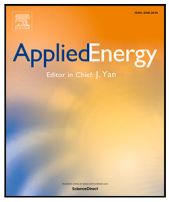
Applied Energy 374 (2024) 124027

Contents lists available at ScienceDirect

Applied Energy

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



Energy consumption forecasting based on spatio-temporal behavioral analysis for demand-side management

Jieyang Peng [a,b], Andreas Kimmig [b], Dongkun Wang [a], Zhibin Niu [c,*], Xiufeng Liu [d], Xiaoming Tao [a], Jivka Ovtcharova [b]

[a] Department of Electronic Engineering Tsinghua University, Beijing 100084, PR China
[b] Institute for Information Management in Engineering, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany
[c] College of Intelligence and Computing, Tianjin university, 300072 Tianjin, PR China
[d] Department of Technology, Management and Economics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Keywords:
Smart grid
Demand side energy management
Visual analytics
Energy demand forecasting
Energy behavior and patterns

ABSTRACT

Understanding user-level energy demand is pivotal for sustainable development and smart grid implementation, as it facilitates optimal resource allocation and energy conservation. Traditional machine learning models, however, often ignored the potential connections between users, limiting comprehension of user-level energy demands. Capturing behavioral correlations among users in high-dimensional temporal data remains challenging. In this paper, a novel scheme is proposed for revealing the implicit energy behavior correlation, which serves as prior knowledge to enhance predictive model performance. In addition, a spatio-temporal feature extraction framework is introduced to fuse spatial and temporal information, capturing the coherence of energy consumption data. This approach captures the correlation between the energy consumption patterns of different users, achieving highly accurate demand forecasting at the user level. In fine-grained demand forecasting experiments, the prediction accuracy of the proposed approach was improved by 14% compared with the baseline model. Based on fine-grained prediction results, an innovative visual analytical interface is also developed to characterize the migration of energy demand over time in both physical and topological space, offering valuable insights into demand-side energy management. In the empirical study, we found that there is an obvious mental inertia in the energy behavior of urban residents, which leads to energy waste. Our research provides critical insights for policymakers and planners in addressing the sustainability challenges of urban energy systems.

1. Introduction

Electricity enables the functioning of various critical components of city life, including lighting, heating, cooling, communication, and industrial processes. As one of the most complex networks in modern society [1], even though power system failures are less frequent, the catastrophic consequences cannot be ignored. For instance, a significant power blackout occurred on August 14, 2003, impacting regions in the Midwest and Northeast United States, as well as Ontario, Canada, affecting around 50 million people and leading to an estimated cost of up to 10 billion U.S. dollars [2]. Subsequently, on November 4, 2006, a comparable blackout, originating in Germany, ultimately caused a widespread power failure across Europe [3].

Accurate forecasting of energy demand patterns allows for proactive planning and resource allocation and enhances the overall resilience of urban energy systems [4]. For example, when a peak demand event is anticipated, such as during extreme weather conditions or when large events are scheduled, the government can proactively engage with consumers through demand response programs, such as sending alerts to residential and industrial customers and encouraging them to reduce their electricity consumption during peak hours by adjusting thermostats, delaying non-essential tasks, or using energy-efficient appliances. Although total energy consumption may not be reduced, these demand response initiatives help to distribute energy usage more evenly throughout the day, reducing the strain on the grid during peak periods. This not only minimizes the risk of energy supply disruptions but also avoids the need for costly infrastructure upgrades to meet sudden spikes in demand.

Investigating the dynamics of residential energy consumption at finely-resolved timescales is increasingly practical with the growing availability of user-grained data. However, existing energy forecasting studies still focus mainly on overall energy demand, e.g., at the city

* Corresponding author.
E-mail addresses: zniu@tju.edu.cn (Z. Niu), taoxm@tsinghua.edu.cn (X. Tao).

https://doi.org/10.1016/j.apenergy.2024.124027
Received 24 April 2024; Received in revised form 30 May 2024; Accepted 23 July 2024
Available online 1 August 2024

level [5] or at the national level [6]. The neglect of user-level energy consumption data prevents a deeper exploration of the correlation between users' energy behaviors, which is significant for predicting peak demand periods [7] (since the overall energy consumption is cumulative among the individuals). Therefore, understanding user-level patterns can reveal important insights into demand surges and help anticipate and mitigate potential grid overloads. Furthermore, disregarding user-specific data hinders the identification of efficiency improvement opportunities. Detailed user data can expose wasteful consumption patterns and suggest targeted energy conservation measures. Specifically, the limitations of current research can be summarized as follows:

- Existing energy demand forecasting models usually lack spatial flexibility, typically focusing on a specific area, such as an entire city, while being unable to provide granular predictions at varying spatial resolutions. It restricts the understanding of energy demand at a finer spatial resolution, which is crucial for targeted energy planning and infrastructure optimization.
- Existing studies usually consider urban residential users as isolated nodes in the Euclidean domain and overlook the complex interactions and patterns inherent between different users (e.g., the synchronization of energy consumption due to seasonal changes, economic status, and electricity pricing policies), leading to inaccurate predictions and, subsequently, flawed decision-making.
- Current forecasting studies typically produce numerical predictions, lacking intuitive visualization and interpretability. This limitation hinders the comprehension of the results and their practical implications, thus impeding their direct application for valuable insights in demand-side energy optimization.

To counter the above knowledge gap, this paper proposes a spatio-temporal feature fusion model based on graph neural networks. The model can explicitly incorporate domain knowledge of energy patterns in the form of graph-structured data with high-dimensional features, thus resulting in interpretable energy forecasting results with high accuracy and fine granularity. The innovation of this paper can be demonstrated as follows:

- We proposed a mapping mechanism that converts high-dimensional time-series energy consumption data from the Euclidean domain to two-dimensional topological space, which enables the quantification of energy consumption behaviors in topological space. Based on this mechanism, we revealed six common patterns of urban energy consumption.
- We propose an energy demand forecasting framework leveraging the spatial and temporal coherence of energy consumption data. Compared with traditional energy forecasting models, the proposed method considers the spatio-temporal correlation between the energy consumption patterns of different users and achieves highly accurate demand forecasting at the user level.
- We present an innovative visual analytical mechanism to characterize the migration of energy demand over time in both physical and topological space. The mechanism provides a graphical solution for observing energy demand migration from both urban space and behavioral pattern perspectives, delivering insights into demand-side energy management.

The remainder of this paper is organized as follows: Section 2 summarizes the current trends in energy demand forecasting research and points out the research gaps from a modeling perspective. Section 3 demonstrates the proposed methodology of spatio-temporal information fusion in detail and justifies the rationality of the methodology. In Section 4, we first introduce the dataset of this paper and then demonstrate the results of the comparison experiments. In Section 5, we provide the visual encoding interface. We discovered the phenomenon of mental inertia in energy behavior based on visual analysis. Finally, we analyze the reasons for mental inertia and provide recommendations.

2. Related work

In this section, we delve into the advancements in energy demand forecasting models, highlighting the transition from traditional time series analysis to the adoption of deep learning techniques. We also identify research gaps in the current methodologies, emphasizing the need for sophisticated models that can handle the complexities of high-dimensional, user-grained energy data.

2.1. Energy demand forecasting model

The evolution of time-series forecasting techniques for electricity demand has been remarkable. Initially, time series analysis has been the norm for predicting overall energy consumption patterns in a city. Such approaches often rely on statistical models like autoregressive integrated moving averages (ARIMA) or exponential smoothing, which excel at capturing trends and seasonal patterns in the data. Tarmanini [8] used ARIMA to predict daily electricity load data for Ireland. Mohanad [9] proposes a novel hybrid model based on empirical mode decomposition with adaptive noise for reliable forecasting of electrical energy demand, finding it particularly suitable for whole-week and monthly forecasts. Fan [10] combined an econometric model with the GARCH model (Generalized Autoregressive Conditional Heteroskedasticity) to forecast electricity consumption and then analyzed electricity usage and economic behaviors using the Nash equilibrium. Time series analysis focuses on using historical data trends and seasonal changes to predict future demand. This method performs well when processing data with an obvious time periodicity [11]. However, it has a weak ability to handle unexpected events and irregular patterns, and it has high requirements for data stationarity.

In recent years, with the rise of machine learning and artificial intelligence technology, more and more researchers have applied these methods to user-level electricity demand forecasting. Machine learning methods, such as neural networks, support vector machines (SVM), and random forests, can automatically learn and identify complex patterns and nonlinear relationships in data without the need for predefined mathematical models. Jiang [12] provided an optimized support vector machine method for electricity forecasting with noise and seasonal factors and demonstrates its potential to improve forecasting performance. Abinet [13] presented a novel support vector machine (SVM) to forecast short-term heat demand in a district heating system. Fan [14] presented a hybrid model that combines random forest (RF) and mean generating function (MGF) to enhance the accuracy of short-term load forecasting, especially during peaks and valleys of highly fluctuating data. Aprillia [15] provided a novel load forecasting model based on random forest, offering an understanding of the outcome risk of the load demand profile. Machine learning makes predictions by finding the optimal hyperplane in high-dimensional space. These methods have certain advantages when dealing with small sample sizes and high-dimensional data but may be affected by overfitting and noisy data.

The emergence of deep learning has further revolutionized this field, offering unprecedented capabilities for handling large-scale, high-dimensional data. Deep learning models like recurrent neural networks (RNNs) and their variants, such as LSTMs, have been particularly effective in modeling sequential data with complex temporal dependencies. Abbasimehr [16] proposed a demand forecasting method based on multi-layer LSTM networks, demonstrating superior performance in predicting highly fluctuating data. Jin [17] integrated singular spectrum analysis (SSA) with long short-term memory (LSTM) neural networks, demonstrating improved computational efficiency in energy consumption forecasting. Jihoon [18] compared the prediction performance of three LSTM models, demonstrating that incorporating operation pattern data can improve the accuracy of energy consumption prediction. Convolutional neural networks (CNN) have proven effective in capturing temporal dependencies and spatial correlations in

demand data. Khan [19] presented a novel hybrid network model 'DB-Net' by incorporating a dilated convolutional neural network (DCNN) with bidirectional long short-term memory (BiLSTM). Atik [20] introduced a new CNN-based forecasting method that incorporates empirical mode decomposition for input signal processing in short-term electricity consumption forecasting.

2.2. Research gaps

Traditional forecasting methods for overall energy demand, while providing valuable insights, often lack the precision necessary for effective energy management at the individual user level. These methods tend to overlook the unique consumption patterns and variabilities inherent in user-specific data [21], limiting their usefulness in tailored energy solutions. With the advent of smart meters and advanced data collection techniques, the focus has shifted towards user-level demand forecasting [22].

