–24:00 (0%)	00:00–24:00 (50%)
	Schedule (HVAC OFF)	00:00–24:00 (0%)	00:00–24:00 (0%)
Equipment	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–08:00 (25%) 08:00–12:00 (100%) 12:00–13:00 (80%) 13:00–19:00 (100%) 19:00–20:00 (50%) 20:00–24:00 (25%)	00:00–24:00 (25%)
	Schedule (HVAC OFF)	00:00–24:00 (25%)	00:00–24:00 (25%)

and forecasting in turn, and an integrated approach that simultaneously performed them. We examined the performance of these two approach under LSTM and GRU structure, the LSTM outperformed GRU in terms of MAE and MRE for 5.0% and 4.1% in average. Moreover, the integrated approach generally showed higher sub-load forecasting performance than the two-step approach (D2F and F2D), with up to 34.9% lower MAE and 18.5% higher F-score. However, in buildings that involve sudden switches in load patterns, the performance deteriorated significantly. In the two-step approach that sequentially performs disaggregation and forecasting, the forecasting step's performance is an issue. Therefore, more advanced models are available for time-series forecasting, such as sequence to sequence models and transformer models. In addition, frequency analysis can be introduced as a means of understanding the characteristics of load waveforms. One of our future works will incorporate these technologies and improve NITLDF approaches.

CRediT authorship contribution statement

Naoya Kaneko: Writing – original draft, Software, Methodology, Conceptualization, Data curation, Formal analysis. **Kazuki Okazawa:** Writing – review & editing, Data curation. **Dafang Zhao:** Writing – review & editing, Project administration. **Hiroki Nishikawa:** Writing – review & editing, Project administration. **Ittetsu Taniguchi:** Project administration, Supervision, Writing – review & editing. **Hiroyuki Murayama:** Data curation, Resources, Writing – review & editing. **Yoshinori Yura:** Data curation, Resources, Writing – review & editing. **Masakazu Okamoto:** Data curation, Resources, Writing – review & editing. **Francky Catthoor:** Writing – review & editing. **Takao Onoye:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP22H0 3697 and DAIKIN Industries, Ltd.

Appendix

The appendix describes the weather in Osaka and Tokyo, which was not explained in detail due to page limitations, as well as the HVAC and occupant schedules for target buildings in Osaka.

Figs. A.1 and A.2 show the outdoor temperatures in Osaka and Tokyo from June 1 to September 30 in 2018. The vertical axis represents the date, and the horizontal axis represents the time. The redder the color is, the higher the outdoor temperature, and the bluer the color is, the lower the outdoor temperature. In both regions, there are days when the outdoor daytime temperature exceeds 30 °C from early July to late August. On rainy and cloudy days, the outdoor temperature

Table A.6
Heat source schedules in building O.

Heat source		Value	
		Weekday	Holiday
Occupants	Maximum	0.1 person/m^2	
	Schedule (HVAC ON)	00:00–06:00 (0%) 06:00–08:00 (16%) 08:00–12:00 (80%) 12:00–13:00 (48%) 13:00–18:00 (80%) 18:00–19:00 (40%) 19:00–20:00 (24%) 20:00–23:00 (16%) 23:00–24:00 (0%)	00:00–24:00 (16%)
	Schedule (HVAC OFF)	00:00–24:00 (0%)	00:00–24:00 (0%)
Lighting	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–06:00 (0%) 06:00–08:00 (50%) 08:00–12:00 (100%) 12:00–13:00 (50%) 13:00–20:00 (100%) 20:00–21:00 (80%) 21:00–23:00 (50%) 23:00–24:00 (0%)	00:00–24:00 (50%)
	Schedule (HVAC OFF)	00:00–24:00 (0%)	00:00–24:00 (0%)
Equipment	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–06:00 (25%) 08:00–12:00 (100%) 12:00–13:00 (80%) 13:00–19:00 (100%) 19:00–21:00 (50%) 21:00–24:00 (25%)	00:00–24:00 (25%)
	Schedule (HVAC OFF)	00:00–24:00 (25%)	00:00–24:00 (25%)

Table A.7
Heat source schedules in building R.

