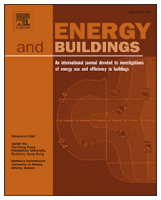
A study of deep learning-based multi-horizon building energy forecasting

Energy & Buildings 303 (2024) 113810

Contents lists available at ScienceDirect

Energy & Buildings

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



A study of deep learning-based multi-horizon building energy forecasting

Zhongjun Ni [a,*], Chi Zhang [b], Magnus Karlsson [a], Shaofang Gong [a]

[a] Department of Science and Technology, Linköping University, Campus Norrköping, Bredgatan 33, Norrköping, 60174, Sweden
[b] Department of Computer Science and Engineering, University of Gothenburg, Universitetsplatsen 1, Gothenburg, 40530, Sweden

ARTICLE INFO

Keywords:
Building energy forecasting
Probabilistic forecast
Deep learning
Quantile regression
Prediction interval

ABSTRACT

Building energy forecasting facilitates optimizing daily operation scheduling and long-term energy planning. Many studies have demonstrated the potential of data-driven approaches in producing point forecasts of energy use. Despite this, little work has been undertaken to understand uncertainty in energy forecasts. However, many decision-making scenarios require information from a full conditional distribution of forecasts. In addition, recent advances in deep learning have not been fully exploited for building energy forecasting. Motivated by these research gaps, this study contributes in two aspects. First, this study has adapted and applied state-of-the-art deep learning architectures to address the problem of multi-horizon building energy forecasting. Eight different methods, including seven deep learning-based ones, were investigated to develop models to perform both point and probabilistic forecasts. Second, a comprehensive case study was conducted in two public historic buildings with different operating modes, namely the City Museum and the City Theatre, in Norrköping, Sweden. The performance of the developed models was evaluated, and the predictability of different scenarios of energy consumption was studied. The results show that incorporating future information on exogenous factors that determine energy use is critical for making accurate multi-horizon predictions. Furthermore, changes in the operating mode and activities held in a building bring more uncertainty in energy use and deteriorate the prediction accuracy of models. The temporal fusion transformer (TFT) model exhibited strong competitiveness in performing both point and probabilistic forecasts. As assessed by the coefficient of variance of the root mean square error (CV-RMSE), the TFT model outperformed other models in making point forecasts of both types of energy use of the City Museum (CV-RMSE 29.7% for electricity consumption and CV-RMSE 8.7% for heating load). When making probabilistic predictions, the TFT model performed best to capture the central tendency and upper distribution of heating load of the City Museum as well as both types of energy use of the City Theatre. The predictive models developed in this study can be integrated into digital twin models of buildings to discover areas where energy use can be reduced, optimize building operations, and improve overall sustainability and efficiency.

1. Introduction

Building energy forecasting is essential for energy efficiency, lowering energy use and greenhouse gas emissions. For example, short-term energy forecasting (for the next several hours or a few days) gives valuable references to facility managers. Maintainers can thus optimize daily operation scheduling [1] and design cost-effective energy-saving methods [2] while still ensuring the functions of a building. Medium- and long-term forecasting are useful for renovating buildings, e.g., examining a design during an early phase [3], as well as government policy-making for energy planning [4]. In addition to the demand side, building energy forecasting is also critical to the supply side. For instance, because of the rising energy demand (in 2021, 30% of the world's total energy use was attributed to the operation of buildings [5]), energy companies must manage energy production more efficiently [6]. Accurate demand forecasting enables these companies to obtain sustainable production plans. Energy forecasting models can further be integrated into a digital twin model of the energy system of a building or a digital twin model of the entire building. Creating a digital twin of a building can combine information and communication technologies, such as Internet of Things, cloud computing, and ontology [7], to model its critical functional areas. By integrating predictive models, the digital twin can simulate energy use in different operating modes and conditions. This can be used to optimize building operations

* Corresponding author.
E-mail addresses: zhongjun.ni@liu.se (Z. Ni), chi.zhang@gu.se (C. Zhang), magnus.b.karlsson@liu.se (M. Karlsson), shaofang.gong@liu.se (S. Gong).

https://doi.org/10.1016/j.enbuild.2023.113810
Received 13 March 2023; Received in revised form 15 September 2023; Accepted 27 November 2023
Available online 1 December 2023

and ultimately result in cost savings, improved human comfort, and a more sustainable built environment.

Accurate and reliable building energy forecasting also has several challenges. On one hand, energy systems of a building or a cluster of buildings can be complex and dynamic due to trend, seasonality and irregularity [4]. On the other hand, exogenous factors, such as outdoor climate, thermal characteristics of a building envelope, and occupants' energy use habits, can affect the energy consumption of a building [8,9]. For example, thermal characteristics of a building envelope, e.g., insulation levels, determine the amount of heat gained or lost through the envelope, which affects the energy required to maintain a comfortable indoor temperature.

Methods for building energy forecasting can be broadly classified into three categories: physical, data-driven, and hybrid approaches that integrate physical and data-driven approaches. Physical approaches adopt thermodynamic rules for precise energy modeling and analysis. They often rely on building energy simulation software, e.g., EnergyPlus [10], to calculate the energy consumption of a building based on characteristics of the building structure, design specifications of heating, ventilation, and air-conditioning (HVAC) systems and lighting systems, operation schedules, as well as indoor or outdoor climate [8]. Physical approaches have benefits in interpreting results and are excellent at simulating energy consumption during the design phase [11]. However, the dependency on building characteristics limits the application scenario of physical approaches since many historic buildings lack such data, and it is labor-intensive to obtain these data or even not allowed to obtain them due to regulations concerning preservation. In contrast to physical approaches, detailed physical characteristics of building structures are not necessary for data-driven approaches [12]. Data-driven approaches leverage historical energy consumption and other data to develop predictive models. These data are becoming more readily available with the digital transformation in buildings, for example, deploying monitoring systems [13] through integration of Internet of Things devices and cloud computing [14,15]. Therefore, it is necessary to fully utilize the accumulated data to create advanced data-driven energy forecasting models for optimizing building operations.

Deep learning methods have emerged among data-driven approaches in recent years due to their enhanced ability to address massive volumes of data, extract features, and model nonlinear processes [1]. Open-source frameworks like PyTorch [16] have also dramatically simplified network implementation and model training. There have been many studies using deep learning techniques, such as recurrent neural network (RNN) and its variants [2,17–19], temporal convolutional network (TCN) [6], and attention mechanism-based network [20,21], for one-step-ahead or multi-horizon building energy forecasting. Nevertheless, most studies focused on point forecast, that is, forecasting the conditional mean or median of future values of the target energy consumption. Only limited cases [22] studied probabilistic forecast.

Recently, probabilistic forecasting has grown in popularity because it can extract deeper information from historical data and better capture future uncertainty [23]. Many decision-making scenarios require more information from a probabilistic forecasting model that returns the full conditional distribution rather than a point forecasting model that merely forecasts the conditional mean [24]. In addition, limited work [19] adopted the operational data, such as opening hours and occupancy, of buildings as features for making multi-horizon predictions. On the one hand, this may be related to the building type of the case study. On the other hand, data collection is also tricky. Nevertheless, scheduling, such as opening hours and activity arrangements, is critical for public buildings because it determines public access and energy consumption. Moreover, in terms of predicted energy use, most deep learning-based studies only predict total energy consumption or one particular type. Further comparisons of the predictability of various types of energy use are needed. Furthermore, previous research mainly selected three types of buildings for case studies: residential, office, and educational. Limited work chose public historic buildings as case studies. However, energy forecasting is equally important for these buildings to optimize daily operations while maintaining functionality and preserving heritage values [25].

This study aims to use state-of-the-art deep learning architectures to address the problem of multi-horizon building energy forecasting. In addition to performing point forecasts, we involve probabilistic forecasts to measure and interpret the uncertainties in forecasts. Moreover, we propose to incorporate future information on exogenous factors, especially the operational data of a building, to improve the accuracy of multi-horizon forecasting. The main contributions of the paper are:

- In addition to linear regression, we adapted and applied seven deep learning architectures, including hierarchical interpolation for time series forecasting (N-HiTS), TCN, Transformer (TF), NLinear, long short-term memory (LSTM), gated recurrent unit (GRU), and temporal fusion transformer (TFT), to the field of multi-horizon building energy forecasting. Among them, N-HiTS and NLinear were improved to support performing probabilistic forecasts based on quantile regression.
- A comprehensive case study was conducted in two public historic buildings with different operating modes to evaluate the performance of developed models. The obtained results provide insights for subsequent studies of public historic buildings with similar operating modes. The findings indicate that involving strong influencing factors makes energy consumption more predictable. Moreover, incorporating future information on exogenous factors that determine energy use is critical for enhancing multi-horizon building energy forecasting. The TFT model shows competitiveness in both point and probabilistic forecasts. Furthermore, involving building operational data, such as opening hours, can improve the prediction accuracy of models.

The remainder of this paper is organized as follows. After discussing related work in Section 2, the detailed methodology for conducting this study is described in Section 3. Then, a case study, including a detailed description of the dataset and experimental setup, is given in Section 4. After that, Section 5 presents and discusses the obtained results. The last section concludes the paper.

2. Related work

Energy consumption of a building is a form of time series, a sequence of values recorded over time (typically at constant intervals) and organized chronologically [26]. Therefore, this section starts with foundational deep learning methods and recent architectures for time series forecasting. Then, studies on deep learning-based building energy forecasting are presented.

2.1. Foundational deep learning methods for time series forecasting

Fully connected networks, like artificial neural networks (ANNs) and deep neural networks (DNNs), have limitations in extracting temporal dependencies of a time series. As a result, more specialized deep learning architectures, such as RNNs, convolutional neural networks (CNNs), and attention mechanism-based networks, began gaining prominence in time series forecasting. Vanilla RNN has a hidden state that serves as a concise summary of previous inputs in a sequence. The hidden state is recursively updated at each time step after processing new inputs. However, the vanishing and expanding gradient problems limit the learning ability of vanilla RNNs. Two variants of RNNs, namely LSTM [27] and GRU [28], address the gradient problems. LSTM employs three gates to retain long-standing essential information while discarding nonessential information. GRU simplifies LSTM and is computationally faster than LSTM since it only has two gates.

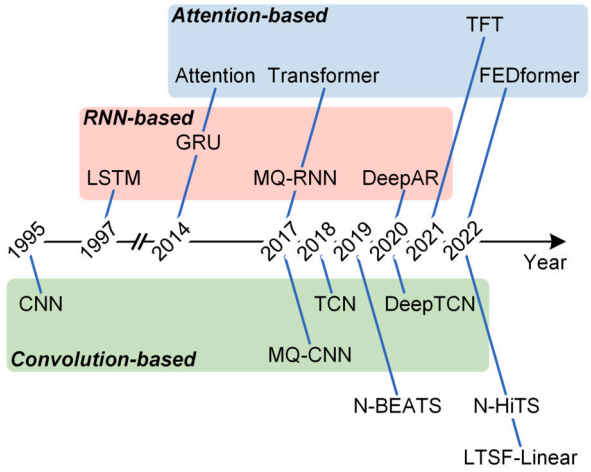


Fig. 1. Selected key research milestones within deep learning architectures for time series forecasting.

Recent research suggests that specific CNNs can achieve state-of-the-art accuracy in various application domains of sequence modeling, such as audio synthesis and autonomous driving [29,30]. CNNs are built with convolution, pooling, and fully connected layers [31]. The convolution layers learn features from input data by filters with a predefined size. Then, pooling layers process the convolution results by average or maximum pooling. Finally, the flattened features produced by the convolutional and pooling layers provide the input for fully connected layers to perform the forecasting. Parallelism is an essential advantage of CNNs, as convolutions can be performed in parallel because the same filter is used in each layer.