User-level demand forecasting necessitates more sophisticated modeling approaches that can account for the unique patterns and variabilities observed at the individual user level. However, user-level energy demand forecasting is challenging, and the difficulties can be summarized as follows:

- High-dimensional user-grained energy data poses significant challenges for accurate analysis and modeling [23]. User-grained energy data frequently exhibits high dimensionality. The high dimensionality arises from the multitude of features inherent in user-level data, each representing different aspects of energy consumption. Mathematically, this translates into a vast feature space, where each feature can potentially influence the target variable (i.e., energy demand). As the dimensionality increases, so does the computational complexity and the risk of overfitting, especially when dealing with limited sample sizes, which require more sophisticated modeling techniques to accurately represent.
- In addition, multiple factors may influence energy usage simultaneously, leading to potential multicollinearity issues [24]. For instance, when multiple users with similar energy consumption behaviors are used as model inputs, it can lead to unstable and uninterpretable coefficient estimates. Specifically, multicollinearity can cause the determinant of the design matrix in a forecasting model to approach zero, leading to large variances in the estimated regression coefficients. This, in turn, can result in wide confidence intervals and make statistical inference unreliable.

3. Methodology

3.1. Requirements and theoretical hypotheses

This paper proposes a user-level energy demand prediction framework. Specifically, this framework receives the historical energy consumption data of all users in a certain area as input and predicts the energy demand for each user by considering the spatio-temporal correlation of energy behaviors among users. In this paper, energy correlation can be defined as the degree of similarity in energy consumption behavior, e.g., if two users have similar temporal distributions of energy consumption peaks, their energy demands are "correlated". The superiority of the proposed method is based on the following theoretical hypotheses:

General Hypotheses: If two users have similar energy behaviors, the prediction accuracy of the energy demand of one user can be improved by considering the historical energy data of the other.

To verify the above hypothesis, we propose several gradually deepening sub-hypotheses:

- ***Similarity Transfer Hypothesis:*** If the energy consumption behaviors of users have shown similarities in the past, this similarity is likely to persist in the future. This hypothesis assumes that past historical energy consumption data can effectively represent and predict future consumption patterns.
- ***Spatio-Temporal Correlation Hypothesis:*** This hypothesis emphasizes that there is a spatiotemporal correlation between the energy consumption behaviors of users. That is, a user's energy consumption is not only influenced by their own historical consumption patterns but also by the consumption patterns of surrounding or similar users.
- ***Generalization Hypothesis:*** The prediction model needs to have good generalization ability, which means the model should not only fit historical data but also effectively predict unseen data (such as new users or new consumption patterns).

In this section, we will justify the Similarity Transfer Hypothesis, while the Spatio-Temporal Correlation Hypothesis and the Generalization Hypothesis will be argued in Sections 3.4 and 4.3, respectively. To justify the similarity transfer hypothesis, we can draw parallels from various fields where historical data has proven to be a valuable predictor of future outcomes. In meteorology, for instance, past weather patterns are routinely used to forecast future weather conditions. Similarly, in the financial markets, historical trends and patterns often inform predictions about future market movements. Applying this logic to energy consumption, it stands to reason that if users have demonstrated similar consumption patterns in the past, it is likely that these patterns will persist, barring any significant external factors that might disrupt them. This is because human behavior tends to be habitual, and energy consumption patterns are often driven by routine activities such as cooking, heating, and lighting needs, which are relatively stable over time.

In fact, the majority of energy demand forecasting studies are based on this theory [25]. Researchers and analysts rely heavily on historical energy consumption data to model and predict future demand [26]. This hypothesis has been widely adopted due to its demonstrated effectiveness in capturing patterns and trends that inform future energy needs.

3.2. Overall framework

The overall architecture of the proposed method in this paper is shown in Fig. 1. With smart grids, user-level energy consumption data is imported into the database in real time, and each user's GPS coordinate is also stored in the database in advance. An example of the energy database is shown in Fig. 2. In Fig. 2, the first column is a unique user ID. The second and third columns are the date and the energy consumption for that day, respectively. The fourth and fifth columns describe the user's geographic location in the form of GPS coordinates. The energy source in this paper refers specifically to electricity; however, the proposed method can be applied to any kind of energy source.

Upon examination, we identified scattered missing values within the data, presumably due to equipment malfunctions or recording errors. To ensure accurate data analysis, we have chosen to use linear interpolation as a method to fill in these missing data points. In our case, for each missing data point, we identify the known values before and after it in the dataset. Then, we use linear interpolation to calculate the missing value based on the known values and their positions.

Then, we extract the temporal features of the electricity behavior of users by temporal clustering, which describes the similarity of users' temporal electricity behavior. The extracted features are depicted in the form of an adjacency matrix. Specifically, the rows and columns in the adjacency matrix represent the users, and the values in the matrix are calculated from the similarity of the temporal energy consumption data. Meanwhile, we adopt the Euclidean distance calculated by the user's GPS coordinates as a spatial correlation feature and record it in the spatial adjacency matrix.

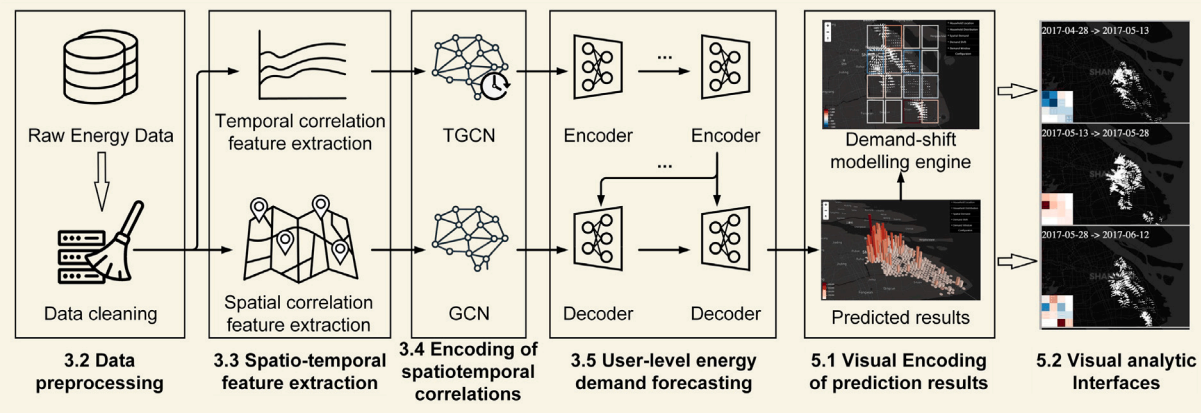


Fig. 1. Overall framework of the proposed method.

CustomerID	pap_r	Date	Latitude	Longitude
210948721	1.65	12/07/2016	121.62291	31.18957621
210948721	8.84	22/08/2016	121.62291	31.18957621
210948721	1.6	16/10/2016	121.62291	31.18957621

Fig. 2. Examples of energy data used in this paper.

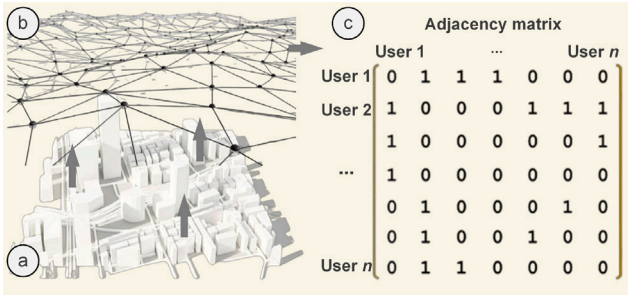


Fig. 3. Construction of urban energy network.

Subsequently, the temporal graph neural network and the graph neural network further encode the spatio-temporal correlation feature matrix, respectively, and input the encoded spatio-temporal features into the transformer-based multidimensional prediction model (with each user as a dimension). The model will output the future energy demand of each user. Finally, we develop a visual analytical interface based on potential flow and quantify the migration of energy demand at different resolutions by fluxes.

3.3. Extraction of spatio-temporal features

After data cleaning, we acquired continuous time-series energy consumption information and spatial location (GPS coordinates) for each user. In this section, we will use this information to extract the spatio-temporal correlation features among users, as shown in Fig. 3. In Fig. 3, the demand-side energy system is considered a network. Each user acts as a node in the network, and the edges in the network are constructed by calculating the spatio-temporal similarity between users (sub-figures a–b). We quantify the temporal and spatial similarity by two different similarity measures, respectively (sub-figures b–c).

For each user, the time-series energy consumption data can be regarded as a one-dimensional probability distribution. Therefore, we can extract the time-series correlation features of the whole energy network using the distance metrics method for probability distributions. Let $x_1, \ldots, x_N$ be the time series data of power consumption, where N is the number of users. The probabilities p_{ij}, which are proportional to the similarity of objects x_i and x_j, are computed as follows:

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2N} \tag{1}$$

$$p_{i|j} = \frac{exp(-\left\|x_i - x_j\right\|^2)}{\sum_{k \neq l} exp(-\left\|x_k - x_l\right\|^2)} \tag{2}$$

Here, $p_{i|j}$ represents the probability of x_i given x_j. It is noteworthy that $p_{ij} = p_{ji}$, $p_{ii} = 0$, and $\sum_{ij} p_{ij} = 1$.

To learn a low-dimensional map $y_1, \ldots, y_N$ that reflects the similarities p_{ij} as well as possible, a t-distribution is used to measure similarities between low-dimensional points, allowing dissimilar objects to be modeled far apart in the map. The distribution q_{ij} is defined as:

$$q_{ij} = \frac{(1 + \left\|y_i - y_j\right\|^2)^{-1}}{\sum_{k \neq l}(1 + \left\|y_k - y_l\right\|^2)^{-1}} \tag{3}$$

To make the distribution in high and low dimensions as similar as possible, the locations of the points y_i in the low-dimensional map are determined by minimizing the Kullback–Leibler divergence [27] of the distribution P from the distribution Q, which can be expressed as:

$$KL(P \,||\, Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{4}$$

The minimization of the Kullback–Leibler divergence concerning the points y_i is performed using gradient descent. The gradient is given by:

$$\frac{\delta KL}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \left\|y_i - y_j\right\|^2)^{-1} \tag{5}$$

This process aims to minimize the Kullback–Leibler divergence between the two similarity distributions p_{ij} and q_{ij} (in high and low dimensions), ensuring that points close in the high-dimensional space remain close in the low-dimensional representation. This optimization preserves the local structure of the data in the low-dimensional space, effectively revealing hidden structures in the original high-dimensional data. As a result, users with similar temporal energy consumption patterns will be positioned closely on the reduced two-dimensional plane, as shown in Fig. 4(a). Simply speaking, each point in Fig. 4(a) represents a user. The more similar the energy distribution is, the closer the points of the users are located in the graph.