Heat source		Value	
		Weekday	Holiday
Occupants	Maximum	0.1 person/m^2	
	Schedule (HVAC ON)	00:00–07:00 (0%) 07:00–08:00 (16%) 08:00–12:00 (80%) 12:00–13:00 (48%) 13:00–18:00 (80%) 18:00–19:00 (40%) 19:00–20:00 (24%) 20:00–24:00 (0%)	00:00–24:00 (16%)
	Schedule (HVAC OFF)	00:00–24:00 (0%)	00:00–24:00 (0%)
Lighting	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–07:00 (0%) 07:00–08:00 (50%) 08:00–12:00 (100%) 12:00–13:00 (50%) 13:00–20:00 (100%) 20:00–24:00 (%)	00:00–24:00 (50%)
	Schedule (HVAC OFF)	00:00–24:00 (%)	00:00–24:00 (0%)
Equipment	Maximum	12 W/m^2	
	Schedule (HVAC ON)	00:00–08:00 (25%) 08:00–12:00 (100%) 12:00–13:00 (80%) 13:00–19:00 (100%) 19:00–20:00 (50%) 20:00–24:00 (25%)	00:00–24:00 (25%)
	Schedule (HVAC OFF)	00:00–24:00 (25%)	00:00–24:00 (25%)

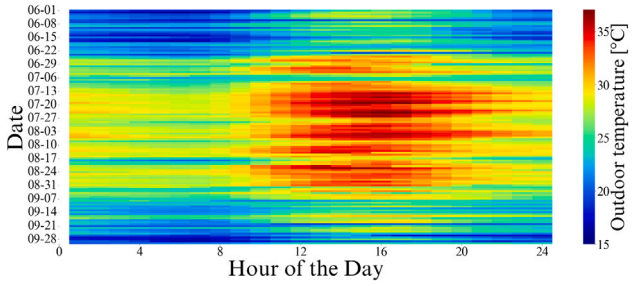


Fig. A.1. Outdoor temperature in Osaka.

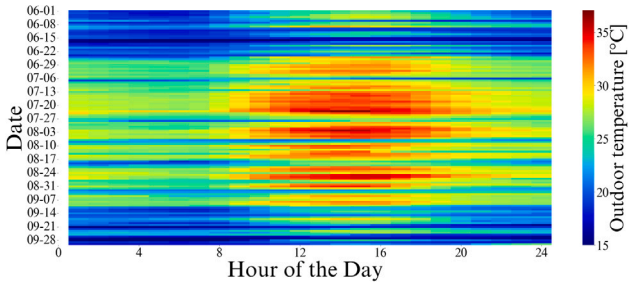
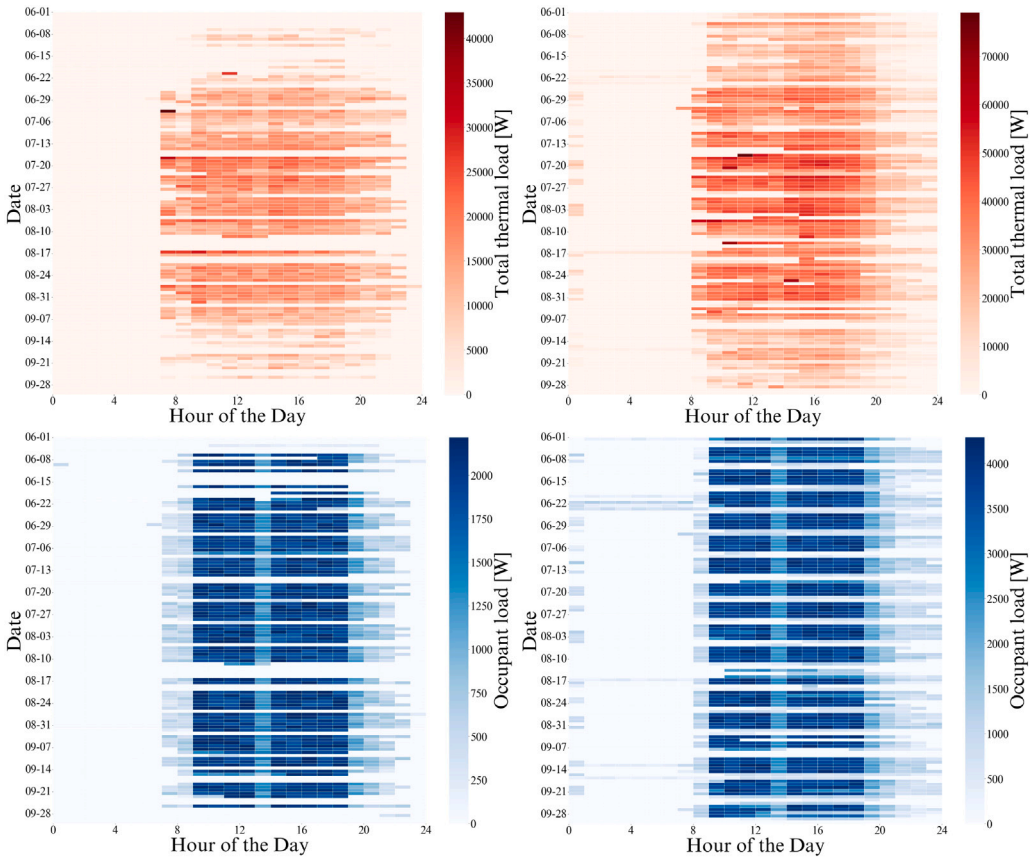