There is a growing interest in understanding how and why a model makes a particular prediction. Based on a better understanding of temporal dynamics and the rationale behind a forecast, decision-makers can improve their actions further. Attention mechanism [32] has become an intrinsic part of sequence modeling in various tasks. The attention mechanism is a key-value lookup technique depending on a provided query. For time series modeling, the output of the attention layer can be interpreted as a weighted average across temporal features. An analysis of attention weights can thus determine the relative importance of features at each time step [33].

2.2. Recent deep learning architectures for time series forecasting

Recent competitive deep learning architectures for time series forecasting, as summarized in Fig. 1, are mainly built on previous advances. RNN (and its variants)-based architectures include MQ-RNN [24] and DeepAR [34]. Both MQ-RNN and DeepAR aim to handle the challenge of large-scale time series forecasting. Rather than predicting each time series individually, they learn a global model from historical data for all time series in a dataset. Meanwhile, they both employ an LSTM for encoding all historical information into hidden states. Unlike DeepAR uses an LSTM as the recursive decoder when generating forecasts, MQ-RNN adopts two multilayer perceptron branches. In addition, while DeepAR is trained using maximum likelihood and teacher forcing (feeding ground truth recursively in training) [34], MQ-RNN uses a more efficient training technique and generates quantile forecasts directly [24].

Convolution-based architectures include MQ-CNN [24], TCN [35], and DeepTCN [23]. MQ-CNN has a similar architecture with MQ-RNN by just replacing the encoder with a stack of dilated causal 1D convolution layers [36]. Bai et al. [35] built TCN by condensing dilated and causal convolution. In addition, TCN employs a generic residual module [37] for stabilization. Based on TCN, Chen et al. [23] proposed DeepTCN as a non-autoregressive probabilistic forecasting framework for large-scale related time series. Like MQ-CNN, DeepTCN follows an encoder-decoder design.

A representative of attention-based architectures is Transformer (TF) [38]. While the attention mechanisms are used together with RNN in many cases, Vaswani et al. [38] proposed the TF, which relies entirely on attention mechanisms to draw global dependencies between input and output. Self-attention enables linking different positions in the sequence, while multi-head attention enables the model to attend to information from distinct representation subspaces at different points [38]. Based on TF, FEDformer [39] further incorporates classical time series analysis techniques like frequency processing through Fourier transformation. Lim et al. [40] created TFT using canonical components, such as gated residual network, LSTM, and multi-head attention. Gating mechanisms enable TFT to skip over any unused components of the architecture, providing adaptive depth and network complexity to accommodate a wide range of datasets and scenarios. Some studies questioned the validity of Transformer-based solutions for long-term time series forecasting tasks. For example, Zeng et al. [41] proposed a simple direct multi-step model through a linear temporal layer named LTSF-Linear. In many cases, it outperforms FEDformer [39] on multi-horizon forecasting of multivariate time series.

Some architectures are based on a deep stack of fully connected layers, such as N-BEATS [49] and its improved version N-HiTS [50]. Oreshkin et al. [49] proposed N-BEATS, an architecture built on backward and forward residual links and a very deep stack of fully connected layers. Challu et al. [50] enhanced the N-BEATS architecture by improving its input decomposition through multi-rate data sampling and its output synthesizer through multi-scale interpolation. N-HiTS adds subsampling layers before fully-connected blocks in N-BEATS. This modification dramatically decreased the required computation and memory usage while retaining the capacity to capture long-term dependencies.

2.3. Deep learning-based building energy forecasting

Many studies have employed deep learning to predict the energy consumption of buildings. As shown in Table 1, the architectures used include RNN and its variants [2,17–19,47,48], convolution-based [6,48], attention-based [20,21], deep belief network (DBN) [43,45], and hybrid models [9,42,44] that combine multiple architectures. Both one-step ahead [3,17,18,42–45] and multi-horizon forecasting [2,6,9,19–21,43,48] were investigated. However, most studies only performed point forecasts. Furthermore, there is a need for studies that apply cutting-edge competitive architectures, such as TFT and N-HiTS, to building energy forecasting and a thorough comparison of these most recent and older architectures.

Most studies used energy data from residential [3,17,21,46], educational [2,9,20,48], and office buildings [18,19,22,44,47]. Models were mainly developed using time granularity of hourly and sub-hourly data, and only a few [44] used daily data. The time span of data in most datasets was less than three years, and only a few studies employed datasets longer than three years [20,44]. For the type of predicted energy use, most studies only predicted one kind or total energy consumption. Few studies used public historic buildings for case studies. However, energy forecasting is also critical for these buildings to optimize daily operations while maintaining their functionalities and preserving heritage values.

Outdoor weather and historical energy consumption were the most commonly used features for predicting energy consumption, regardless of building type. In contrast, data about occupants' behavior [19] was rarely utilized. This preference for features is primarily due to more readily available outdoor weather data. Outdoor weather, for example, can be gathered from many public databases. However, privacy policies make obtaining features such as occupants' behavior challenging [11]. Other used features include indoor environmental parameters, such as room temperature and relative humidity, as well as temporal features, e.g., the type of day (weekday, weekend, or holiday) and the type of hour (daytime or nighttime) [8]. Building operational data like opening hours and scheduling of HVAC systems were rarely used. Nevertheless,

Table 1
A summary of related work that employed deep learning to predict energy consumption of buildings. For comparison with previous studies, this work is also listed. Missing information is represented by a dash -.

Study	Dataset				Deep learning architecture	Forecast setup	
	Building type	Predicted energy	Time span	Time granularity		Horizon	Type
[17]	residential	electricity	3 months	30 minutes	LSTM	1	point
[20]	educational	electricity/ cooling/heating	36 months	1 hour	TF	24	point
[3]	residential	total	12 months	1 year	DNN	1	point
[9]	educational	electricity	12 months	15 minutes	CNN/LSTM	4	point
[2]	educational	cooling	12 months	30 minutes	RNN/ LSTM/GRU	48	point
[21]	residential	electricity	11 months	1 hour	TFT	1/24	point
[42]	industrial	electricity	17 months	1 hour	TCN	1	point
[6]	infrastructure	electricity	15 months	1 hour	TCN	48	point
[43]	non-residential	electricity	4 months	1 hour	DBN	1	point
[44]	office	electricity	36 months	1 day	LSTM/TF	1	point
[19]	office	electricity	2 months	1 hour	LSTM	24	point
[18]	office	cooling	5 months	1 hour	LSTM/GRU	1	point
[45]	-	total	12 months	30 minutes	DBN	1	point
[22]	office	hot water	3 months	1 hour	ANN	1	probabilistic
[46]	residential	electricity	4.5 months	10 minutes	CNN/GRU	1	point
[47]	office	electricity	12 months	1 hour	LSTM/GRU	1	point
[48]	educational/ commercial	electricity	12 months	1 hour	CNN/RNN	1/24	point
This work	public historic	electricity/ heating	48 months	1 hour	N-HiTS/TF/ TCN/NLinear/ LSTM/GRU/ TFT	24	point/ probabilistic

involving available operational data can potentially increase the prediction accuracy of models, especially for public buildings where energy consumption is highly correlated with held activities.

2.4. Innovation of this study

In order to contribute to addressing aforementioned research gaps, this study adapted and applied state-of-the-art deep learning architectures to multi-horizon building energy forecasting. While previous studies mainly focused on point forecasts, we also investigated probabilistic forecasts. Quantile regression was adopted to achieve a complete understanding of the distribution of energy consumption. A comprehensive case study was conducted in two public historic buildings to compare the performance of various models. Public historic building is a rare building type in previous research. Regarding features, we proposed incorporating future information on factors that determine energy use, especially data related to building operations. Involving operational data like scheduling of activities has the potential to improve prediction accuracy of models since activities held in public historic buildings could considerably affect their energy consumption. Furthermore, the predictability of different types of energy consumption inside the same building and between buildings with different operating modes was studied. These efforts could bring inspiration to predicting energy use and optimizing energy efficiency of public buildings, especially historic buildings.

3. Methodology

This section begins by formulating the problem of multi-horizon building energy forecasting. Then, the encoder-decoder architecture is described. After that, seven deep learning architectures for comparison are given. Finally, the loss function for model training and metrics for evaluating model performance are introduced.

3.1. Problem formulation

This study considers the problem of multi-horizon forecasting for energy consumption of buildings. We denote a specific type of energy use, i.e., the target variable, as a non-negative real variable $y \in \mathbb{R}_+$. Predictor variables that might affect the energy use are divided into two parts: observable in the past (i.e., before (including) a forecast origin, see Fig. 2) and observable in the future (i.e., after the forecast origin). The former is denoted as a real row vector $\mathbf{x_b} \in \mathbb{R}^k$ while the latter is denoted as a real row vector $\mathbf{x_f} \in \mathbb{R}^m$. All target and predictor variables are assumed to be observed across time at constant intervals and organized chronologically. At time t, the observed value of the target variable is denoted as y_t. Similarly, observed values of predictor variables are denoted as $\mathbf{x_b}_t = [x_{b1,t}, x_{b2,t}, ..., x_{bk,t}]$ and $\mathbf{x_f}_t = [x_{f1,t}, x_{f2,t}, ..., x_{fm,t}]$, respectively.

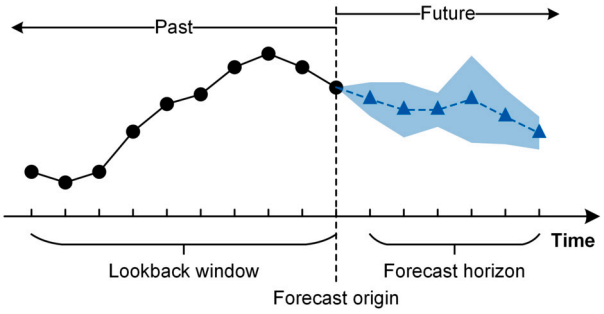


Fig. 2. An illustration of multi-horizon building energy forecasting. In this example, black dots are observed values of a specific type of energy consumption over a lookback window ($size = 10$) in the past. Blue triangles are predicted values (conditional mean or median) of the energy consumption over a forecast horizon ($size = 6$) in the future. The shadow area represents a particular prediction interval. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Then, a point energy forecasting model takes the form

$$\hat{y}_{t+1:t+h} = f_\theta(y_{t-w+1:t}, \mathbf{x_b}_{t-w+1:t}, \mathbf{x_f}_{t+1:t+h}), \quad (1)$$

where $\hat{y}_{t+1:t+h} = [\hat{y}_{t+1}, \hat{y}_{t+2}, ..., \hat{y}_{t+h}] \in \mathbb{R}_+^h$ are model forecasts for mean values of the target variable over a forecast horizon h, $y_{t-w+1:t} = [y_{t-w+1}, y_{t-w+2}, ..., y_t] \in \mathbb{R}_+^w$ as well as $\mathbf{x_b}_{t-w+1:t} = \{\mathbf{x_b}_{t-w+1}, \mathbf{x_b}_{t-w+2}, ..., \mathbf{x_b}_t\}$ are observations of the target and predictor variables over a lookback window w, $\mathbf{x_f}_{t+1:t+h} = \{\mathbf{x_f}_{t+1}, \mathbf{x_f}_{t+2}, ..., \mathbf{x_f}_{t+h}\}$ are observations of predictor variables over the forecast horizon h, and $f_\theta(.)$ is the prediction function learnt by the model.