Then, an adjacency matrix M is constructed, where each entry M_{ij} represents the distance between user i and user j in the two-dimensional plane. However, a larger value in the matrix indicates a weaker association between the two users due to their increased distance. To transform this representation into one where higher values

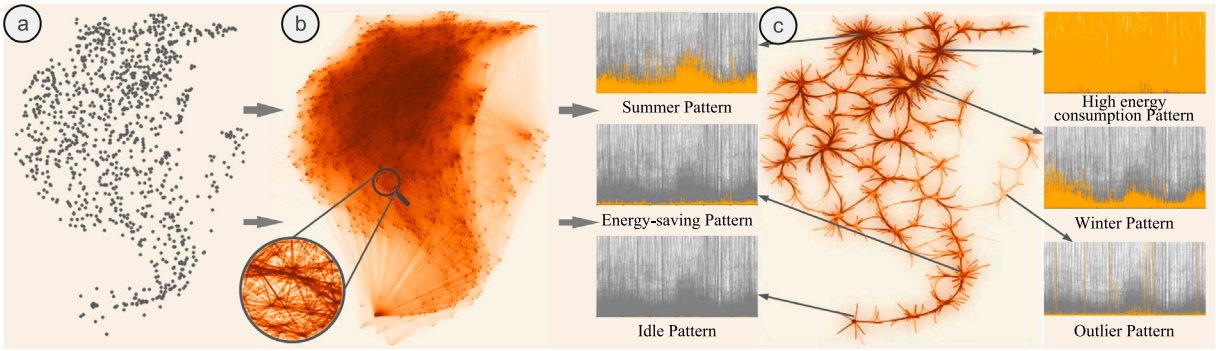


Fig. 4. Temporal correlation extraction between users.

indicate stronger associations, we normalize all the values in the matrix M to the range $(0, 1)$, as shown in the following formula:

$$M_{ij} = \frac{M_{ij} - \min(M)}{\max(M) - \min(M)} \tag{6}$$

where M'_{ij} is the normalized value of the original distance M_{ij}, and $\min(M)$ and $\max(M)$ represent the minimum and maximum values in the matrix M, respectively. The next step is to take the reciprocal of each entry in M' to obtain a new matrix N, where higher values now indicate stronger associations between users:

$$N_{ij} = \frac{1}{M'_{ij}} \tag{7}$$

A threshold τ is then introduced to sparsify the matrix. If any value in N falls below this threshold, indicating insufficient correlation between the corresponding users, it is replaced with 0:

$$A_{ij}^{temporal} = \begin{cases} N_{ij}, & \text{if } N_{ij} \geq \tau \\ 0, & \text{if } N_{ij} < \tau \end{cases} \tag{8}$$

The threshold serves a dual purpose. From an efficiency perspective, pruning the matrix by eliminating weak connections helps reduce computational complexity when performing matrix multiplication or eigenvalue decompositions, which can be computationally intensive on dense matrices. From an accuracy perspective, removing weak associations can help focus analysis on the most significant relationships between users and filter out potentially misleading or irrelevant information, allowing for a clearer understanding of the underlying patterns.

Fig. 4(b) visualizes the constructed temporal correlation network. In Fig. 4(b), visual clutter becomes a serious problem because there are numerous nodes and edges, as shown in the subgraph. Whereas, in energy networks, node positions have well-defined meanings and cannot be modified. Therefore, the edge bundling approach [28] is applied to mitigate the visual clutter and reveal potential energy patterns.

The basic idea of edge bundling is to group and bundle together edges that share similar trajectories. Given two edges e_i and e_j, their similarity s_{ij} can be defined based on various metrics, such as the Hausdorff distance between their trajectories. Mathematically, this can be expressed as:

$$s_{ij} = f(\text{Hausdorff}(e_i, e_j)) \tag{9}$$

where f is a decreasing function of the Hausdorff distance, ensuring that edges with closer trajectories have higher similarity scores. Based on the calculated similarities, edges are clustered using algorithms like hierarchical clustering, DBSCAN, or spectral clustering. The goal is to group geometrically similar edges. Let C represent the set of clusters, where each cluster $c \in C$ contains a set of edges. For each cluster c in C, a representative path or "bundle" is computed. This bundle should approximate the common trajectory of the edges in the cluster. Our approach to computing the bundle is to use a spline fitting method:

$$B_c(t) = \sum_{i=0}^{n} a_i \phi_i(t) \tag{10}$$

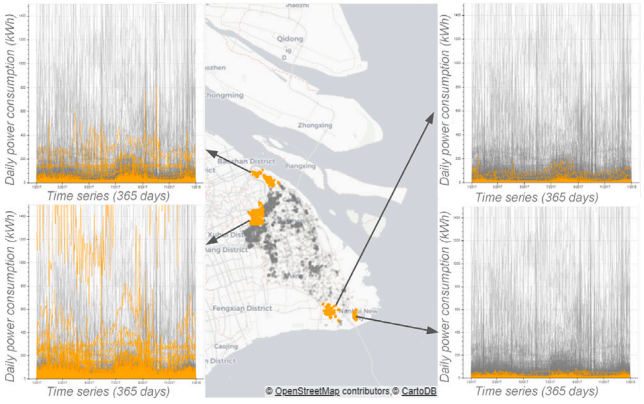


Fig. 5. Spatial correlation extraction between users.

Where $\phi_i(t)$ are basis functions (e.g., B-splines), a_i are the control points, and t parametrizes the bundle path. The control points a_i are optimized to minimize the distance between the bundle and the edges of the cluster.

Finally, the bundles $B_c(t)$ are rendered in the visualization instead of the individual edges, which significantly reduces the visual clutter and allows for a clearer representation of the underlying data structure, as shown in Fig. 4(c). The horizontal axis of each of the six subgraphs corresponds to 365 days (1 year), and the vertical axis represents the power consumption on each day. Those focal points in Fig. 4(c) (orange-red areas) have many users clustered around them and thus can represent common patterns of urban energy consumption. For instance, the most prevalent patterns observed were the summer and winter patterns, which are attributable to the increased energy use for heating during the sun-deprived winters and for cooling during the sweltering summers. At the two extremes of energy consumption lie high energy consumption and energy-saving patterns. Additionally, a significant portion is attributed to the idle pattern, which is likely due to a considerable number of unoccupied houses. The outlier pattern, on the other hand, could stem from equipment malfunctions, human errors, or electricity theft.

The process of spatial feature extraction is similar to that of temporal feature extraction. The main difference is that when constructing the spatial adjacency matrix $A_{ij}^{spatial}$, the values in the adjacency matrix are determined by the distance between the two households (calculated from the GPS coordinates). We also performed normalization and inversion operations on the adjacency matrix, such that the closer the two households are, the stronger their spatial correlation is considered to be (the larger the value in the matrix).

The spatial distribution of users in certain areas is shown in Fig. 5, and a spatial adjacency matrix can be constructed based on the GPS information. As can be seen in Fig. 5, energy behavior also exhibits

aggregation properties at the geographical level. For example, users in the city center demonstrate a tendency to consume more energy (bottom left subfigure), while users in the suburbs generally have lower energy demand (two subfigures on the right).

3.4. Encoding of spatiotemporal correlations

In this section, we elaborate on the encoding process using Temporal Graph Convolutional Networks (TGCN) and Graph Convolutional Networks (GCN) to capture spatiotemporal correlations in users' energy consumption patterns.

Temporal Graph Convolutional Network (TGCN) Encoding

Firstly, we crafted a TGCN model to capture the temporal dependencies inherent in users' energy consumption behaviors by integrating graph convolution and gated recurrent units (GRUs).

At each time step t, the energy data of all users can be expressed as a graph signal $X_t \in \mathbb{R}^{N\times D}$, where N represents the number of nodes (signifying users) and D is the feature dimension (the length of the time series). This graph signal transforms a graph convolution operation at each step to extract crucial information about the users' energy consumption patterns.

Then, we construct an adjacency matrix, $A^{temporal}$, which is based on the observed correlation between users' energy consumption behaviors (refer to Section 3.3). The feature matrix X_t associated with users at any given time t is transformed using the following formula:

$$H_t = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}^{temporal}\tilde{D}^{-\frac{1}{2}}X_t W\right) \tag{11}$$

In the above equation, $\tilde{A}^{temporal} = A^{temporal} + I_N$, where I_N is the identity matrix, ensuring that each node considers its features during convolution. The matrix $\tilde{D}$ represents the degrees of nodes in the enhanced adjacency matrix $\tilde{A}^{temporal}$, while W is a learnable weight matrix that fine-tunes the convolution process. The activation function σ, such as ReLU, introduces non-linearity, essential for capturing complex patterns.

The graph convolution operation can aggregate valuable information from neighboring nodes based on their predefined correlation at $A^{temporal}$. This aggregation is crucial as it helps uncover temporal patterns hidden within users' energy consumption data.

The output from this graph convolution layer, denoted as H_t, is then fed into a GRU, a type of recurrent neural network specifically designed to model temporal dependencies. The GRU operations include the Reset and Update Gates, which control the flow of information within the GRU, deciding which information to retain or discard, and are computed as follows:

$$r_t = \sigma(W_r \cdot H_t + U_r \cdot H_{t-1} + b_r) z_t = \sigma(W_z \cdot H_t + U_z \cdot H_{t-1} + b_z) \tag{12}$$

These gates control the flow of information within the GRU, deciding which information to retain and which to discard. Furthermore, the candidate hidden state is computed, considering the current input and the previous hidden state modulated by the reset gate:

$$\hat{H}_t = \tanh(W_h \cdot H_t + U_h \cdot (r_t \odot H_{t-1}) + b_h) \tag{13}$$

This step computes a candidate's hidden state, considering the current input and the previous hidden state, modulated by the reset gate. Lastly, the hidden state is updated, incorporating elements from both the previous hidden state and the candidate hidden state, as determined by the update gate:

$$H'_t = (1 - z_t) \odot H_{t-1} + z_t \odot \hat{H}_t \tag{14}$$

The final hidden state is a blend of the previous state and the candidate state, controlled by the update gate. This mechanism allows the GRU to effectively capture the sequential dependencies that are key to understanding users' energy consumption patterns over time.

After processing the entire sequence through the GRU, we obtain the final hidden state, H'_T. This state encapsulates the rich temporal features learned from the evolving spatiotemporal correlations in the users' energy consumption data. To extract a meaningful temporal encoding representation, $V_{temporal}$, for the entire sequence, we use the following transformation:

$$V_{temporal} = W_o \cdot H'_T + b_o \tag{15}$$

In this way, $V_{temporal}$ encapsulates the temporal features learned from the evolving spatiotemporal correlations present in users' energy consumption data. This comprehensive summary of dynamic energy consumption behaviors observed across all users serves as valuable input for further analysis.