Fig. A.2. Outdoor temperature in Tokyo.

is low throughout the day. Looking at the outdoor temperatures in Osaka (Fig. A.1), they rise significantly during the day and drop in the evening. On the other hand, in Tokyo (Fig. A.2), the outdoor temperature is low throughout the day, and there are many rainy and cloudy days, especially in June and September, where there is no noticeable increase. However, since the temperature and humidity are given for the integrated and two-step approaches, these are taken into account in the inference model. The difference in weather between Osaka and Tokyo does not have a large effect on the superiority of the NITLDF approaches.

Fig. A.3 shows the total thermal and occupant loads of buildings A and Y in Osaka. Similar to Fig. 5 in Section 4.3, the vertical axis is the date, and the horizontal axis is the time. The total thermal load is shown as a red heat map, and the occupant load is shown as a blue heat map. In building A, HVAC basically operates between 7:00 am and 11:00 pm, and the occupants are also present during these hours. There are almost no days of exceptional behavior. In building Y, the HVAC operates from 7:00 am to 1:00 am, and people are in the room. However, building Y exhibits exceptional behavior in some periods, such as around June 22 and September 14. Although this behavior may have increased the error of the NITDF approach, the integrated approach still shows a better sub-load forecasting performance than the two-step approach.

Tables A.4–A.7 describe the heat source schedules in building Y, O, N, R, respectively.



(a) Building A (b) Building Y

Fig. A.3. Total thermal load and occupant load of buildings A and Y

References

[1] Europe Comission. In focus: Energy efficiency in buildings. 2023, https://energyplus.net/. [Accessed on 24 July 2023].
[2] IEA. 2020 Global status report for buildings and construction. 2020.
[3] Fischer Corinna. Feedback on household electricity consumption: A tool for saving energy? Energy Effic 2008;1:79–104.
[4] Ridi Antonio, et al. A survey on intrusive load monitoring for appliance recognition. In: Proc. of 2014 22nd international conference on pattern recognition. 2014, p. 3702–7.
[5] Hart George W. Nonintrusive appliance load monitoring. Proc IEEE 1992;80:1870–91.
[6] Xiao Ziwei, et al. Cooling load disaggregation using a NILM method based on random forest for smart buildings. Sustainable Cities Soc 2021;74:103202.
[7] Enríquez Ricardo, et al. Towards non-intrusive thermal load monitoring of buildings: BES calibration. Appl Energy 2017;191:44–55.
[8] Erickson Varick L, et al. Occupancy based demand response HVAC control strategy. In: Proc. of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building. 2010, p. 7–12.
[9] Beltran Alex, et al. Optimal HVAC building control with occupancy prediction. In: Proc. of the 1st ACM conference on embedded systems for energy-efficient buildings. 2014, p. 168–71.
[10] Kato Kenshiro, et al. Scheduling for multiple HVAC systems with electrical power allocation. In: Proc. of the 2022 IEEE 11th global conference on consumer electronics. 2022, p. 226–7.
[11] Kaneko Naoya, et al. RNN-based non-intrusive thermal load disaggregation and forecasting for HVAC systems. In: Proc. of the 10th ACM international conference on systems for energy-efficient buildings, cities, and transportation. 2023, p. 292–3.
[12] Makonin Stephen, et al. Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring. IEEE Trans Smart Grid 2016;7:2575–85.
[13] Kong Weicong, et al. A hierarchical hidden Markov model framework for home appliance modeling. IEEE Trans Smart Grid 2018;9:3079–90.
[14] Bonfigli Roberto, et al. Non-intrusive load monitoring by using active and reactive power in additive factorial hidden Markov models. Appl Energy 2017;208:1590–607.
[15] He Kanghang, et al. Non-intrusive load disaggregation using graph signal processing. IEEE Trans Smart Grid 2018;9:1739–47.
[16] Zhao Bochao, et al. Improving event-based non-intrusive load monitoring using graph signal processing. IEEE Access 2018;6:53944–59.
[17] Jia Dongning, et al. Non-intrusive load identification using reconstructed voltage–current images. IEEE Access 2021;9:77349–58.
[18] Ahmed cShamim, et al. Edge computed NILM: A phone-based implementation using MobileNet compressed by tensorflow lite. In: Proc. of the 5th international workshop on non-intrusive load monitoring. 2020, p. 44–8.
[19] Todic Tamara, et al. An active learning framework for the low-frequency non-intrusive load monitoring problem. Appl Energy 2023;341:121078.
[20] Jiang Jie, et al. Deep learning-based energy disaggregation and On/Off detection of household appliances. ACM Trans Knowl Discov Data 2021;15:50.
[21] Shastri Hetvi, et al. Neural network approaches and dataset parser for NILM toolkit. In: Proc. of the 8th ACM international conference on systems for energy-efficient buildings, cities, and transportation. 2021, p. 172–5.
[22] Yadav Akhilesh, et al. NILM based energy disaggregation algorithm for dairy farms. In: Proc. of the 5th international workshop on non-intrusive load monitoring. 2020, p. 16–9.
[23] Krystalakos Odysseas, et al. Sliding window approach for online energy disaggregation using artificial neural networks. In: Proc. of the 10th Hellenic conference on artificial intelligence. 2018, p. 1–6.
[24] Kelly Jack, et al. Neural NILM: Deep neural networks applied to energy disaggregation. In: Proc. of the 2nd ACM international conference on embedded systems for energy-efficient built environments. 2015, p. 55–64.
[25] Xiao Ziwei, et al. Comparison between artificial neural network and random forest for effective disaggregation of building cooling load. Case Stud Therm Eng 2021;28:101589.
[26] Okazawa Kazuki, et al. Exploring of recursive model-based non-intrusive thermal load monitoring for building cooling load. In: Proc. of the 14th ACM international conference on future energy systems. 2023, p. 120–4.
[27] Fang Tingting, et al. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Appl Energy 2016;179:544–52.
[28] Zhao Jing, et al. A hybrid method of dynamic cooling and heating load forecasting for office buildings based on artificial intelligence and regression analysis. Energy Build 2018;174:293–308.
[29] Jie Zhang, et al. Thermal load forecasting based on PSO-SVR. In: Proc. of the 2018 IEEE 4th international conference on computer and communications. 2018, p. 2676–80.