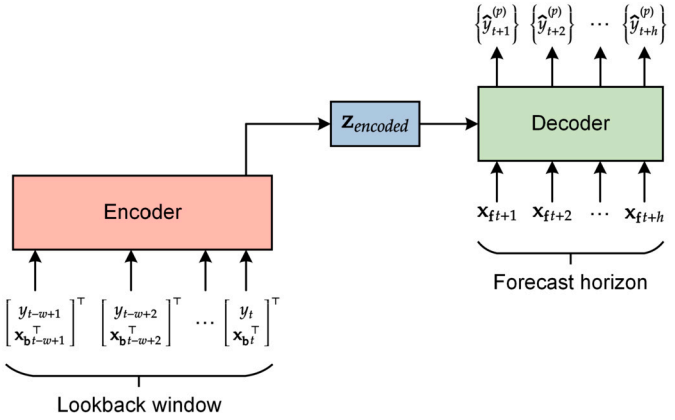


Fig. 3. A high-level illustration of the encoder-decoder architecture. The encoder takes observations of a target variable y and predictor variables $\mathbf{x}_\mathbf{b}$ over the loopback window as input. At each time step in the lookback window, the observations of target and predictor variables are concatenated. The encoder outputs a representation of the information in $\mathbf{z}_{encoded}$. The decoder takes the representation as well as observations of predictor variables $\mathbf{x}_\mathbf{f}$ over the forecast horizon as input and outputs the quantile forecasts. At each time step during the forecast horizon, the decoder outputs a set of predetermined quantile forecasts.

For developing probabilistic forecasting models, we do not assume that energy consumption follows some distributions but develop models that generate interested quantiles directly. Quantile forecasts are performed through quantile regression [51]. The pth quantile denotes the value where the cumulative distribution function crosses p [52]. Thus, quantiles can specify any position of a distribution.

Given a predetermined set of quantiles $\mathcal{Q} \subset (0,1)$, a quantile energy forecasting model takes the form

$$\hat{y}^{(p)}_{t+1:t+h} = g_\theta(y_{t-w+1:t}, \mathbf{x}_{\mathbf{b}t-w+1:t}, \mathbf{x}_{\mathbf{f}t+1:t+h}), \tag{2}$$

where p is an element of the set $\mathcal{Q}$, $\hat{y}^{(p)}_{t+1:t+h} = \left[\hat{y}^{(p)}_{t+1}, \hat{y}^{(p)}_{t+2}, ..., \hat{y}^{(p)}_{t+h}\right] \in \mathbb{R}^h_+$ are the model forecasts for the pth quantile of the target variable over a predicting horizon h, $y_{t-w+1:t}$, $\mathbf{x}_{\mathbf{b}t-w+1:t}$ and $\mathbf{x}_{\mathbf{f}t+1:t+h}$ have the same definition as in the point forecasting model, and $g_\theta(.)$ is the prediction function learnt by the model.

3.2. The encoder-decoder architecture

Most competitive sequential transduction models employ an encoder-decoder architecture [38]. This design decouples handling inputs and generating outputs into two separate stages and works much better. The encoder converts an input sequence to a sequence of representations called hidden states. Given hidden states, the decoder generates an output sequence of target variables. Specific to the problem of building energy forecasting, Fig. 3 illustrates a design that supports incorporating past and future information for predicting multi-horizon energy consumption. The encoder takes observations of the target and predictor variables over the loopback window as input. The target and predictor variables are concatenated as input at each time step in the look-back window. Then, the encoder outputs a summary of past information. The decoder takes the summary and observations of predictor variables over the forecast horizon as input and outputs the quantile forecasts. At each time step during the forecast horizon, the decoder outputs a set of predetermined quantile forecasts. Point forecast adopts the same architecture, except it outputs only one predicted value (conditional mean) at each time step during the forecast horizon.

3.3. Deep learning architectures for comparison

In addition to linear regression (LR), seven deep learning architectures, namely N-HiTS [50], TCN [35], Transformer (TF) [38], NLinear [41], LSTM [27], GRU [28], and TFT [40], were investigated to develop predictive models and compare their performance in multi-horizon building energy forecasting. For simplicity, this paper does not give the detailed design of these architectures. The LR model makes predictions based on a linear relationship between the target variable and past and future values of some predictor variables. Among the seven deep learning architectures, N-HiTS, TCN, and TF only support incorporating past values of target and predictor variables for making predictions. Other four architectures, NLinear, LSTM, GRU, and TFT, support incorporating past values of target variables and past and future values of predictor variables. Moreover, architectures, such as N-HiTS and NLinear, were improved to support producing probabilistic forecasts based on quantile regression.

Ensemble methods are not used in this study since any deep learning algorithm can profit from model averaging at the expense of extra computation and memory [53].

3.4. Loss function and evaluation metrics

Point forecasting models were trained on a training set to minimize the total squared error, which leads to forecasts of the mean [54]. The training squared error for a set $S = \{(y_{t-w+1:t}, \mathbf{x}_{\mathbf{b}t-w+1:t}, \mathbf{x}_{\mathbf{f}t+1:t+h}, y_{t+1:t+h})\}_{t=w}^{n+w-1}$ is denoted by $L_s(\theta)$, and

$$L_s(\theta) = \sum_{t=w}^{n+w-1} \sum_{i=1}^{h} \left(\hat{y}_{t+i} - y_{t+i}\right)^2, \tag{3}$$

where n denotes the number of training samples and definitions of other variables are as in Eq. (1).

Probabilistic forecasting models, i.e., quantile forecasting models in this study, were trained to minimize the total quantile loss. As in studies [24,40], the pth quantile loss for one prediction at one time step is calculated as

$$\ell(\hat{y}, y, p) = (1-p)(\hat{y} - y)_+ + p(y - \hat{y})_+, \tag{4}$$

where $(.)_+ = max(0, .)$. Then, the training quantile loss for a set $S = \{(y_{t-w+1:t}, \mathbf{x}_{\mathbf{b}t-w+1:t}, \mathbf{x}_{\mathbf{f}t+1:t+h}, y_{t+1:t+h})\}_{t=w}^{n+w-1}$ is denoted by $L_q(\theta)$, and

$$L_q(\theta) = \sum_{t=w}^{n+w-1} \sum_{j=1}^{|\mathcal{Q}|} \sum_{i=1}^{h} \ell\left(\hat{y}^{(p_j)}_{t+i}, y_{t+i}, p_j\right), \tag{5}$$

where $\mathcal{Q}$ denotes a predetermined set of quantiles and p_j is an element of $\mathcal{Q}$.

The performance of developed models was compared through two aspects: prediction accuracy and computational cost. The computational cost was expressed as the training time of each model in seconds. As suggested by the ASHRAE Guideline 14-2014 [55], the prediction accuracy of point forecasting models was evaluated by two scale-independent metrics, namely coefficient of variation of the root mean square error (CV-RMSE) and normalized mean bias error (NMBE), over the entire test set. They are calculated by Eq. (7) and (9).

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}, \tag{6}$$

$$CV\text{-}RMSE = \frac{RMSE}{\overline{y}} \times 100, \tag{7}$$

$$MBE = \frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t), \tag{8}$$

$$NMBE = \frac{MBE}{\overline{y}} \times 100, \tag{9}$$

where n denotes the size of forecast horizon, y_t is the actual value of a target variable at time t, $\hat{y}_t$ is the predicted value of the target variable at time t, and $\overline{y}$ is the mean actual value of the target variable over the forecast horizon.

The CV-RMSE measures the variation between the actual values and the predictions of a model [55]. NMBE normalizes the mean bias error and gives the global difference between the actual and predicted values [56]. A positive NMBE value means that the model over-predicts actual energy consumption, and a negative one means under-prediction. For both CV-RMSE and NMBE, a closer value to zero represents better prediction accuracy. When making comparisons, we mainly focused on the CV-RMSE if the NMBE of a model is within the required range. As suggested by the ASHRAE Guideline 14-2014 [55], an applicable predictive model for energy use of whole building should have a CV-RMSE $\leq$ 30% and an NMBE within $\pm$10% when using hourly data for training models.

As in studies [34,40], the ρ-risk, which normalizes quantile losses across the entire forecast horizon, was used for evaluating the performance of probabilistic forecasting models. ρ-risk at pth quantile is calculated by

$$\rho\text{-}risk(p) = \frac{2 \times \sum_{t=1}^{n} \ell\left(\hat{y}_t^{(p)}, y_t, p\right)}{\sum_{t=1}^{n} y_t}, \tag{10}$$

where n denotes the size of forecast horizon, y_t is the actual value of a target variable at time t, $\hat{y}_t^{(p)}$ denotes the predicted pth quantile value at time t, and $\ell\left(\hat{y}_t^{(p)}, y_t, p\right)$ is the pth quantile loss calculated by Eq. (4).

4. Case study

To verify the performance of different deep learning architectures, a case study was conducted to develop predictive models for the energy consumption of two public historic buildings. This section describes details of the used dataset and experimental setup. The obtained results and discussion will be presented in Section 5.

4.1. Dataset

The dataset consists of two parts. One is the historical energy consumption data from two public historic buildings: the City Museum (Fig. 4a) and the City Theatre (Fig. 4b) in Norrköping, Sweden. The other is the meteorological data of Norrköping. The energy consumption data are electricity use and heating load provided by the building maintainer. Heating energy comes from the district heating system. Both types of energy use data are of the entire building. The meteorological data include dry-bulb temperature, relative humidity, dew point temperature, precipitation, air pressure, wind speed, and global irradiance. The meteorological data are obtained through open application programming interfaces (APIs) provided by the Swedish Meteorological and Hydrological Institute. The global irradiance is collected according to the latitude and longitude of the buildings, while other meteorological data are from a weather station located ~2 km away from the two buildings. All energy consumption and meteorological data range from 01:00 on January 1, 2016 to 00:00 on January 1, 2020, with a time granularity of one hour. Hours appearing in this paper are expressed in 24-hour format and are all in local time (Greenwich Mean Time (GMT) +1 for summer time and GMT +2 for winter time). The time span of the collected data is before the pandemic of COVID-19, which means that it excludes the impact of COVID-19 on public activities held in these two buildings.

These two public historic buildings have different operating modes. The normal operation of the City Museum is to maintain an appropriate indoor climate for preservation of collections and human comfort of staff and visitors. As shown in Table 2, the City Museum has regular opening hours. For the City Theatre, the operation mainly serves the delivery of live shows to audiences. For example, sound and light equipment and the ventilation system should work during a show or rehearsal. The performed shows have some seasonality. Several shows

Fig. 4. A case study was conducted in two public historic buildings: (**a**) the City Museum and (**b**) the City Theatre in Nörrkping, Sweden.

Table 2
Opening hours of the City Museum. Normally, it is open for six days, from Tuesday to Sunday every week. The opening hours will change in some holidays. On the day before Christmas Eve (December 23) and Epiphany (January 6), it opens from 11:00 to 16:00. On Christmas Eve (December 24), Christmas Day (December 25), and New Year's Day (January 1), it is closed.

	June–August	In other months
Monday	closed	closed
Tuesday	12:00–16:00	11:00–17:00
Wednesday	12:00–16:00	11:00–17:00
Thursday	12:00–20:00	11:00–20:00
Friday	12:00–16:00	11:00–17:00
Saturday	12:00–16:00	11:00–16:00
Sunday	12:00–16:00	11:00–16:00

of the same production are typically performed in adjacent 2–4 weeks. For example, 16 shows of the production *Farmor och Vår Herre* were performed during the period of February 24 to March 18, 2018. If a show is performed on one day, the start time is usually 19:00 on working days, 18:00 on Saturdays, and 16:00 on Sundays. Long shows last around three hours. In addition, shows are generally not performed on Mondays. The different operating modes of the two buildings could verify the adaptability of different predictive models to some extent.