Graph Convolutional Network (GCN) Encoding

The GCN takes as input the same historical energy consumption information used by TGCN. However, unlike TGCN, GCN focuses on spatial relationships between nodes. To achieve this, we compute the normalized adjacency matrix $\hat{A}^{spatial}$ from the spatial distance matrix $A^{spatial}$:

$$\hat{A}^{spatial} = D_2^{-\frac{1}{2}} A^{spatial} D_2^{-\frac{1}{2}} \tag{16}$$

In this formula, D_2 is the diagonal matrix of degrees corresponding to $A^{spatial}$. This normalization ensures that the influence of each node on its neighbors is weighted based on their connectivity, providing a more balanced representation of spatial relationships. The graph convolution operation for each node i is then performed as follows:

$$H_i = \sigma\left(\sum_{j\in\mathcal{N}(i)\cup\{i\}} \frac{1}{\sqrt{\hat{d}_i \hat{d}_j}} X_j W\right) \tag{17}$$

Here, H_i represents the hidden representation of node i after applying the graph convolution. The function σ is a non-linear activation function that introduces nonlinearity to the model, allowing it to capture more complex patterns. The set $\mathcal{N}(i)$ denotes the neighboring nodes of i, while X_j is the feature vector associated with node j. The trainable weights W of the graph convolution layer are learned during the training process to optimize the model's performance. The normalized degrees $\hat{d}_i$ and $\hat{d}_j$ balance the influence of each neighboring node based on their connectivity. Extending this operation to all nodes simultaneously, we obtain the overall graph convolution expression:

$$Z = \sigma(\hat{A}^{spatial} X W_s) \tag{18}$$

In this formula, $W_s \in \mathbb{R}^{F\times F''}$ represents the trainable spatial graph convolutional weights. The output feature matrix $Z \in \mathbb{R}^{N\times F''}$ captures the spatial patterns and dependencies among nodes after applying the graph convolution. The non-linear activation function σ adds complexity to the model, enabling it to learn more intricate spatial relationships.

Finally, we introduce a linear transformation to encapsulate the spatial patterns and dependencies among nodes. Specifically, after obtaining the node features $Z_i, i = 1, 2, \ldots, N$ from the final layer of the Graph Convolutional Network (GCN), we compute a representative feature Z. To derive the unified graph-level representation $V_{spatial}$, we then apply a linear transformation:

$$V_{spatial} = W_g \cdot Z + b_g \tag{19}$$

Here, W_g and b_g are learnable parameters that are optimized during training. By introducing this linear layer, the model can learn to extract the most relevant information from Z to form a meaningful graph representation. This approach allows the model to capture the rich spatial characteristics encoded in the GCN's node features.

By combining the encodings V_1 from the TGCN and $V_{spatial}$ from the GCN, we obtain a comprehensive representation that captures both temporal and spatial correlations in users' energy consumption patterns. $V_{temporal}$ reflects the dynamic temporal evolution influenced by user behaviors, while $V_{spatial}$ encapsulates spatial dependencies shaped

by geographical factors. The comprehensive encoding process offers a robust framework for understanding and exploiting spatiotemporal correlations in energy consumption data, which is vital for applications such as demand forecasting and energy management.

The significance of spatiotemporal features

As can be seen from the above process, the historical spatiotemporal correlations among users are fed as features to the prediction model to enhance accuracy. An important prerequisite for the above process is that a user's energy consumption is not only influenced by their historical consumption patterns but also by the consumption patterns of surrounding users, i.e., the spatial–temporal correlation hypothesis (cf. Section 3.1).

From the perspective of conditional probability theory, the spatiotemporal correlation hypothesis is reasonable, even in the absence of a direct causal relationship between their energy behaviors (that is, the energy consumption of User A does not physically affect the behavior of User B). Conditional probability allows us to quantify the likelihood of an event given the presence of another related event. The essence of this approach is that the similarity in energy consumption patterns between Users A and B serves as a form of "conditional" information. If the correlation is strong, the probability distribution of User A's future energy consumption is influenced by the observed data of User B. Mathematically, this can be expressed as:

$$P(X_A|X_B) \neq P(X_A) \tag{20}$$

Here, $P(X_A|X_B)$ represents the conditional probability of User A's energy consumption given User B's energy consumption, while $P(X_A)$ represents the unconditional probability of User A's energy consumption. The inequality indicates that knowing User B's energy consumption history alters the probability distribution of User A's energy consumption, indicating a dependency between the two users.

This does not imply causation but indicates a conditional dependency. Essentially, this method leverages the concept of conditional probability to enrich the dataset with patterns that might not be evident from User A's data alone but emerge from the aggregated behavior observed across multiple users with similar consumption patterns.

3.5. User-level energy demand forecasting

In this section, we delineate the execution steps of the Transformer model for user-level energy demand forecasting. The Transformer model is suitable for temporal prediction tasks due to its ability to capture complex dependencies across sequences. The self-attention mechanism allows the model to attend to different parts of the input sequence simultaneously, enabling it to understand and model temporal relationships effectively, especially those requiring the identification of long-term dependencies and complex patterns in time series data [29].

Our model incorporates two key encoded vectors: $V_{temporal}$ and $V_{spatial}$, derived from the Temporal Graph Convolutional Network (TGCN) and Graph Convolutional Network (GCN), respectively. These vectors encapsulate historical energy consumption patterns and geographical proximity information. We start with the utilization of $V_{temporal}$ within the transformer encoder, aiming to capture temporal dependencies in energy consumption behaviors.

The encoder begins by employing a self-attention mechanism on $V_{temporal}$, enabling the model to weigh different temporal dependencies dynamically. This operation calculates attention scores by projecting V_{temporal} into queries, keys, and values:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{21}$$

Then, the self-attention operation is repeated with multiple sets of parameters, facilitating the capture of diverse temporal dependencies. The results are concatenated and linearly transformed:

$$MultiHead(Q,K,V) = Concat(\text{head}_1,\ldots,\text{head}_h)W \tag{22}$$

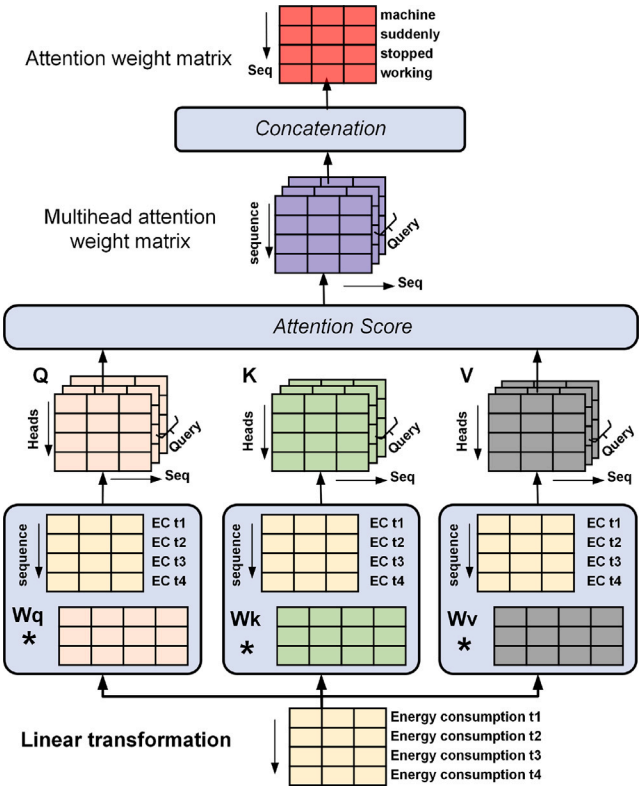


Fig. 6. Procedure of encoder operation.

Following attention computation, a position-wise feed-forward network is applied to introduce non-linearity and enhance the model's capacity to capture complex temporal patterns.

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{23}$$

To ensure effective gradient flow and stabilize training, residual connections and layer normalization are applied after each sub-layer, which can be mathematically represented as:

$$y = LayerNorm(x + SubLayer(x)) \tag{24}$$

In this equation, y represents the output after applying the residual connection and layer normalization, x is the input to the sub-layer, $SubLayer(x)$ denotes the processing function of the sub-layer, and $LayerNorm$ is the layer normalization operation. This equation describes the process of passing the input x through the sub-layer, adding the processed result to the original input x to form a residual connection, and finally applying layer normalization to the entire result. The entire procedure of the encoder operation is shown in Fig. 6.

Multiple layers of the aforementioned operations are stacked to capture intricate temporal dependencies effectively:

$$Encoder(V_{temp}) = LayerNorm(MultiHead(LayerNorm(V_{temp}))) \tag{25}$$

Moving to the Transformer Decoder, we integrate the encoded vector $V_{spatial}$ from the GCN to predict future energy consumption patterns while considering geographical distance:

Similar to the encoder, the decoder applies self-attention with masking to ensure causality, preventing future information from influencing predictions prematurely.

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}} + Mask\right)V \tag{26}$$

Besides self-attention, the decoder also attends to the encoder's output to incorporate relevant temporal information into the forecasting

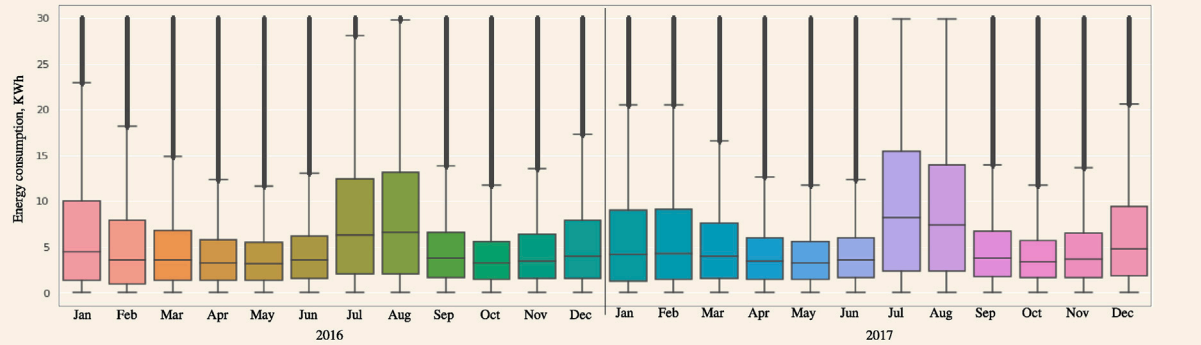


Fig. 7. Energy consumption data for Shanghai.

process.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{27}$$

Lastly, a linear transformation followed by softmax generates the probability distribution over future energy consumption levels for each user, leveraging both temporal and spatial features. By applying the softmax function across the temporal dimension, we obtain a probability distribution over future energy consumption levels for users. Mathematically, this can be represented as:

$$\hat{Y} = softmax(V_{spatial}W + b) \tag{28}$$

Where $\hat{Y}$ is a matrix of size $n \times T$, with T being the number of future time steps being predicted. Each row of $\hat{Y}$ corresponds to the predicted energy consumption distribution for a particular user over the T future time steps. The matrices W and b represent the weights and biases of the linear transformation, respectively.