[30] Fan Chengliang, et al. Development of a cooling load prediction model for air-conditioning system control of office buildings. Int J Low-Carbon Technol 2019;14:70–5.
[31] Fan Cheng, et al. A short-term building cooling load prediction method using deep learning algorithms. Appl Energy 2017;195:222–33.
[32] Lin Xinyi, et al. Short-term forecast model of cooling load using load component disaggregation. Appl Therm Eng 2019;157:113630.
[33] Benalcazar Pablo, et al. Short-term heat load forecasting in district heating systems using artificial neural networks. IOP Conf Ser: Earth Environ Sci 2019;214:012023.
[34] Wang Lan, et al. Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach. Appl Energy 2018;228:1740–53.
[35] Leiprecht Stefan, et al. A comprehensive thermal load forecasting analysis based on machine learning algorithms. Energy Rep 2021;7:319–26.
[36] Geysen Davy, et al. Operational thermal load forecasting in district heating networks using machine learning and expert advice. Energy Build 2018;162:144–53.
[37] Meng Huixian, et al. Data center cooling load prediction and analysis based on weather data clustering. In: Proc. of the 2023 12th international conference on informatics, environment, energy and applications. 2023, p. 1–7.
[38] Kingma Diederik P, et al. Adam: A method for stochastic optimization. 2017, arXiv:1412.6980.
[39] Open source software. 2023, https://energyplus.net/. [Accessed on 24 July 2023].
[40] WEBPRO. 2024, https://building.app.lowenergy.jp/. [Accessed On 18 March 2024].
[41] Bergstra James, et al. Algorithms for hyper-parameter optimization. In: Proc. of the 24th international conference on neural information processing systems. 2011, p. 2546–54.
[42] TensorFlow v2.5.0. 2024, https://www.tensorflow.org/versions/r2.5/api_docs. [Accessed On 18 March 2024].
[43] CUDA toolkit 11.2. 2024, https://developer.nvidia.com/cuda-11.2.0-download-archive. [Accessed On 18 March 2024].
[44] Hyndman Rob J, et al. Another look at measures of forecast accuracy. Int J Forecast 2006;22:679–88.
[45] Green Christy, et al. Non-intrusive load monitoring of water heaters using low-resolution data. In: Proc. of the 5th international workshop on non-intrusive load monitoring. 2020, p. 54–8.