4.2. Exploratory data analysis

Fig. 5 is time plot of hourly electricity consumption and heating load of the two buildings in the dataset. As revealed from the time plot, there is no notable trend in energy consumption for both buildings. No long-term increase or decrease can be inspected from both type of energy consumption. Nevertheless, a yearly seasonality exists. Both electric-

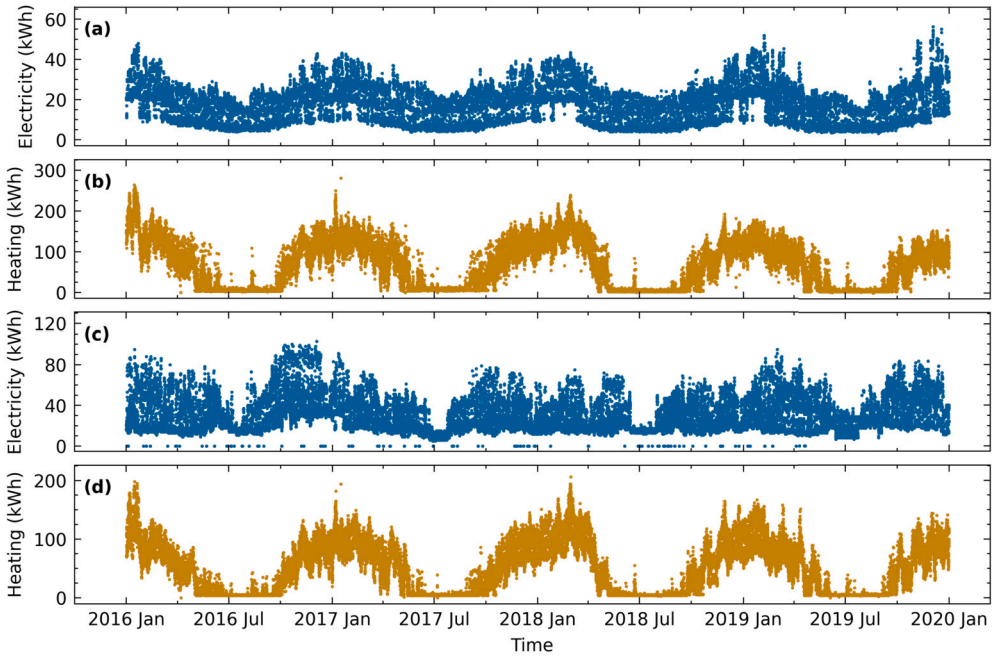


Fig. 5. Historical hourly (**a**) electricity consumption and (**b**) heating load of the City Museum, as well as (**c**) electricity consumption and (**d**) heating load of the City Theatre in Norrköping, Sweden from 01:00 on January 1, 2016 to 00:00 on January 1, 2020. Hours appearing in this paper are expressed in 24-hour format and are all in local time.

ity consumption and heating load are lower in summer and higher in winter. The distinctions in operating modes can be reflected in the electricity consumption of the two buildings. Due to regular opening hours, there is a strong yearly seasonality in electricity consumption of the City Museum (see Fig. 5a). However, the irregularity of show arrangements makes electricity consumption of the City Theatre (see Fig. 5c) vary from year to year. Compared to the considerable dissimilarity in the pattern of electricity consumption, the pattern in heating load of the two buildings (see Fig. 5b and 5d) has a high similarity. This similarity is mainly because both buildings employ adaptive district heating, which is driven by the difference between indoor and outdoor temperatures.

There is also a weekly seasonality in electricity consumption of the City Museum. As shown in Fig. 6, in each week, electricity consumption on weekdays is usually greater than on weekends. The electricity use on each day could basically reflect the opening hours on that day. Meanwhile, there are differences in electricity consumption between months. In winter months, the City Museum consumed more electricity than in summer months. In general, electricity consumption pattern of the City Museum is similar from year to year.

The yearly seasonality in electricity consumption of the City Theatre, as depicted in Fig. 7, is weaker than that of the City Museum. Electricity consumption varies considerably from year to year. For example, the electricity use during the shows held between October and December 2016 was greater than during shows held in other years. This dissimilarity in electricity consumption is because shows held in different periods differed. Different shows have distinct durations, and the use of lighting and sound equipment is also diverse among shows. Nevertheless, the electricity use can reflect how shows were scheduled. For instance, shows were typically not arranged on Mondays or in summer, and shows held on weekends began earlier than on weekdays. Therefore, incorporating show arrangements could help improve the prediction accuracy of electricity use of the City Theatre.

As both buildings employ adaptive district heating, there is a strong linear correlation between heating load and outdoor dry-bulb temperature (see Fig. 8). Lower outdoor temperatures result in higher heating loads to maintain a suitable indoor temperature for both buildings. However, when the dry-bulb temperature is less than −10 °C, the variation of heating load of the City Theatre (see Fig. 8b) is greater than that of the City Museum (see Fig. 8a). This larger variation might indicate that predicting heating load of the City Theatre is more difficult.

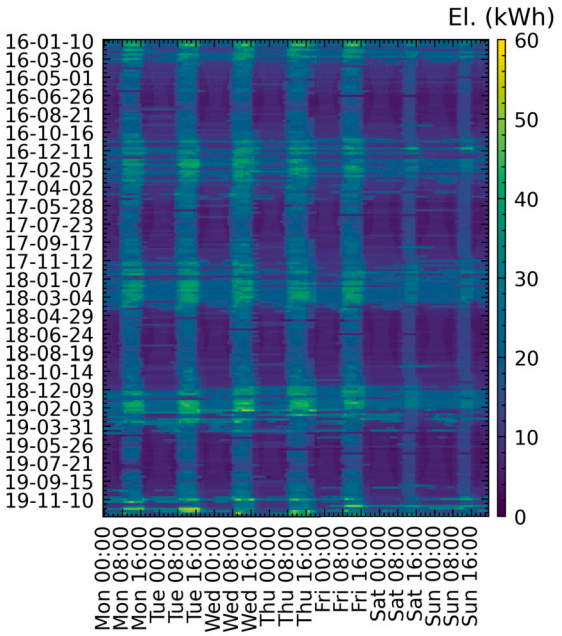


Fig. 6. The heat map of hourly electricity consumption of the City Museum from 00:00 on January 4, 2016 to 23:00 on December 29, 2019. Each row shows 168 data points, i.e., the energy consumption for each hour of one week from Monday (Mon) 00:00 to Sunday (Sun) 23:00. Date is represented as the format of YY-MM-DD. Electricity is abbreviated as El.

4.3. Data preprocessing

Data preprocessing aims to convert the raw data into a format that can be easily handled and understood by models. In this study, data pre-

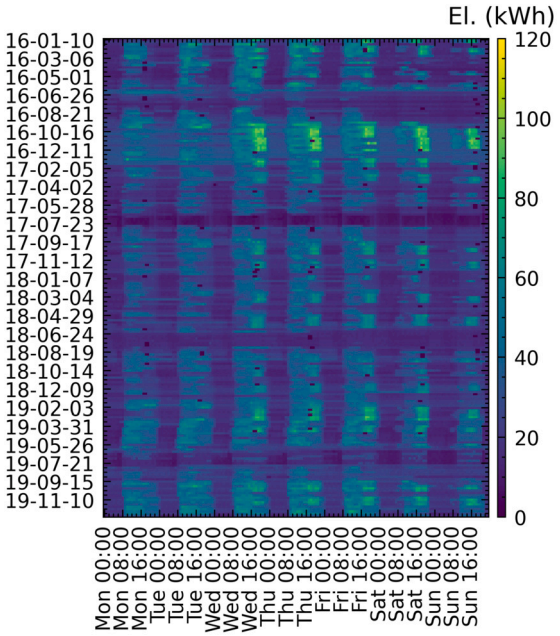


Fig. 7. The heat map of hourly electricity consumption of the City Theatre from 00:00 on January 4, 2016 to 23:00 on December 29, 2019.

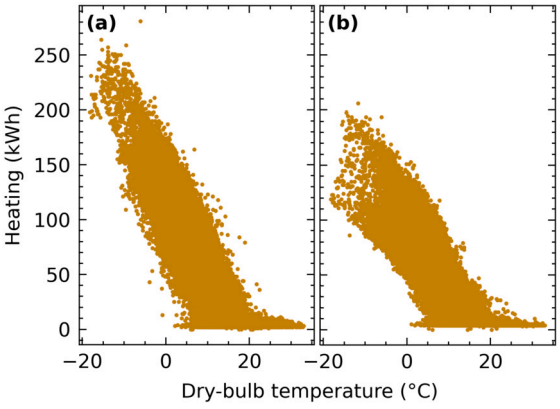


Fig. 8. The scatter plot of historical hourly heating load of (**a**) the City Museum and (**b**) the City Theatre versus outdoor dry-bulb temperature from 01:00 on January 1, 2016 to 00:00 on January 1, 2020.

processing includes data cleaning, dataset splitting, feature preparation, and data transformation.

4.3.1. Data cleaning and dataset splitting

First, missing values in meteorological data were interpolated linearly. Then, the dataset was divided into three subsets: a training set for learning the parameters of models, a validation set for tuning hyperparameters and preventing overfitting, and a test set for evaluating the performance of models. The dataset splitting roughly follows the empirical ratio of 80:10:10, where 38 months of data from January 1, 2016 to February 28, 2019 are used as the training set, five months of data from March 1, 2019 to July 31, 2019 are used as the validation set, and five months of data from August 1, 2019 to December 31, 2019 are used as the test set. The three subsets do not overlap in time, avoiding information leakage from the future. We did not identify and address outliers in meteorological data since the provider has ensured their validity. For the energy data, only the training set was inspected to avoid information leakage from the test set. As shown in Fig. 9, many outliers in electricity consumption of both buildings and one outlier in heating load of the City Theatre are identified. After inspecting the occurrence

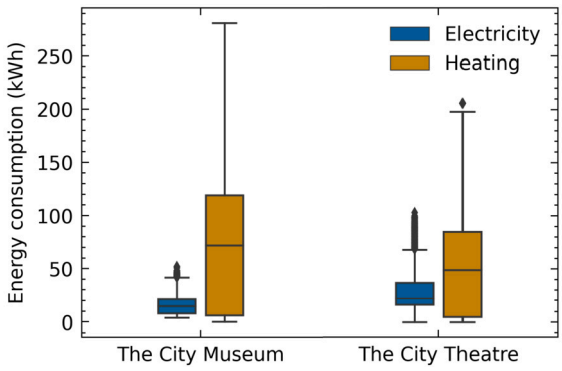


Fig. 9. The boxplot of hourly electricity consumption and heating load of the City Museum and the City Theatre from January 1, 2016 to February 28, 2019. Data points that are more than 1.5 box lengths from the edge of their box are classified as outliers, illustrated as diamond dots.

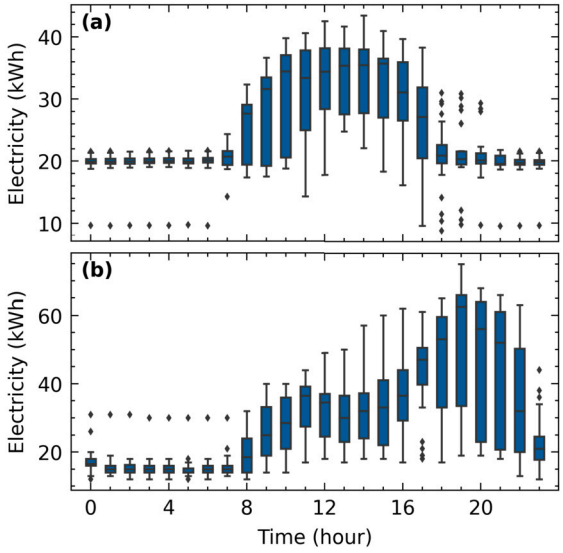


Fig. 10. The boxplot of electricity consumption per hour of (**a**) the City Museum and (**b**) the City Theatre from 00:00 on February 19 to 23:00 on March 18, 2018.

of these outliers, they are kept in the training set as these outliers are high energy consumption due to activities held in the buildings and are not anomalies.

4.3.2. Feature preparation

Feature preparation includes extracting temporal features from timestamps, generating features from operating modes of buildings, and reducing redundant features. Four temporal features are extracted: two binary and two cyclical variables. The binary variables include one called *is holiday* to indicate if a day is a Sweden public holiday and another called *is weekend* to indicate if a day is a weekend. The cyclical variables are *hour* (integer value from 0 to 23) and *weekday* (integer value from 0 to 6, each value represents a day in a week, starting from Monday). In addition to the temporal features, one feature called *is open* with a binary value is added to reflect the occupancy of a building for a given hour. For the City Museum, *is open* indicates that if it is open to visitors. For the City Theatre, *is open* indicates that if there is a show performed.