Each step in the Transformer model contributes significantly to capturing the complex dependencies within the energy consumption data while integrating spatial information, thus enabling accurate forecasting at the user level. Through the fusion of temporal and spatial features, our model offers a holistic approach to user-level energy demand forecasting.

4. Experiment and evaluation

In this chapter, we first introduce the dataset of this paper, then we introduce the comparison model and parameter settings, and finally, we analyze the experimental results.

4.1. Description of the dataset

All energy data used in this study comes from the Shanghai Municipal Electric Power Bureau, including electricity consumption data from about 10,000 households in Shanghai from 2016 to 2018. The dataset is usually a detailed time series of peak and trough values at daily granularity. In this experiment, we used the 2017 energy consumption data to train the network parameters and used the 2018 energy consumption data as a test dataset to evaluate the performance and predictive ability of the model. The electricity consumption data of Shanghai in 2016 and 2017 is shown in Fig. 7.

Fig. 7 visualizes the energy consumption patterns in Shanghai from January 2016 to December 2017. The data is represented through a series of box plots, with each month corresponding to a specific box plot derived from the energy consumption statistics of 10,000 users for that particular month. Several notable patterns emerge from this visualization. Firstly, the energy consumption in December, January, July, and August is significantly higher than in the other months of the year. This can be attributed to various factors. December and January, being the colder months, see an increase in energy usage due to heating needs, especially in a city like Shanghai where winters can be chilly. On the other hand, the elevated consumption in July and August is undoubtedly linked to the widespread use of air conditioning during the hot summer months. Another interesting observation is that the upper limits of the box plots for July and August are distinctly higher than those of January and December, which suggests that during the summer peak, there are more extreme energy users due to increased reliance on air conditioning in hotter weather.

4.2. Parameters and evaluation metrics

In this study, we introduced the spatio-temporal feature fusion model (ST Fusion Model), which leverages GCN and TGCN to extract spatial and temporal information from energy data. To validate its efficacy, we compared it against six traditional approaches. To be specific, the HA Model predicts future values based on historical averages. It does not require complex parameter tuning, making it a straightforward benchmark. ARIMA Model [30] captures temporal patterns effectively with an autoregression order of 2, a differencing degree of 1, and a moving average order of 1. We also used an SVR [31] model with a regularization parameter set to 100, balancing model complexity and training errors. The kernel function chosen was RBF, and the penalty term is 0.001. The GCN [32] model featured two graph convolutional layers, with the first layer utilizing 64 filters and the second layer employing 32 filters. This configuration allowed for effective capture of spatial dependencies within the network. GRU Model [33] consisted of two layers, each with 128 GRU units, and a learning rate of 0.001, optimized for capturing temporal dependencies. The LSTM model [34] was configured with two layers, 128 LSTM cells in each, a learning rate of 0.001, and the dropout rate for each LSTM function was set to 0.3 to avoid over-fitting. This setup was chosen to efficiently handle long-term dependencies.

The proposed TGCN model incorporates specific parameters to optimize its performance [35]. The learning rate is set to 0.001 for efficient gradient descent, the batch size is configured to 64 for effective computation, and the model undergoes 3000 training epochs for thorough learning. Additionally, the hidden units are set to 100, allowing the model to capture complex patterns and relationships within the data. An Adam optimizer [36] was also used to train our model. To avoid gradient explosion, we set the maximum gradient to 2. To quantify the performance of the models, we applied four indicators: root-mean-square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R squared (R^2). All evaluations

Table 1
Prediction results of LSTM algorithm in different patterns.

T	Metric	HA Model	ARIMA Model	SVR Model	GCN Model	GRU Model	LSTM Model	TGCN	ST Fusion Model
10 days	RMSE	1.482881	1.599869	1.379108	1.772607	0.831712	0.903139	0.739673	0.631355
	MAE	1.145063	1.166924	1.034687	1.307548	0.632416	0.686895	0.571507	0.490554
	MAPE	13.24871	13.20661	11.62666	14.5131	7.250702	7.919646	6.542047	5.662308
	R2	0.74873	0.707519	0.782667	0.640951	0.820955	0.806795	0.837481	0.854451
	Training time (s)	/	/	/	7237	1657	1825	9221	14 560
30 days	RMSE	3.228759	2.930029	2.667984	3.323084	1.634792	1.926825	1.701278	1.510002
	MAE	2.413861	2.205412	1.947403	2.543281	1.205457	1.459668	1.357918	1.093568
	MAPE	26.55238	25.66597	21.74066	28.5795	13.57049	16.71365	15.41527	12.40923
	R2	−0.19124	−0.26186	0.186617	0.68992	0.594611	0.575758	0.669266	0.739454
	Training time (s)	/	/	/	14 560	2203	2386	12 488	19 706

were conducted on a consistent hardware setup: Ubuntu 18.04.6 LTS, Intel(R) Xeon(R) Gold 6248 CPU, with a Tesla-V100 GPU and 128 G of memory size. These indicators can be defined as follows:

$$\begin{aligned} \mathrm{RMSE} &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i' - y_i\right)^2} \\ \mathrm{MAE} &= \frac{1}{N}\sum_{i=1}^{N}\left|y_i' - y_i\right| \\ \mathrm{MAPE} &= \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i'}{y_i}\right| \times 100\% \\ R^2 &= 1 - \frac{\sum_{i=1}^{N}(y_i - y_i')^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \end{aligned} \tag{29}$$

Where N is the length of the time series, y_i is the actual data, and y_i' is the predicted value. The experimental results are shown in Table 1.

4.3. Experimental results

The experimental results, as shown in the table, depict the performance of various energy demand forecasting models across different prediction horizons: 10 days and 30 days. With the experiments in this paper, we verify the generalization ability of the model (generalization hypothesis). In the 10 day forecast scenario, the ST Fusion Model demonstrates exceptional accuracy, showcasing the lowest RMSE (0.631) and MAE (0.490) among all models. This superior accuracy is complemented by reduced errors and enhanced precision, as evidenced by the lowest MSLE and MAPE. The ST Fusion Model also achieves the highest R2 score (0.854), indicating its proficiency in capturing intricate spatio-temporal patterns effectively.

When extending the forecast horizon to 30 days, the ST Fusion Model maintains its superiority over alternative methods. Despite the increased prediction horizon, it sustains superior accuracy with the lowest RMSE (1.510) and MAE (1.094) compared to traditional and neural network-based approaches. Consistently, the model exhibits the lowest MSLE and MAPE, indicative of robust performance over an extended period. Notably, the ST Fusion Model achieves an impressive R2 score of 0.739454, affirming its capability to capture complex spatio-temporal dynamics inherent in energy consumption patterns.

In comparison to other models, the HA Model demonstrates the poorest performance, as expected, due to its simplistic nature of forecasting based solely on historical averages. The ARIMA Model, while capturing some temporal patterns, struggles to adapt to the complex dynamics of energy consumption. The SVR Model, while showing competitive results, falls short in capturing intricate patterns embedded within the data due to its linear nature. Although the GCN Model is adept at capturing spatial correlations, its inability to capture temporal features results in less-than-optimal performance.

Among the neural network-based models, the GRU model and the LSTM model perform reasonably well, especially in capturing temporal dependencies. The LSTM Model, known for its ability to capture long-term dependencies, exhibits competitive performance but is surpassed by the ST Fusion Model, indicating the importance of integrating spatial features. The GRU model lacks explicit mechanisms to incorporate spatial correlations among energy consumption patterns. Unlike the ST Fusion Model, which integrates Graph Convolutional Networks (GCN) to capture spatial dependencies, the GRU Model primarily focuses on temporal relationships within the data. This limitation impedes the GRU Model's ability to discern spatial influences on energy consumption, particularly in scenarios where geographical proximity plays a significant role. The LSTM Model, known for its ability to capture long-term dependencies, exhibits competitive performance but is surpassed by the ST Fusion Model. Similar to the GRU model, the LSTM model lacks explicit mechanisms to incorporate spatial correlations among energy consumption patterns. It primarily focuses on capturing temporal relationships within the data, neglecting important spatial influences that may impact energy consumption dynamics.

The superior performance of the ST Fusion Model can be attributed to its unique fusion of spatial and temporal features facilitated by GCN and TGCN modules, respectively. By integrating graph convolutional networks for spatial correlations and temporal graph convolutional networks for temporal dependencies, our model accurately captures the intricate dependencies inherent in energy demand. This comprehensive approach enhances predictive capacity, enabling more accurate and reliable forecasts compared to traditional methods and single-modal neural network architectures. Moreover, the consistent performance of the ST Fusion Model across both short-term (10 days) and long-term (30 days) forecasts underscores its durability and dependability. This consistency is vital in energy demand forecasting, where precise predictions over extended periods are indispensable for efficient resource allocation and decision-making processes.

We also analyzed the energy consumption patterns of six selected users from a multi-user energy dataset, as shown in Fig. 8. These users exhibit distinct consumption behaviors, categorized into three main patterns: high energy consumption, bimodal patterns, and energy-saving patterns.

Two users demonstrated a high energy consumption pattern, characterized by a consistent daily energy consumption of around 200 kWh. The predictive accuracy for these users is notably high. This can be attributed to the stability in their daily energy usage, likely stemming from a diverse range of energy sources contributing to their consumption. Consequently, their energy consumption exhibits less variability, leading to more accurate predictions. Three other users follow a "bimodal pattern", characterized by higher energy use during the summer and winter seasons and lower consumption during the spring and fall. Their daily energy consumption ranges from 4 to 30 kWh. While the seasonal variations pose a challenge for the prediction model, the algorithm still manages to provide reasonably accurate forecasts, albeit with slightly more variance compared to the high-consumption users.

One user displays an energy-saving pattern characterized by daily energy consumption not exceeding 1 kWh. Our prediction accuracy suffers the most in this case. The reason behind this reduced accuracy lies in the high variability of energy use from a percentage standpoint. Small absolute changes in energy consumption represent significant percentage shifts, making it difficult for the algorithm to capture these

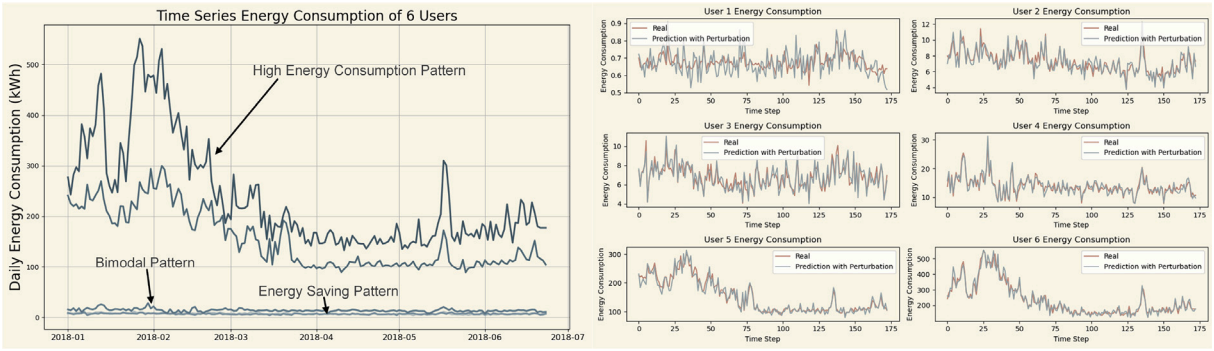


Fig. 8. Case studies of predicted results.

fluctuations accurately. From a probabilistic perspective, the energy-saving user's consumption behavior is highly stochastic. Minor changes in daily activities, such as turning on the air conditioning for a longer duration, can lead to a several-fold increase in energy use for that particular day. This unpredictable nature of energy consumption poses significant challenges for any prediction model.