These features could help predict energy use. For example, there is usually a daily seasonality for operating a building. Fig. 10 shows the distribution of electricity consumption per hour of the two buildings during four weeks of the training set. The electricity consumption of the City Museum (see Fig. 10a) was stable, and the distribution was narrow

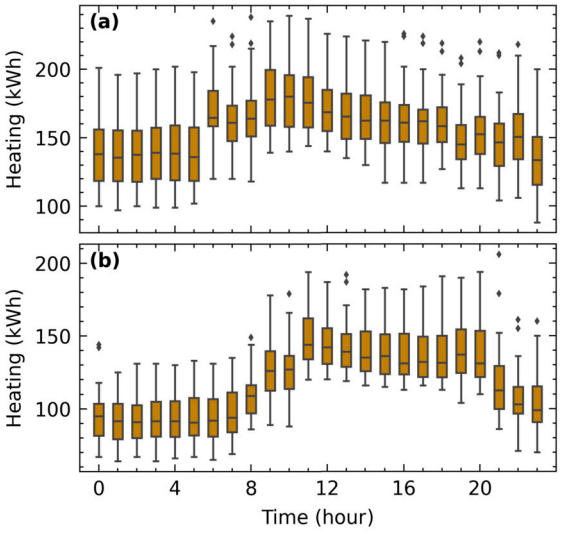


Fig. 11. The boxplot of heating load per hour of (**a**) the City Museum and (**b**) the City Theatre from 00:00 on February 19 to 23:00 on March 18, 2018.

Table 3
Pearson correlation coefficient (r) between variables for the City Museum in the training set. Electricity and Heating are target variables, while others are features. Temperature is abbreviated as temp. Except for the dry-bulb temperature, the coefficients between other features are not shown in the table because their $|r| < 0.7$.

	Electricity	Heating	Dry-bulb temp.
Dry-bulb temp.	−0.466	**−0.887**	1.000
Relative humidity	0.088	**0.384**	−0.562
Dew point temp.	**−0.517**	−0.843	**0.861**
Precipitation	−0.005	−0.017	0.001
Air pressure	0.019	0.000	−0.020
Wind speed	0.152	0.081	0.055
Global irradiance	0.127	**−0.332**	0.547
Is holiday	−0.090	0.035	−0.038
Is weekend	−0.216	−0.025	−0.004
Is open	**0.435**	0.102	0.101

Table 4
Pearson correlation coefficient (r) between variables for the City Theatre in the training set. Electricity and Heating are target variables, while others are features. Except for the dry-bulb temperature, the coefficients between other features are not shown in the table because their $|r| < 0.7$.

	Electricity	Heating	Dry-bulb temp.
Dry-bulb temp.	−0.102	**−0.850**	1.000
Relative humidity	−0.057	**0.336**	−0.562
Dew point temp.	−0.152	−0.826	**0.861**
Precipitation	−0.011	−0.022	0.001
Air pressure	0.046	0.011	−0.020
Wind speed	0.155	0.087	0.055
Global irradiance	0.053	**−0.309**	0.547
Is holiday	−0.090	0.041	−0.038
Is weekend	−0.152	−0.037	−0.004
Is open	**0.734**	0.153	−0.066

Table 5
A summary of used features for predicting each target variable.

Target variable	Features	Observable in
The City Museum		
Electricity	electricity	past
	dew point temp. is open hour weekday	past and future
Heating	heating	past
	dry-bulb temp. relative humidity global irradiance hour weekday	past and future
The City Theatre		
Electricity	electricity	past
	is open hour weekday	past and future
Heating	heating	past
	dry-bulb temp. relative humidity global irradiance hour weekday	past and future

before 8:00 and after 20:00. Between 8:00 and 17:00, hourly electricity consumption rose significantly due to the work of staff and the opening to the public, and the data was distributed wider. Between 18:00 and 20:00, although the median electricity consumption decreased, many outliers appeared because the City Museum remained open until 20:00 on Thursdays. A similar phenomenon can be observed in the electricity consumption of the City Theatre. As depicted in Fig. 10b, the highest median electricity consumption is at 19:00 because shows performed on working days started at that time. During the show time, the distribution of electricity consumption also became wider. The heating load of the two buildings also has a similar correlation with the hour of a day (see Fig. 11), although it is not as strong as for the electricity consumption. Therefore, extracting temporal features such as hour and weekday from timestamps helps predict energy consumption. Furthermore, opening hours can also provide information for making predictions.

A filter method based on finding the correlation between variables was employed to select critical features and reduce redundant features. The Pearson correlation coefficient (r) was used for measuring the linear relationship between two variables. As general rules of thumb, a threshold of $|r| \geq 0.3$ was employed to filter out critical features with at least moderate correlation with a target variable. To reduce redundant features, when two features are highly correlated ($|r| \geq 0.7$), the one holding larger $|r|$ with the target variable was kept to avoid duplicate information.

As coefficients shown in Table 3, to predict electricity consumption of the City Museum, *dew point temperature*, *is open*, and extracted temporal features *hour* and *weekday* were used. To predict heating load of the City Museum, *dry-bulb temperature*, *relative humidity*, *global irradiance*, *hour*, and *weekday* were used. Similarly, according to coefficients shown in Table 4, to predict electricity consumption of the City Theatre, *is open*, *hour*, and *weekday* were used. To predict heating load of the City Theatre, *dry-bulb temperature*, *relative humidity*, *global irradiance*, *hour*, and *weekday* were used.

The features that are used to predict the electricity use and heating load of the two buildings are summarized in Table 5. For predicting each target variable, past observations of itself, as well as past and future observations of predictor variables, are used. Given a forecast origin, temporal features in a forecast horizon such as *hour* and *weekday* are naturally known in advance. Operational features like *is open* in the forecast horizon can be retrieved from APIs provided by the maintainer of a building. Values of meteorological features in the forecast horizon are also considered available, as short-term weather forecasts, i.e., for the next 24 hours, are highly accurate nowadays. Many organizations provide APIs to access them. However, it is worth noting that this study used actual meteorological data to train and evaluate the model. Therefore, the prediction performance of models might have some degradation when these models are deployed in real applications due to the use of forecasted meteorological data.

4.3.3. Data transformation

Data transformation aims to change raw features into a more suitable representation for model learning. For observations of target vari-

ables electricity and heating over the lookback window, as well as meteorological features such as dry-bulb temperature, relative humidity, global irradiance, and dew point temperature, a min-max normalization was performed to scale each of them to a range of $[0, 1]$. All min-max scalers were fitted on the training set, then used for transforming validation and test sets. Cyclical features *hour* and *weekday* were transformed into two dimensions using a sine-cosine transformation. Binary features like *is open* were not transformed.

4.4. Experimental setup

Models are developed for predicting hourly electricity consumption and heating load of the City Museum and the City Theatre 24 steps ahead. In other words, given a forecast origin, models should predict electricity consumption and heating load of the two buildings for each hour of the following 24 hours. The maximum lookback window size was determined by the partial autocorrelation function. Both electricity consumption and heating load of the two buildings on the training set are non-stationary as assessed by the Kwiatkowski–Phillips–Schmidt–Shin test ($p < .01$). Consequently, we took a difference of lag 24 followed by a first difference for each target variable. The result suggested a maximum lookback window size of 168 hours, i.e., seven days, for all target variables. Therefore, given a forecast origin, we use the observed values of the target and predictor variables over the past 168 hours to predict the value of the target variable over the next 24 hours.

Based on the seasonal naïve (SN) method [54], two models, namely SN-24 and SN-168, were prepared as baselines because electricity consumption and heating load are highly seasonal. SN-24 model aims to utilize daily seasonality, and each forecast of a target variable was set to be its value observed 24 hours ago. Similarly, for the SN-168 model, each forecast was set to be the value observed 168 hours ago to use weekly seasonality.

All models except for the two baseline models were trained according to the following four cases:

- Case 1, point forecast: train the eight models (LR, N-HiTS, TCN, TF, NLinear, LSTM, GRU, and TFT) by only using past values (in the lookback window) of target and predictor variables (see Table 5).
- Case 2, point forecast: train the five models (LR, NLinear, LSTM, GRU, and TFT) by incorporating past values (in the lookback window) of target and predictor variables, as well as future values of predictor variables (in the forecast horizon).
- Case 3, probabilistic forecast: train the six models that achieved highest prediction accuracy in Cases 1 and 2 with all information they support using. The predefined set of quantiles is $\{0.1, 0.5, 0.9\}$.
- Case 4, point forecast: train the eight models with all information they support using except for the feature related to operating buildings.

Case 1 compares the performance of models in making multi-horizon predictions when only using past information. Case 2 looks into the impact of incorporating future information on model prediction accuracy. Case 3 is to investigate the performance of models in probabilistic forecast, and we are interested in evaluating ρ-risk on 0.5th and 0.9th quantiles. Case 4 is a sensitivity analysis to examine how operating mode-related features affect model prediction accuracy.

Models were trained by minimizing an appropriate loss function (as defined in Section 3.4) on the training set. A sliding window with a step size of one hour was used to generate training samples. All deep learning models were trained by a maximum of 100 epochs. An early-stopping training technique was employed to avoid overfitting. This technique stops the training process when the resulting accuracy in the validation set stops rising after a specified number of iterations (30 epochs in this study). For each deep learning architecture, the model that has the lowest loss for the validation set was selected. Finally, the selected models were evaluated and compared by predefined metrics (see Section 3.4) on the test set. A sliding window with a step size of 24 hours was used to generate all forecasts on the test set. Since the heating energy of both buildings is from the district heating system, the criterion recommended by the ASHRAE Guideline 14-2014 was applied separately to electricity use and heating load.

All experiments were conducted on a computer with Ubuntu 20.04 operating system. The computer has an Intel Xeon W-2145 CPU, a total of 16 GB memory, and a graphics card NVIDIA GeForce GTX 1080. Models were implemented by Python (v3.8.16) programming language based on libraries PyTorch [16] (v1.12.0), darts [57] (v0.23.1), and scikit-learn [58] (v1.2.1).

Table 6
The prediction accuracy of point forecasts for different models on the test set when not incorporating future information (Case 1). For both CV-RMSE and NMBE, a closer value to zero represents better prediction accuracy. Electricity is abbreviated as El.

	The City Museum		The City Theatre	
	El.	Heating	El.	Heating
CV-RMSE (%)				
SN-24	45.1	24.4	30.1	27.0
SN-168	55.3	41.2	31.2	46.2
N-HiTS	**32.2**	22.0	27.7	23.5
TCN	37.6	22.0	25.5	25.0
TF	40.3	23.0	32.2	29.6
LR	36.6	20.7	**23.6**	24.2
NLinear	33.8	29.1	30.6	22.9
LSTM	36.1	18.5	29.4	22.4
GRU	37.7	20.9	31.5	23.3
TFT	32.6	**17.4**	24.7	**21.9**
NMBE (%)				
SN-24	−0.5	−1.1	−0.1	−1.2
SN-168	−4.9	−8.1	0.1	−8.9
N-HiTS	−2.2	−0.4	−2.4	−5.9
TCN	−4.7	−0.6	−6.0	−0.3
TF	−10.0	0.1	−2.3	−3.3
LR	−2.8	−0.3	−0.9	−0.8
NLinear	7.5	18.3	5.9	0.8
LSTM	−6.4	−0.2	−6.0	0.8
GRU	−5.3	−7.3	−2.0	0.6
TFT	−1.3	2.1	−3.5	0.3

5. Results and discussion

The presentation and discussion of results include three parts. First is the quantitative analysis of the results based on predefined metrics. Then, the results are qualitatively analyzed through the exploratory data analysis approach. Finally, a discussion about integrating the predictive models developed in this study into applying a digital twin model was given.