The observed differences in predictive accuracy among the three patterns stem from the inherent characteristics of each consumption pattern. Users with stable and consistent energy consumption behaviors, such as those in the High Energy Consumption Pattern, exhibit higher predictive accuracy due to the reduced variability in their data. Conversely, users with erratic and unpredictable energy usage behaviors, like those in the Energy Saving Pattern, present challenges for accurate prediction due to high randomness and variability in their consumption data.

To assess the complexity of our models, we have used the training time required for the models to reach convergence as a metric. This metric provides a reliable estimate of the complexity, considering that the training process encapsulates the overall computational workload of the model, as shown in Table 1. While our proposed model requires the longest training time, it offers advantages in terms of prediction accuracy and interpretability. Regarding interpretability, our model not only outputs energy prediction results but also identifies potential energy behavior patterns, which allows domain experts to gain a deeper understanding of the system and make informed decisions accordingly. Moreover, we believe that the trade-off between training time and model performance is reasonable in our context. The longer training time allows our model to learn more complex patterns from the data, resulting in improved accuracy and interpretability. This advantage outweighs the potential disadvantage of a longer training time, especially when considering the benefits it brings to decision-making and system optimization.

5. Visual encoding and analytics

In the field of energy demand management, experts and decision-makers strive to quantify energy consumption, comprehend locational disparities, and discern consumer preferences through behavioral analysis. Nevertheless, a significant portion of the research in this domain has predominantly centered on examining temporal variations in energy consumption within a specific region, ignoring the exploration of energy demand at different spatial and temporal scales [37]. To this end, we developed a demand-side energy management visualization interface by encoding energy data on a variable spatio-temporal scale, as shown in Fig. 9.

5.1. Modeling of spatio-temporal fluctuations of energy consumption

Fig. 9c depicts the spatio-temporal fluctuations of urban energy demand in a flow field diagram. To be specific, we observed that geospatial energy demands evolve continuously over time, forming a simply connected, irrotational continuum in the time dimension. Drawing inspiration from fluid dynamics and continuum mechanics, we propose modeling this continuum as a potential flow. Formally, the spatiotemporal demand shift, denoted as $\upsilon = \nabla\varphi$, represents the flow velocity υ — a vector field quantifying the rate of demand change. Here, φ signifies the spatial energy demand variation across selected moments or periods.

To model spatial energy demand, we introduce a kernel density estimation-based approach. This method transforms discrete household energy consumption data into a continuous representation, facilitating the generation of a smoother vector field. Specifically, our model is defined as $\hat{\phi}_h(x) = \sum_{i=1}^{n} c_i K_H\left(x - x_i\right)$, where x_i represents household locations, c_i is the normalized energy demand used to adjust demand strength, H is the bandwidth (or smoothing) matrix — a symmetrical and positive definite $d \times d$ matrix, and K is the kernel function. In our study, we opted for the Gaussian kernel. The spatial energy demand change φ is then modeled as the difference between the estimated $\hat{\phi}$ at different time points.

Fig. 10 illustrates the above modeling process intuitively. Fig. 10a, b visualize historical energy consumption data and future energy demand, respectively. The subplots in the lower left corner show the results of discrete energy data smoothed by kernel density estimation. Fig. 10c exhibits the spatio-temporal fluctuations of urban energy demand based on potential flow modeling. The vector field map depicted in Figg 10c illustrates the spatiotemporal demand-shift model, comparing future energy demands to a specific past period. In this map, the vectors indicate the direction of the energy demand shift, while the magnitude of each vector signifies the intensity of this shift. In addition, we divided the entire region into zones based on rectangles. Then, we quantitatively characterize the intensity of fluctuations in energy demand within each zone by calculating the fluxes of the fields within each zone separately and encoding the intensity with colors (the darker the color, the stronger the energy fluctuations). For example, the dark red rectangular area in Fig. 10c (high-density residential area, Sanlin Town) has the largest fluctuations in energy demand (over 80,000 kWh). Energy experts can further observe the fluctuation of energy demand flowing in different directions through the zoom-in function to anticipate potential load fluctuations. Fig. 10d displays the fluctuations in energy demand between different quarters.

5.2. Case study — Mental inertia in energy behavior

Figs. 10a, b visualize the time-series characteristics of energy demand from different perspectives. Fig. 10a presents the distribution of daily energy consumption and energy consumption per user in the region. Fig. 10b exhibits temporal energy demand and furnishes key interactive functionalities for spatial demand and demand-shift analyses. Specifically, the temporal energy demand view employs a stream graph to visualize peak and trough energy demands. Users can activate the

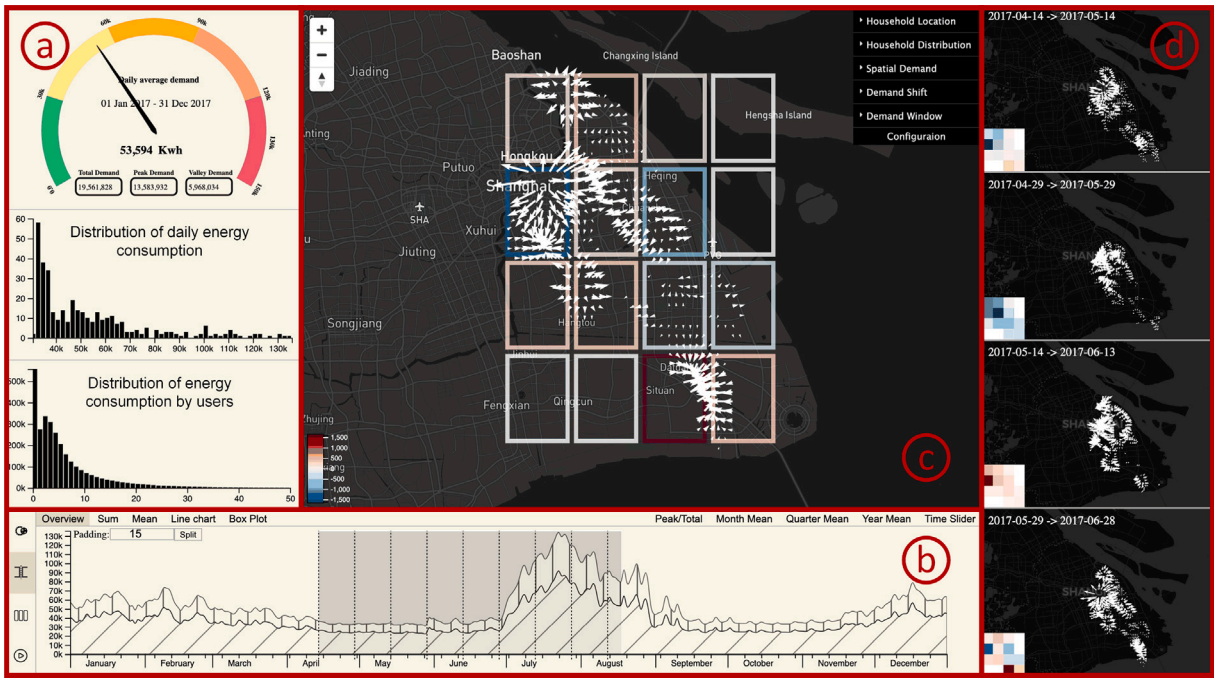


Fig. 9. Case studies of predicted results.

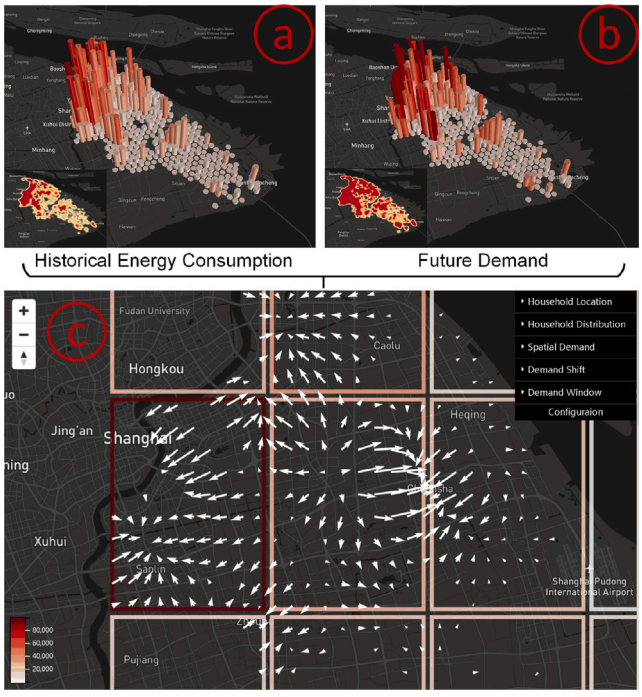


Fig. 10. Case studies of predicted results.

auxiliary analysis line (displaying yearly, quarterly, and monthly average demands and peak-to-trough ratios) to pinpoint a specific period for deeper temporal analysis. The demand-shift analysis commences via the functional buttons. Once a user identifies a period of interest through the preceding steps, they can specify various demand-shift exploration tasks across diverse periods: peak-trough, regularly divided intervals, or multiple periods. Users can toggle the relevant button, select their desired period(s) by brushing, and then activate the compute button to produce and enumerate the results in the demand-shift visual index.

In empirical studies, we revealed a distinct bimodal pattern, with a significant demand surge in summer and a lesser one in winter. This trend aligns with the fact that in Shanghai, electricity consumption for air conditioning in the summer exceeds that for heating in the winter. Upon examining the peak-to-total ratio curve, we made the following discoveries: (1) A weekly periodic pattern, implying that the life pattern of the local population is cyclical, i.e., working on weekdays and resting on weekends. (2) A notably lower peak-to-total ratio in summer compared to other seasons indicates that the power system is operating at full capacity, prompting maintenance engineers to be vigilant for any potential malfunctions. (3) Despite similar temperatures in Shanghai during May, June (24 °C–28 °C), September, and October (27 °C–23 °C), we noticed that total energy demand in the latter period was approximately 200,000 kWh higher. This suggests that the general population is influenced by a strong mental inertia in their energy behavior.

Mental inertia refers to the tendency of individuals to maintain their current behaviors or habits, even in the face of changing environmental conditions. As the weather cools down slightly in September and October, despite the similar temperatures, people may still maintain their cooling habits due to psychological inertia. They may continue to use air conditioning or other energy-consuming devices at the same level as during the hotter months, resulting in higher energy demand during this period.

The above phenomenon can be explained by the following theories. The perceived effort required to make changes to energy consumption behaviors, such as adjusting thermostat settings or adopting energy-saving practices, may outweigh the perceived benefits for some individuals. This cost–benefit analysis often leads to inertia, as individuals prioritize maintaining the status quo over investing time and effort on behavioral adjustments. Additionally, psychological inertia may also be influenced by social norms and expectations. If it is socially acceptable or even expected to use air conditioning during these months, individuals may feel pressure to conform to these norms, further driving up energy demand.

The discovery of the influence of psychological inertia on energy consumption patterns has significant implications for energy efficiency and conservation efforts. It underscores the importance of raising awareness among the public about the impact of their daily habits

on energy usage and the environment. By understanding the role of psychological inertia, we can design more effective energy conservation campaigns that encourage people to overcome their ingrained behaviors and adopt more sustainable practices.

By recognizing the role of psychological inertia, policymakers and energy stakeholders can design more effective strategies to encourage behavior change. These strategies may include targeted education campaigns aimed at raising awareness about the impact of energy consumption habits and providing practical tips for reducing energy usage. Additionally, implementing incentives such as rebates or rewards for adopting energy-saving behaviors can help motivate individuals to overcome inertia and take action. Moreover, technological innovations such as smart energy monitoring systems can provide real-time feedback on energy usage, empowering individuals to make informed decisions and track their progress toward energy-saving goals. By combining technological solutions with targeted behavioral interventions, stakeholders can effectively address psychological barriers and promote a culture of sustainable energy consumption.

5.3. Case study — Demand side management

In the context of a growing energy demand, the management of electricity consumption has become a critical issue. Particularly, peak load periods pose significant challenges to the stability and efficiency of the power grid. The objective of this case study is to estimate the potential for shifting peak electricity consumption to off-peak periods, thereby reducing the overall peak load and enhancing grid stability.

To achieve this goal, we propose a data-driven approach that leverages historical electricity consumption data from 10,000 users in Shanghai. For each user, we plot a histogram of their daily electricity usage data, with the horizontal axis representing the range of electricity usage and the vertical axis representing the number of days, as shown in Fig. 11. As can be seen in Fig. 11, the distribution of energy demand exhibits a normal distribution characteristic, i.e., there are fewer days with high and low energy consumption, with the majority of days having a centralized energy demand within specific intervals (the intervals vary for different energy consumption patterns). In addition, users in summer/winter patterns and energy-saving patterns exhibit significant bimodal characteristics. The presence of two peaks in the electricity consumption distribution suggests that there are two periods within the year where energy demand is significantly higher. This can be attributed to the increased use of air conditioning systems during the hot summer months and heating systems during the cold winter months (refer to Section 3.3). The higher energy consumption during these seasons is a direct response to the need for temperature regulation in homes and workplaces. In contrast, the single-peak distribution observed among high-energy-consuming users indicates a more consistent high level of energy use throughout the year, possibly due to larger living spaces, higher energy-consuming appliances, or less efficient energy management practices.

In addition, we also noticed a distinctive long-tail characteristic in the distribution of daily electricity consumption. The long-tail characteristic indicates that while most days are associated with a relatively average or moderate level of electricity consumption, there is a small but consistent number of days when users consume significantly higher amounts of electricity. This characteristic is particularly evident among users in the high-energy consumption pattern, suggesting that the occurrence of high-usage days is not an isolated event but a recurring pattern.

The long-tail days with exceptionally high electricity consumption suggest that there are inefficiencies or unnecessary usage within the energy-consuming systems. These spikes could be due to outdated equipment, inefficient operational practices, or a lack of proper energy management strategies. Therefore, the high-energy-consuming users represented in the long tail offer significant potential for energy transfer or optimization. For instance, these users may be able to utilize energy-efficient technologies or equipment that reduces their overall energy demand. Additionally, strategies such as demand response or energy storage could be implemented to shift peak demand to off-peak hours, reducing the need for additional generation capacity and improving grid stability.

To quantify the potential for peak transfer, we fit a normal distribution curve to the histogram. This involves estimating the mean μ and standard deviation σ of the user's daily electricity consumption. Then, we defined a threshold for identifying days with excessive electricity usage. After discussions with energy experts, we propose using a threshold at $\mu + 2\sigma$, which corresponds to the point where the probability of exceeding this value is less than 5% under the normal distribution assumption, as shown in Fig. 11. For each identified day, we calculate the excess consumption as the difference between the daily consumption and the mean consumption μ. This excess consumption is assumed to be the amount that can potentially be shifted to off-peak times. Finally, we sum up the excess consumption of all users to estimate the total potential load that can be shifted from peak to off-peak periods.

In our empirical study, we calculated the energy transfer potential of different patterns. We found that the higher the energy consumption, the greater the potential for energy transfer. Specifically, users in the idle pattern have no energy transfer potential because they do not consume any energy. Users in the energy-saving pattern have an energy transfer potential of 5.89%, meaning that 5.9% of their total energy consumption can be transferred. Users in summer and winter patterns have the potential for energy transfer of 7.41% and 7.26%, respectively. Moreover, users in the high-energy consumption pattern have an energy transfer potential of 11.08%. Overall, the total energy transfer rate in Shanghai is 10.21%. This is approximately equivalent to expert estimates (previous studies have shown that about 10%–15% of Shanghai's energy consumption is wasted).

According to the statistics, the average daily electricity consumption of 10000 users in Shanghai is 84,383.55 kWh, and the highest daily electricity consumption equals 176,674.89 kWh, i.e., the peak-to-average ratio is 2.094. The purpose of the energy transfer is to shave the peaks and fill in the valleys. To calculate the peak-to-average ratio reduced after implementing demand management, we consider two scenarios, i.e., an optimistic schema and a conservative schema. In the optimistic schema, we assume that the energy transfer potential can be saved, i.e., we only implement peak shaving (without valley filling). In this schema, the peak-to-average ratio is 1.88. In the conservative schema, we keep the total energy demand constant and shift the peak energy to an equal number of valley days (peak shaving and valley filling). In this case, the peak-to-average ratio is 1.79.

The optimistic schema of demand management prioritizes peak shaving, aiming to reduce energy demand during peak hours. One effective approach is to raise awareness through education or propaganda that highlights the importance of energy conservation. Incentive-based policies, such as tax breaks or rebates for energy-efficient appliances, can also encourage residents to adopt more efficient technologies. In this scenario, the total energy demand is reduced. However, aggressive peak shaving measures may inconvenience consumers, who may need to reduce energy usage at inconvenient times. The conservative schema of demand management takes a more balanced approach, aiming to optimize grid efficiency by balancing peak shaving and valley filling. For example, energy companies can offer time-of-use pricing, where electricity costs more during peak hours, incentivizing consumers to shift their energy usage to off-peak times. This schema offers several benefits. It improves grid utilization by shifting some load to off-peak hours, reducing the waste of renewable energy resources. Additionally, it allows consumers to maintain their energy usage patterns while still contributing to demand management, ensuring consumer convenience. However, the conservative mode may achieve a less significant reduction in peak demand compared to the optimistic mode.

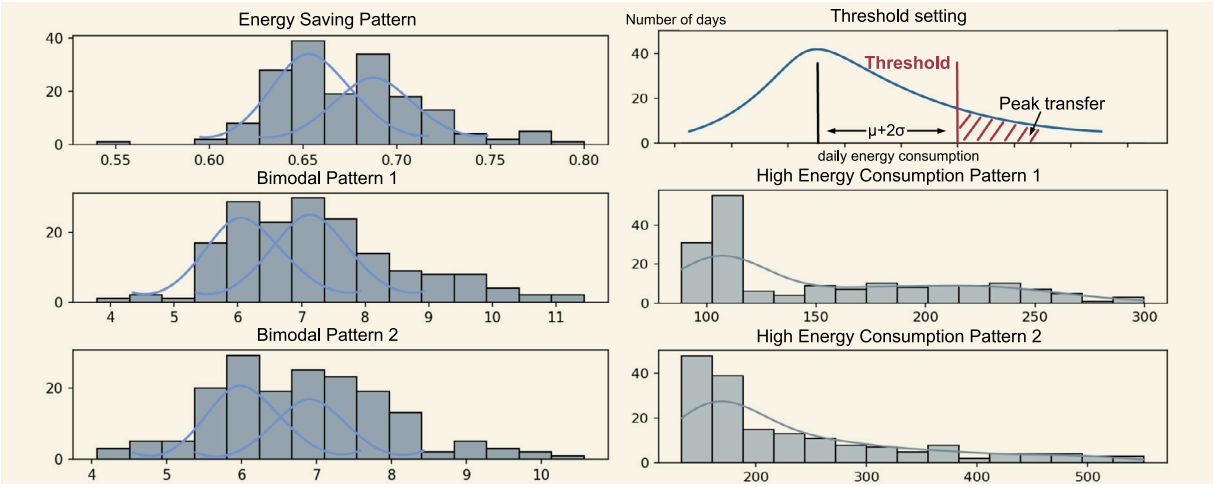


Fig. 11. Case study of demand side management. The horizontal coordinate is the range of daily energy consumption and the vertical coordinate is the number of days. The figure demonstrates the distribution of daily electricity consumption for several typical patterns in a statistical histogram.

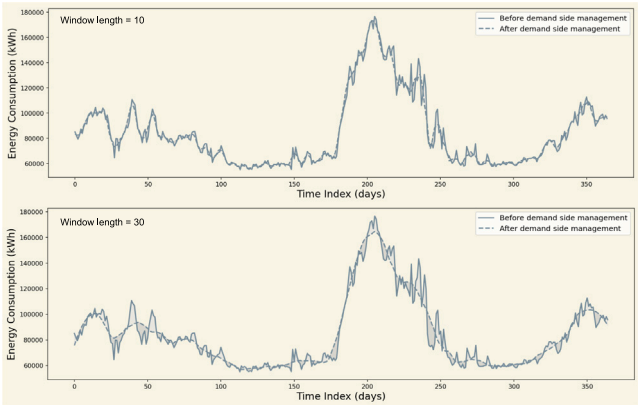


Fig. 12. Comparison of energy consumption before and after demand side management.