5.1. Quantitative analysis

The quantitative analysis is to analyze the predictability of various energy use and the performance of models under the four cases, including both point and probabilistic forecasts.

5.1.1. Comparison of predictability of electricity and heating

Both electricity consumption and heating load of the two buildings exhibit a stronger daily seasonality than weekly seasonality as the SN-24 model performed better than the SN-168 model on all metrics (see Table 6). Meanwhile, heating load has stronger daily seasonality than electricity consumption for both buildings since the SN-24 model obtained a lower CV-RMSE on predictions of heating load than that of electricity consumption. Furthermore, prediction accuracy of the SN-24 model suggests that electricity consumption is less predictable than

heating load for both buildings. Interestingly, the SN-24 model provides a strong baseline, especially for heating load of both buildings, as its performance on predicting heating loads has met the criterion (30% for CV-RMSE and ±10% for NMBE) of the ASHRAE Guideline 14-2014 [55].

In addition to baseline models, the performance of the other eight models also indicates higher predictability in heating load. As shown in Table 6, except for the performance of the LR model on the City Theatre, the other seven models achieved a lower CV-RMSE on predictions of heating load than electricity consumption. Moreover, all eight models have met the criterion of the ASHRAE Guideline 14-2014 in predicting heating load of the City Museum, while seven of the eight models (except for the NLinear model, which has an NMBE of 18.3%) have met the criterion in predicting heating load of the City Theatre. However, no model achieved a CV-RMSE ≤ 30% when predicting electricity consumption of the City Museum. This result indicates that it is difficult to make an accurate prediction of the hourly electricity consumption of the City Museum for the next 24 hours by relying only on past information. The situation becomes better when predicting electricity consumption of the City Theatre, where five of the eight models obtained a CV-RMSE ≤ 30%. The higher predictability in heating load is attributed to the fact that the two buildings employ adaptive heating, which is driven by the difference between indoor and outdoor temperatures.

For the same type of energy consumption in different buildings, electricity consumption of the City Museum is less predictable than that of the City Theatre, while heating load is more predictable for the City Museum than the City Theatre. As shown in Table 6, all eight models achieved higher values of CV-RMSE on predictions of electricity consumption of the City Museum than the City Theatre. At the same time, seven of the eight models (except for the NLinear model) achieved lower values of CV-RMSE on heating load of the City Museum than the City Theatre. This phenomenon is attributed to the arrangement of shows in the City Theatre as adjacent days typically performed shows of the same production. For example, 11 shows of the production *Faust II* were performed during the period of September 28 to October 12, 2019. Such an arrangement leads to a high similarity in the operating mode of the City Theatre in neighboring days. However, gathering many audiences in a place for a long time also caused more considerable fluctuations in heating load as more audiences lead to higher internal heat gain [7].

However, most of the eight models did not obtain a remarkably improved prediction accuracy, e.g., a decrease of 10% in CV-RMSE, over the baseline SN-24 model when not incorporating future information. For predicting electricity consumption, the N-HiTS model performed best for the City Museum (CV-RMSE 32.2%), and the LR model performed best for the City Theatre (CV-RMSE 23.6%). The TFT model obtained the best performance on both buildings for predicting heating load (CV-RMSE 17.4% for the City Museum and CV-RMSE 21.9% for the City Theatre). Nevertheless, some models, such as the NLinear model for predicting the heating load of the City Museum and the TF model for predicting the electricity consumption of the City Theatre, cannot even outperform the SN-24 model.

5.1.2. The impact of incorporating future information

Including future values of predictor variables in the forecast horizon increases the prediction accuracy of models by providing information on factors that determine energy consumption. As shown in Table 7, all five models achieved improved performance on the CV-RMSE metric for predicting both types of energy consumption of the two buildings. Also, the improvements of most models (except for NLinear and GRU models on the City Theatre) in predicting heating load are more evident than in predicting electricity consumption. The TFT model performed best on both types of energy consumption of the City Museum (CV-RMSE 29.7% on electricity consumption and CV-RMSE 8.7% on heating load). For the City Theatre, the LR model performed best on electricity consumption (CV-RMSE 17.9%), while the LSTM model performed best on heating load (CV-RMSE 12.5%).

Table 7
The prediction accuracy of point forecasts for different models on the test set after incorporating future information (Case 2). The values in brackets represent the change in corresponding performance compared to Table 6, and negative values represent improvements.

	The City Museum		The City Theatre	
	El.	Heating	El.	Heating
CV-RMSE (%)				
LR	34.5	12.4	**17.9**	15.8
	(−2.1)	(−8.3)	(−5.7)	(−8.4)
NLinear	32.8	23.8	24.4	20.6
	(−1.1)	(−5.3)	(−6.3)	(−2.3)
LSTM	35.7	10.9	22.2	**12.5**
	(−0.3)	(−7.6)	(−7.2)	(−9.9)
GRU	32.6	9.8	21.4	13.5
	(−5.1)	(−11.1)	(−10.1)	(−9.8)
TFT	**29.7**	**8.7**	20.3	12.9
	(−2.9)	(−8.7)	(−4.4)	(−9.0)
NMBE (%)				
LR	−3.4	1.6	−0.2	−0.5
	(+0.6)	(+1.3)	(−0.8)	(−0.3)
NLinear	−3.0	11.5	4.7	2.7
	(−4.4)	(−6.8)	(−1.2)	(+1.9)
LSTM	−6.7	−2.4	0.8	−1.2
	(+0.2)	(+2.2)	(−5.2)	(+0.4)
GRU	−6.6	−2.5	0.0	−3.8
	(+1.3)	(−4.8)	(−2.0)	(+3.2)
TFT	−6.2	−0.1	−0.2	−1.8
	(+4.9)	(−2.0)	(−3.3)	(+1.5)

The highest prediction accuracy achieved with the LR model for predicting electricity consumption of the City Theatre brings some inspiration. If there exist strong linear correlations between some predictor variables and a specific type of energy consumption, a basic linear model might provide a strong baseline for energy forecasting. As illustrated in Table 4 of Section 4.3.2, a strong linear correlation (Pearson's $r = 0.734$) exists between the operating mode-related feature *is open* and electricity consumption of the City Theatre. The LR model managed to extract this correlation and generate accurate predictions.

5.1.3. The performance of probabilistic forecast

Based on the results in Cases 1 and 2, the best six models for performing point forecasts are N-HiTS, LR, NLinear, LSTM, GRU, and TFT. Among the six models, the TFT model dominates the probabilistic forecast. It performed best to capture the central tendency and upper distribution of heating load of the City Museum as well as both energy consumption of the City Theatre as it achieved the lowest ρ-risk at the 0.5th and 0.9th quantiles as in Table 8. For predicting electricity consumption of the City Museum, the GRU model performed best to capture the central tendency of the electricity consumption as it achieved the lowest ρ-risk at the 0.5th quantile (ρ-risk(0.5) = 0.182). The N-HiTS model, on the other hand, performed best to capture the upper distribution of electricity consumption of the City Museum (ρ-risk(0.9) = 0.142) and might be useful for predicting extreme values or identifying outliers. Among all models, the LR model performed worst when producing probabilistic forecast. The authors speculated that it is because the LR model assumed that the residuals are normally distributed, which may not hold true when estimating quantiles, as the distribution of the residuals can be skewed.

The probabilistic forecasts also reflect that heating load has higher predictability than electricity consumption. Except for the NLinear model on heating load of the City Theatre and the LR model, other models achieved lower ρ-risk at 0.5th quantile when predicting heating load than electricity consumption for both buildings. Meanwhile, the uncertainties in electricity consumption are larger than heating load since

Table 8
The ρ-risk at 0.5th and 0.9th quantiles of probabilistic forecasts for different models on the test set (Case 3). For each metric, lower values represent better performance.

	The City Museum		The City Theatre	
	El.	Heating	El.	Heating
ρ-risk (0.5)				
N-HiTS	0.199	0.133	0.199	0.164
LR	0.520	0.662	0.410	0.683
NLinear	0.241	0.177	0.187	0.197
LSTM	0.186	0.055	0.140	0.086
GRU	**0.182**	0.059	0.143	0.091
TFT	0.192	**0.054**	**0.136**	**0.081**
ρ-risk (0.9)				
N-HiTS	**0.142**	0.074	0.120	0.116
LR	0.266	0.296	0.210	0.268
NLinear	0.159	0.136	0.099	0.122
LSTM	0.143	0.029	0.072	0.058
GRU	0.148	0.027	0.074	0.067
TFT	0.165	**0.027**	**0.070**	**0.056**

Table 9
The computational cost of training different models in seconds. For point forecasting models, TCN, TF, and N-HiTS are from Case 1, while the other six models are from Case 2. Probabilistic forecasting models are all from Case 3.

	Training time (s)			
	The City Museum		The City Theatre	
	El.	Heating	El.	Heating
Point				
TCN	664	620	575	590
TF	4646	4889	4714	4712
N-HiTS	1318	1222	1121	1170
LR	**3**	**3**	**1**	**2**
NLinear	688	603	676	451
LSTM	1353	2044	1617	1531
GRU	1314	1391	2355	1468
TFT	5757	6348	3274	3852
Probabilistic				
N-HiTS	1389	1335	1150	1334
LR	**144**	**180**	**119**	**175**
NLinear	707	753	258	782
LSTM	1958	2901	2018	3668
GRU	2405	2440	1595	2709
TFT	5529	7220	3653	4602

these models also achieved higher ρ-risk at 0.9th quantile when predicting electricity consumption than heating load for the two buildings. Nevertheless, The uncertainty in predicting electricity consumption also implies that, on the one hand, it is favorable to enhance certainty by optimizing electricity usage while still ensuring the regular functionality of a building. On the other hand, additional operating model-related features that determine electricity use should be involved for a better forecast.

5.1.4. Computational cost

More complex models typically require more training time to extract patterns in data. As shown in Table 9, the basic LR model consumed far less training time than other deep learning models. Among deep learning models, the TCN and the NLinear models consumed less training time since they processed the entire data sequence in parallel. Recurrent models LSTM and GRU process temporal data sequentially, leading to a slower training process. The computational cost of the N-HiTS model is better than recurrent models. Models employing the self-attention mechanisms, such as TF and TFT, require quadratic time complexity concerning the length of the input sequence, making them less efficient than recurrent models for processing long sequences.

Table 10
The change of prediction accuracy of point forecasts when not incorporating opening hours (Case 4). Values in brackets after each metric reflect the performance change (negative values represent improvements, and positive values represent deterioration) versus the model using opening hours. The N-HiTS, TCN, and TF models are compared with Table 6 while the other models are compared with Table 7.

	The City Museum	The City Theatre
	Electricity	Electricity
CV-RMSE (%)		
N-HiTS	33.8 (+1.7)	26.1 (−1.6)
TCN	38.2 (+0.6)	25.0 (−0.5)
TF	38.0 (−2.3)	31.9 (−0.3)
LR	36.0 (+1.6)	**23.4** (+5.5)
NLinear	35.3 (+2.5)	29.8 (+5.4)
LSTM	33.8 (−2.0)	30.4 (+8.2)
GRU	33.2 (+0.6)	28.1 (+6.7)
TFT	**32.5** (+2.7)	31.4 (+11.1)
NMBE (%)		
N-HiTS	−9.0 (+6.8)	−5.9 (+3.4)
TCN	−6.6 (+1.9)	−4.6 (−1.5)
TF	−4.3 (−5.8)	−2.0 (−0.2)
LR	−3.3 (−0.1)	−0.9 (+0.8)
NLinear	4.6 (+1.6)	−5.2 (+0.5)
LSTM	−1.1 (−5.6)	−4.2 (+3.4)
GRU	−2.8 (−3.8)	−1.4 (+1.4)
TFT	−7.5 (+1.3)	−0.7 (+0.5)

It is worth the training time to develop recurrent models like LSTM and GRU, as well as the TFT model, which integrates LSTM and attention mechanisms. The three models exhibited better performance based on the findings of point and probabilistic forecasts. According to the prediction accuracy of point forecasts (Table 7), the TFT model outperformed other models in predicting both types of energy use of the City Museum (CV-RMSE 29.7% for electricity consumption and CV-RMSE 8.7% for heating load) and the LSTM model performed best in predicting heating load of the City Theatre (CV-RMSE 12.5%). For performing probabilistic forecasts (Table 8), LSTM, GRU, and TFT models obtained lower ρ-risks than other models.

5.1.5. Sensitivity analysis

Removing the operating mode-related feature, i.e., *is open*, has a greater impact on predicting electricity consumption of the City Theatre than the City Museum. As shown in Table 10, all five models that support incorporating future information obtained a deteriorated CV-RMSE on forecasts of electricity consumption of the City Theatre. The performance of the TFT (CV-RMSE 31.4%) and LSTM (CV-RMSE 30.4%) models failed to outperform the baseline SN-24 model (CV-RMSE 30.4%). For the City Museum, the performance change of the models varied. Some increased and some decreased, but the overall change was smaller than that of the City Theatre. This is mainly because the linear correlation between the City Theatre and opening hours (Pearson's $r = 0.734$ as in Table 4) is greater than that between the City Museum and opening hours (Pearson's $r = 0.435$ as in Table 3).

No model met the criterion (30% for CV-RMSE) of the ASHRAE Guideline 14-2014 for predicting electricity consumption of the City Museum when not incorporating operational features. This suggests that more factors that determine electricity consumption, such as the scheduling of ventilation and lighting systems, should be involved to improve prediction accuracy.

5.2. Qualitative analysis

The qualitative analysis interprets the impact of operating modes and activities on energy use.

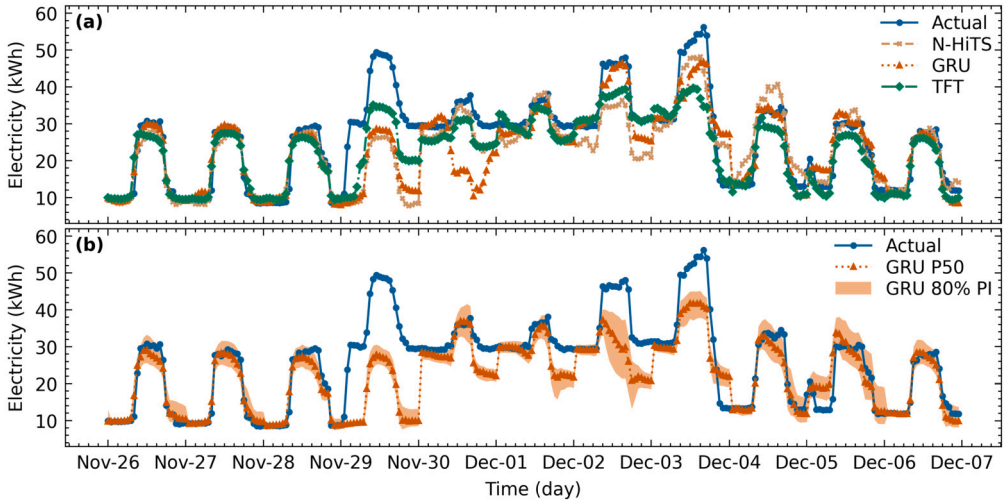


Fig. 12. The actual and predicted hourly electricity consumption of the City Museum from November 26 to December 6, 2019. (**a**) Point forecasts of the best three models and (**b**) probabilistic forecast of the GRU model. The predicted median is P50, and the 80% prediction interval (PI) is from 0.1th to 0.9th quantile.

5.2.1. Changes in operating mode of the City Museum

Previous quantitative analysis suggested that electricity consumption of the City Museum is less predictable than that of the City Theatre. The lower predictability was partly due to changes in the operating mode of the City Museum on some days in November and December 2019. Fig. 12a shows such a change. During the five days, from November 29 to December 3, the hourly energy consumption in the nighttime was even higher than in the daytime of the previous days. Meanwhile, the operating mode in these five days was also different. Among them, the hourly electricity consumption in the daytime was relatively high on November 29, December 2, and December 3, while it was relatively low on November 30 and December 1. From December 4, the operating mode changed to the pattern before November 29.

The changes in operating mode degrade the prediction accuracy of models during these days. On November 29 (the first day when the operating mode started to change), all three models believed the original operating mode would be maintained. Therefore, the forecasts followed the pattern in the previous operating mode (see Fig. 12a). The good news is that models adjusted their forecasts to adapt to the new operating mode from November 30 (the second day that the operating mode changed). After the operating mode changed to the old pattern before November 29, the forecasts of models also adapted to the change. The changes also introduce more uncertainty into forecasts. As shown in Fig. 12b, the 80% prediction interval (from 0.1th quantile to 0.9th quantile) during the daytime of the five days from November 29 to December 3 was relatively higher than during the daytime of the days before the operating mode changed. Furthermore, the higher uncertainty for forecasts during the daytime continued after the operating mode changed to the old pattern (see December 4 and 5).

Similar changes in the operating mode of the City Museum caused the prediction accuracy of models to deteriorate in November and December compared to the previous three months in the test set. As assessed by inspecting the boxplot (see Fig. 13) of CV-RMSE per 24 hours of the best five models, the median CV-RMSE of most models increased in November and December, and the distribution became broader in the two months.

5.2.2. Uncertainty brought by performing shows in the City Theatre

Activities held in a building might bring some uncertainty to electricity consumption. Fig. 14 depicts the actual and predicted hourly electricity consumption of the City Theatre in four days of October 2019. Each of the first three days had a show performed, and each show lasted for three hours. The hourly electricity consumption was higher during the show, and these predictive models can make good predictions (Fig. 14a). However, there is more uncertainty in forecasts when there are shows performed (Fig. 14b). Interestingly, on October 7 (Monday), the electricity consumption was expected to drop after 18:00, according to the probabilistic forecast (P50). However, it remained high until 23:00. Such information can be further investigated to better understand building energy use.

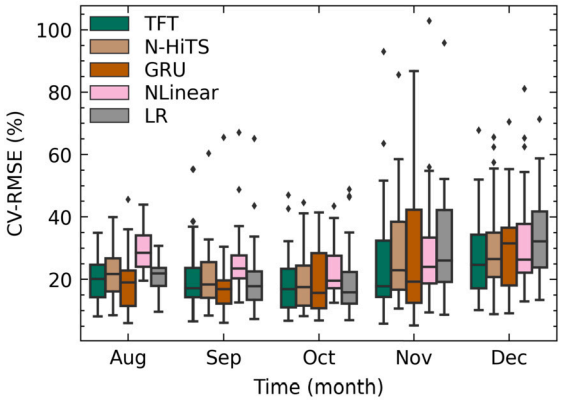


Fig. 13. The boxplot of the metric CV-RMSE per 24 hours of the best five models for predicted electricity consumption of the City Museum on the test set.

5.2.3. Heating is more predictable than electricity

Previous quantitative analysis indicates that heating load is more predictable than electricity consumption. On the one hand, higher predictability is attributed to strong influencing factors like dry-bulb temperature being involved in making predictions. On the other hand, the heating load is less affected by the change in operating mode. As shown in Fig. 15a, even on November 29 and 30, the two days when the operating mode changed, the best three models still made good predictions. Similarly, the uncertainty in predictions was greater during the daytime than during the nighttime (see Fig. 15b). The authors speculated that higher uncertainty in the daytime is likely due to more heat exchange between the indoor and the outdoor environment brought by staff and visitors.

6. Conclusion

This study set out to adapt and apply state-of-the-art deep learning architectures to address the problem of multi-horizon building energy forecasting. Eight methods, including seven deep learning architectures,

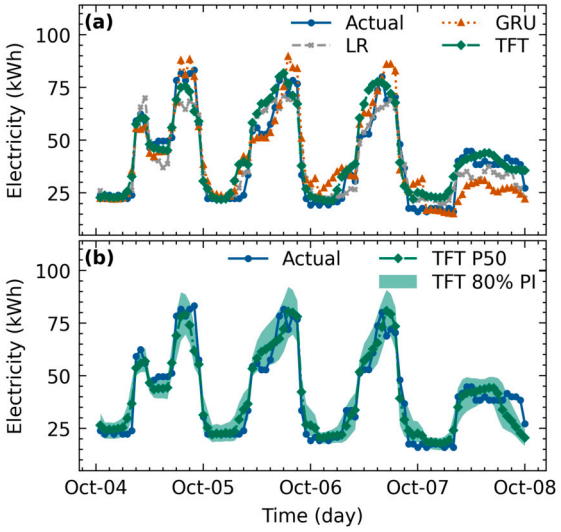


Fig. 14. The actual and predicted hourly electricity consumption of the City Theatre from October 4 (Friday) to October 7 (Monday), 2019. (**a**) Point forecasts of the best three models and (**b**) probabilistic forecast of the TFT model. During the first three days, one show was performed each day. No show was performed on the last day.

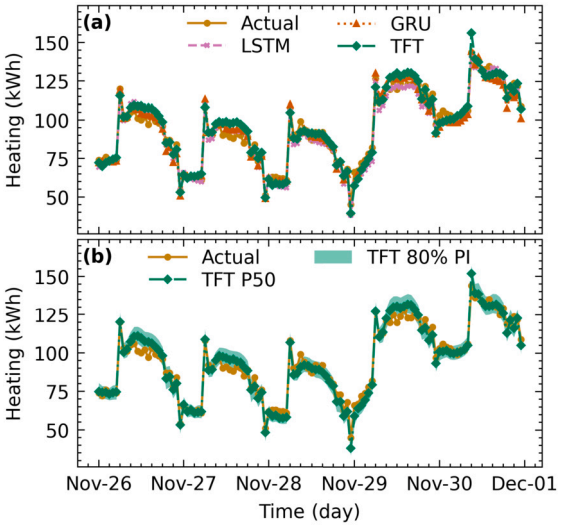


Fig. 15. The actual and predicted hourly heating load of the City Museum from November 26 to November 30, 2019. (**a**) Point forecasts of the best three models and (**b**) probabilistic forecast of the TFT model.

were studied to develop models for point and probabilistic forecasts. A comprehensive case study was conducted on two public historic buildings in Norrköping, Sweden, to evaluate the performance of these models and investigate factors that affect the predictability of energy consumption.

The results show that incorporating future information that determines coming energy consumption is critical for making multi-horizon predictions. Moreover, changes in the operating mode of a building and activities held in a building bring more uncertainty in energy consumption and deteriorate the performance of point forecasts. For point forecast, the TFT model performed best on both types of energy consumption of the City Museum (CV-RMSE 29.7% on electricity consumption and CV-RMSE 8.7% on heating load). The LR model performed best on electricity consumption of the City Theatre (CV-RMSE 17.9%), while the LSTM model performed best on heating load of the City Theatre (CV-RMSE 12.5%). The TFT model dominated the probabilistic forecast. Meanwhile, recurrent models like LSTM and GRU can make competitive quantile forecasts.

For future work, more features, especially occupancy and building operational data, might be included to study their impact on prediction accuracy. Also, the predictive models developed in this study could be integrated into a digital twin model of a building to reduce energy use while keeping the expected functionalities of the building.

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

Data will be made available on request.

Acknowledgement

The Swedish Energy Agency is acknowledged for the financial support of this study (project number: 50043-1). The authors thank Johan Björhn and his colleagues at Norrevo Fastigheter AB in Norrköping for providing access to the City Museum and the City Theatre as well as offering historical energy consumption data for these two buildings. Anna Donarelli at Uppsala University helps collect information about shows performed in the City Theatre. The Swedish Meteorological and Hydrological Institute is acknowledged for providing application programming interfaces to access meteorological data.

References

[1] J. Runge, R. Zmeureanu, A review of deep learning techniques for forecasting energy use in buildings, Energies 14 (3) (2021) 608, https://doi.org/10.3390/en14030608.
[2] C. Fan, J. Wang, W. Gang, S. Li, Assessment of deep recurrent neural network-based strategies for short-term building energy predictions, Appl. Energy 236 (2019) 700–710, https://doi.org/10.1016/J.APENERGY.2018.12.004.
[3] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, S. Ajayi, Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques, J. Build. Eng. 45 (2022) 103406, https://doi.org/10.1016/j.jobe.2021.103406.
[4] L. Zhang, J. Wen, Y. Li, J. Chen, Y. Ye, Y. Fu, W. Livingood, A review of machine learning in building load prediction, Appl. Energy 285 (2021) 116452, https://doi.org/10.1016/J.APENERGY.2021.116452.
[5] C. Delmastro, T.D. Bienassis, T. Goodson, K. Lane, J.-B.L. Marois, R. Martinez-Gordon, M. Husek, Buildings, Tech. Rep., The International Energy Agency, 9 Rue de la Fédération, 75739 Paris Cedex 15, France, 9 2022, https://www.iea.org/reports/buildings.
[6] P. Lara-Benítez, M. Carranza-García, J.M. Luna-Romera, J.C. Riquelme, Temporal convolutional networks applied to energy-related time series forecasting, Appl. Sci. 10 (7) (2020) 2322, https://doi.org/10.3390/app10072322.
[7] Z. Ni, Y. Liu, M. Karlsson, S. Gong, Enabling preventive conservation of historic buildings through cloud-based digital twins: a case study in the City Theatre, Norrköping, IEEE Access 10 (2022) 90924–90939, https://doi.org/10.1109/ACCESS.2022.3202181.
[8] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renew. Sustain. Energy Rev. 81 (2018) 1192–1205, https://doi.org/10.1016/j.rser.2017.04.095.
[9] N. Somu, G.R.M. R, K. Ramamritham, A deep learning framework for building energy consumption forecast, Renew. Sustain. Energy Rev. 137 (2021) 110591, https://doi.org/10.1016/j.rser.2020.110591.
[10] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisher, M.J. Witte, et al., Energyplus: creating a new-generation building energy simulation program, Energy Build. 33 (4) (2001) 319–331.
[11] Q. Qiao, A. Yunusa-Kaltungo, R.E. Edwards, Towards developing a systematic knowledge trend for building energy consumption prediction, J. Build. Eng. 35 (2021) 101967, https://doi.org/10.1016/j.jobe.2020.101967.
[12] M. Khalil, A.S. McGough, Z. Pourmirza, M. Pazhoohesh, S. Walker, Machine learning, deep learning and statistical analysis for forecasting building energy consumption — a systematic review, Eng. Appl. Artif. Intell. 115 (2022) 105287, https://doi.org/10.1016/j.engappai.2022.105287.

[13] Z. Ni, Y. Liu, M. Karlsson, S. Gong, A sensing system based on public cloud to monitor indoor environment of historic buildings, Sensors 21 (16) (2021) 5266, https://doi.org/10.3390/s21165266.
[14] Y. Liu, Z. Ni, M. Karlsson, S. Gong, Methodology for digital transformation with Internet of things and cloud computing: a practical guideline for innovation in small- and medium-sized enterprises, Sensors 21 (16) (2021) 5355, https://doi.org/10.3390/s21165355.
[15] Z. Ni, Y. Liu, M. Karlsson, S. Gong, Link historic buildings to cloud with Internet of things and digital twins, in: The 4th International Conference on Energy Efficiency in Historic Buildings, 2022, pp. 229–235.
[16] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: an imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems, vol. 32, Curran Associates Inc., 2019.
[17] W. Kong, Z.Y. Dong, Y. Jia, D.J. Hill, Y. Xu, Y. Zhang, Short-term residential load forecasting based on LSTM recurrent neural network, IEEE Trans. Smart Grid 10 (1) (2019) 841–851, https://doi.org/10.1109/TSG.2017.2753802.
[18] C. Zhang, J. Li, Y. Zhao, T. Li, Q. Chen, X. Zhang, A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process, Energy Build. 225 (2020) 110301, https://doi.org/10.1016/J.ENBUILD.2020.110301.
[19] C.H. Kim, M. Kim, Y.J. Song, Sequence-to-sequence deep learning model for building energy consumption prediction with dynamic simulation modeling, J. Build. Eng. 43 (2021) 102577, https://doi.org/10.1016/J.JOBE.2021.102577.
[20] C. Wang, Y. Wang, Z. Ding, T. Zheng, J. Hu, K. Zhang, A transformer-based method of multienergy load forecasting in integrated energy system, IEEE Trans. Smart Grid 13 (4) (2022) 2703–2714, https://doi.org/10.1109/TSG.2022.3166600.
[21] H. Dong, J. Zhu, S. Li, W. Wu, H. Zhu, J. Fan, Short-term residential household reactive power forecasting considering active power demand via deep transformer sequence-to-sequence networks, Appl. Energy 329 (2023) 120281, https://doi.org/10.1016/j.apenergy.2022.120281.
[22] Z. O'Neill, C. O'Neill, Development of a probabilistic graphical model for predicting building energy performance, Appl. Energy 164 (2016) 650–658.
[23] Y. Chen, Y. Kang, Y. Chen, Z. Wang, Probabilistic Forecasting with Temporal Convolutional Neural Network, Neurocomputing, vol. 399, Elsevier B.V., 2020, pp. 491–501.
[24] R. Wen, K. Torkkola, B. Narayanaswamy, D. Madeka, A multi-horizon quantile recurrent forecaster, https://doi.org/10.48550/arxiv.1711.11053, Nov. 2017.
[25] Z. Ni, P. Eriksson, Y. Liu, M. Karlsson, S. Gong, Improving energy efficiency while preserving historic buildings with digital twins and artificial intelligence, IOP Conf. Ser. Earth Environ. Sci. 863 (1) (2021) 012041, https://doi.org/10.1088/1755-1315/863/1/012041.
[26] J.F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, A. Troncoso, Deep learning for time series forecasting: a survey, Big Data 9 (1) (2021) 3–21, https://doi.org/10.1089/big.2020.0159.
[27] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, MIT Press Journals.
[28] K. Cho, B.v. Merriënboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: encoder-decoder approaches, in: Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation, Association for Computational Linguistics (ACL), ISBN 9781937284961, 2014, pp. 103–111.
[29] C. Zhang, C. Berger, M. Dozza, Social-iwstcnn: a social interaction-weighted spatio-temporal convolutional neural network for pedestrian trajectory prediction in urban traffic scenarios, in: 2021 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2021, pp. 1515–1522.
[30] C. Zhang, C. Berger, Learning the pedestrian-vehicle interaction for pedestrian trajectory prediction, in: 2022 8th International Conference on Control, Automation and Robotics (ICCAR), IEEE, 2022, pp. 230–236.
[31] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, et al., Comparison of learning algorithms for handwritten digit recognition, in: International Conference on Artificial Neural Networks, Perth, Australia, vol. 60, 1995, pp. 53–60.
[32] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv:1409.0473, 2014.
[33] B. Lim, S. Zohren, Time-series forecasting with deep learning: a survey, Philos. Trans. R. Soc. A, Math. Phys. Eng. Sci. 379 (2194) (2021) 20200209, https://doi.org/10.1098/rsta.2020.0209, The Royal Society Publishing.
[34] D. Salinas, V. Flunkert, J. Gasthaus, T. Januschowski, DeepAR: probabilistic forecasting with autoregressive recurrent networks, Int. J. Forecast. 36 (3) (2020) 1181–1191, https://doi.org/10.1016/j.ijforecast.2019.07.001, Elsevier.
[35] S. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv:1803.01271, Mar. 2018.
[36] A.v.d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, WaveNet: a generative model for raw audio, https://doi.org/10.48550/arxiv.1609.03499, Sep. 2016.
[37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[39] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, R. Jin, Fedformer: frequency enhanced decomposed transformer for long-term series forecasting, in: International Conference on Machine Learning, PMLR, 2022, pp. 27268–27286.
[40] B. Lim, S.Ö. Arık, N. Loeff, T. Pfister, Temporal fusion transformers for interpretable multi-horizon time series forecasting, Int. J. Forecast. 37 (4) (2021) 1748–1764.
[41] A. Zeng, M. Chen, L. Zhang, Q. Xu, Are transformers effective for time series forecasting?, https://doi.org/10.48550/arxiv.2205.13504, May 2022.
[42] Y. Wang, J. Chen, X. Chen, X. Zeng, Y. Kong, S. Sun, Y. Guo, Y. Liu, Short-term load forecasting for industrial customers based on TCN-LightGBM, IEEE Trans. Power Syst. 36 (3) (2021) 1984–1997, https://doi.org/10.1109/TPWRS.2020.3028133.
[43] L. Lei, W. Chen, B. Wu, C. Chen, W. Liu, A building energy consumption prediction model based on rough set theory and deep learning algorithms, Energy Build. 240 (2021) 110886, https://doi.org/10.1016/j.enbuild.2021.110886.
[44] Y. Gao, Y. Ruan, Interpretable deep learning model for building energy consumption prediction based on attention mechanism, Energy Build. 252 (2021) 111379, https://doi.org/10.1016/j.enbuild.2021.111379.
[45] G. Zhang, C. Tian, C. Li, J.J. Zhang, W. Zuo, Accurate forecasting of building energy consumption via a novel ensembled deep learning method considering the cyclic feature, Energy 201 (2020) 117531, https://doi.org/10.1016/j.energy.2020.117531.
[46] M. Sajjad, Z.A. Khan, A. Ullah, T. Hussain, W. Ullah, M.Y. Lee, S.W. Baik, A novel CNN-GRU-based hybrid approach for short-term residential load forecasting, IEEE Access 8 (2020) 143759–143768, https://doi.org/10.1109/ACCESS.2020.3009537.
[47] E. Skomski, J.-Y. Lee, W. Kim, V. Chandan, S. Katipamula, B. Hutchinson, Sequence-to-sequence neural networks for short-term electrical load forecasting in commercial office buildings, Energy Build. 226 (2020) 110350.
[48] M. Cai, M. Pipattanasomporn, S. Rahman, Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques, Appl. Energy 236 (2019) 1078–1088.
[49] B.N. Oreshkin, D. Carpov, N. Chapados, Y.B. Mila, N-BEATS: neural basis expansion analysis for interpretable time series forecasting, https://doi.org/10.48550/arxiv.1905.10437, May 2019.
[50] C. Challu, K.G. Olivares, B.N. Oreshkin, F. Garza, M. Mergenthaler-Canseco, A. Dubrawski, N-HiTS: neural hierarchical interpolation for time series forecasting, https://doi.org/10.48550/arxiv.2201.12886, Jan. 2022.
[51] R. Koenker, G. Bassett, Regression quantiles, Econometrica 46 (1) (1978) 33, https://doi.org/10.2307/1913643, JSTOR.
[52] L. Hao, D.Q. Naiman, D.Q. Naiman, Quantile Regression, Sage, 2007.
[53] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[54] R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice, OTexts, 2018.
[55] Ashrae Guideline 14-2014: Measurement of Energy, Demand, and Water Savings, Standard, ASHRAE, 1791 Tullie Circle, NE, Atlanta, GA, 2014.
[56] G. Ramos Ruiz, C. Fernandez Bandera, Validation of calibrated energy models: common errors, Energies 10 (10) (2017) 1587.
[57] J. Herzen, F. Lässig, S.G. Piazzetta, T. Neuer, L. Tafti, G. Raille, T.V. Pottelbergh, M. Pasieka, A. Skrodzki, N. Huguenin, J. Kościszkościsz, D. Bader, F. Gusset, M. Benheddi, C. Williamson, M. Kosinski, M. Petrik, G. Grosch, Darts: user-friendly modern machine learning for time series maxime dumonal †, J. Mach. Learn. Res. 23 (2022) 1–6.
[58] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.