In practical situations, it is plausible that both the optimistic and conservative scenarios may occur concurrently, meaning that a portion of peak energy consumption is conserved while another portion is shifted to off-peak hours or redistributed across different time periods. By implementing a combination of peak shaving and valley filling, the energy grid can achieve a more stable and efficient operation. The optimistic measures can lead to a reduction in overall energy consumption and lower peak loads, which is beneficial for the environment and can help reduce energy costs in the long run. On the other hand, by shifting some of the load to off-peak hours, the grid can better utilize its capacity and reduce the need for expensive peak power generation. This can also help to integrate renewable energy sources more effectively, as excess energy produced during off-peak hours can be stored and used later. This dual strategy allows for greater flexibility in managing energy demand, accommodating the varying needs and preferences of consumers. It recognizes that not all consumers may be able or willing to change their energy usage patterns significantly and thus provides a more inclusive approach to demand management.

The statistical histogram of energy consumption data over a specific period is initially utilized to estimate the potential for energy shifting, with the length of the time window significantly influencing the estimation results. As presented in Fig. 12, we have considered two scenarios: 10 day and 30 day windows. The figure depicts the energy consumption before and after energy demand management, with solid and dashed lines, respectively. The light and dark gray areas represent the transferable energy and its shifted positions. Notably, the 30 day time window reveals a greater potential for energy shifting compared to the 10 day window.

6. Conclusions

In conclusion, this study delves into the intricate dynamics of urban energy consumption patterns, emphasizing the critical role of accurate demand forecasting in enhancing the resilience and efficiency of urban energy systems. By analyzing historical data and identifying spatio-temporal behavioral patterns, we reveal insights into the drivers of energy demand and propose innovative solutions for demand-side energy management.

- We propose a spatio-temporal pattern mining scheme for energy behavior, which enables the quantification of energy consumption behaviors. The scheme intuitively demonstrates implicit energy behavior patterns from both time series and spatial distribution perspectives. We employ energy behavior patterns as a priori knowledge for improving the performance of predictive models. Specifically, we reveal six basic energy consumption patterns. These patterns enable us to identify urban energy consumption behavioral mechanisms and target optimization. Although energy consumption patterns vary across regions and climates, the proposed energy pattern discovery method can be applied to any region to discover more potential patterns.
- We provide a framework for spatio-temporal feature extraction based on the proposed hypotheses and utilize the transformer model to implement spatio-temporal information fusion. Our framework considers the spatial and temporal coherence of energy consumption data, capturing the correlation between the energy consumption patterns of different users. This approach achieves highly accurate demand forecasting at the user level, surpassing traditional models. Specifically, compared to traditional models such as LSTM, the MAE of the proposed method is reduced by 28%. Compared to SOTA models such as TGCN, the MAE of the proposed method is reduced by 14%.
- We developed a visual analytical mechanism to characterize the migration of energy demand over time in both physical and topological space. This graphical solution provides insights into demand-side energy management by observing energy demand migration. Based on this interface, we explored the spatial and temporal distribution of energy consumption in Shanghai from 2016 to 2018 and identified mental inertia in the energy behavior of urban residents. We further explore the causes of mental inertia

and offer solutions. In addition, we estimate the energy transfer potential of Shanghai to be approximately 10.21% based on three years of historical energy consumption data.

By addressing these contributions, our research advances the field of energy demand forecasting by providing more accurate, interpretable, and actionable insights into urban energy systems. This study also lays the groundwork for future research in demand-side energy management and provides valuable insights for policymakers, city planners, and energy stakeholders seeking to navigate the complex challenges of urban energy sustainability. Future research can build upon our methodology to further enhance energy forecasting accuracy and develop more effective demand-side energy management strategies.

Funding

The research is partially supported by the EU H2020 Research and Innovation Program under the Marie Sklodowska-Curie Grant Agreement (Project-DEEP, Grant No. 101109045), the National Natural Science Foundation of China (No. NSFC 61925105 and 62171257), the Tsinghua University-China Mobile Communications Group Co., Ltd. Joint Institute, and the Fundamental Research Funds for the Central Universities, China (No. FRF-NP-20-03). Additionally, the research is partially funded by the German Federal Ministry of Education and Research (BMBF) (project AITT, AI-assisted Technology Transfer, No. 03LB3058B).

CRediT authorship contribution statement

Jieyang Peng: Writing – original draft, Software. **Andreas Kimmig:** Visualization, Methodology. **Dongkun Wang:** Software, Investigation. **Zhibin Niu:** Methodology, Data curation, Conceptualization. **Xiufeng Liu:** Methodology, Investigation, Conceptualization. **Xiaoming Tao:** Supervision, Project administration. **Jivka Ovtcharova:** Writing – review & editing, Supervision, Conceptualization.

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.

References

[1] Veloza OP, Santamaria F. Analysis of major blackouts from 2003 to 2015: Classification of incidents and review of main causes. Electr J 2016;29(7):42–9.
[2] Allen EH, Stuart RB, Wiedman TE. No light in August: Power system restoration following the 2003 North American blackout. IEEE Power Energy Mag 2013;12(1):24–33.
[3] Van der Vleuten E, Lagendijk V. Transnational infrastructure vulnerability: The historical shaping of the 2006 European "blackout". Energy Policy 2010;38(4):2042–52.
[4] Peng J, Kimmig A, Niu Z, Wang J, Liu X, Ovtcharova J. A flexible potential-flow model based high resolution spatiotemporal energy demand forecasting framework. Appl Energy 2021;299:117321.
[5] Kazemzadeh M-R, Amjadian A, Amraee T. A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting. Energy 2020;204:117948.
[6] Raza MA, Khatri KL, Israr A, Haque MIU, Ahmed M, Rafique K, et al. Energy demand and production forecasting in Pakistan. Energy Strategy Rev 2022;39:100788.
[7] Satre-Meloy A, Diakonova M, Grünewald P. Cluster analysis and prediction of residential peak demand profiles using occupant activity data. Appl Energy 2020;260:114246.
[8] Tarmanini C, Sarma N, Gezegin C, Ozgonenel O. Short term load forecasting based on ARIMA and ANN approaches. Energy Rep 2023;9:550–7.
[9] Al-Musaylh MS, Deo RC, Li Y, Adamowski JF. Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. Appl Energy 2018;217:422–39.
[10] Fan G-F, Wei X, Li Y-T, Hong W-C. Forecasting electricity consumption using a novel hybrid model. Sustainable Cities Soc 2020;61:102320.
[11] Antonopoulos I, Robu V, Couraud B, Kirli D, Norbu S, Kiprakis A, et al. Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review. Renew Sustain Energy Rev 2020;130:109899.
[12] Jiang P, Li R, Liu N, Gao Y. A novel composite electricity demand forecasting framework by data processing and optimized support vector machine. Appl Energy 2020;260:114243.
[13] Eseye AT, Lehtonen M. Short-term forecasting of heat demand of buildings for efficient and optimal energy management based on integrated machine learning models. IEEE Trans Ind Inf 2020;16(12):7743–55.
[14] Fan G-F, Zhang L-Z, Yu M, Hong W-C, Dong S-Q. Applications of random forest in multivariable response surface for short-term load forecasting. Int J Electr Power Energy Syst 2022;139:108073.
[15] Aprillia H, Yang H-T, Huang C-M. Statistical load forecasting using optimal quantile regression random forest and risk assessment index. IEEE Trans Smart Grid 2020;12(2):1467–80.
[16] Abbasimehr H, Shabani M, Yousefi M. An optimized model using LSTM network for demand forecasting. Comput Ind Eng 2020;143:106435.
[17] Jin N, Yang F, Mo Y, Zeng Y, Zhou X, Yan K, et al. Highly accurate energy consumption forecasting model based on parallel LSTM neural networks. Adv Eng Inform 2022;51:101442.
[18] Jang J, Han J, Leigh S-B. Prediction of heating energy consumption with operation pattern variables for non-residential buildings using LSTM networks. Energy Build 2022;255:111647.
[19] Khan N, Haq IU, Khan SU, Rho S, Lee MY, Baik SW. DB-Net: A novel dilated CNN based multi-step forecasting model for power consumption in integrated local energy systems. Int J Electr Power Energy Syst 2021;133:107023.
[20] Atik I. A new CNN-based method for short-term forecasting of electrical energy consumption in the COVID-19 period: The case of Turkey. IEEE Access 2022;10:22586–98.
[21] Guo Z, O'Hanley JR, Gibson S. Predicting residential electricity consumption patterns based on smart meter and household data: A case study from the Republic of Ireland. Util Policy 2022;79:101446.
[22] Dileep GJRE. A survey on smart grid technologies and applications. Renew Energy 2020;146:2589–625.
[23] Chen Z, Freihaut J, Lin B, Wang CD. Inverse energy model development via high-dimensional data analysis and sub-metering priority in building data monitoring. Energy Build 2018;172:116–24.
[24] Bayman EO, Dexter F. Multicollinearity in logistic regression models. Anesth Analg 2021;133(2):362–5.
[25] Ahmad T, Zhang H, Yan B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustainable Cities Soc 2020;55:102052.
[26] Nti IK, Teimeh M, Nyarko-Boateng O, Adekoya AF. Electricity load forecasting: A systematic review. J Electr Syst Inf Technol 2020;7:1–19.
[27] Ji S, Zhang Z, Ying S, Wang L, Zhao X, Gao Y. Kullback–Leibler divergence metric learning. IEEE Trans Cybern 2020;52(4):2047–58.
[28] Wu J, Sun J, Xie X, Gao T, Pan Y, Yu H. Accelerating web-based graph visualization with pixel-based edge bundling. In: 2023 IEEE international conference on big data. IEEE; 2023, p. 6005–14.
[29] Li S, Jin X, Xuan Y, Zhou X, Chen W, Wang Y-X, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. Adv Neural Inf Process Syst 2019;32.
[30] Shumway RH, Stoffer DS, Shumway RH, Stoffer DS. ARIMA models. In: Time series analysis and its applications: With R examples. Springer; 2017, p. 75–163.
[31] Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput 2004;14:199–222.
[32] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. 2016, arXiv preprint arXiv:1609.02907.
[33] Cho K, Van Merriënboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: Encoder–decoder approaches. 2014, arXiv preprint arXiv:1409.1259.
[34] Peng J, Kimmig A, Wang J, Liu X, Niu Z, Ovtcharova J. Dual-stage attention-based long-short-term memory neural networks for energy demand prediction. Energy Build 2021;249:111211.
[35] Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, et al. T-GCN: A temporal graph convolutional network for traffic prediction. IEEE Trans Intell Transp Syst 2019;21(9):3848–58.
[36] Chandriah KK, Naraganahalli RV. RNN/LSTM with modified adam optimizer in deep learning approach for automobile spare parts demand forecasting. Multimedia Tools Appl 2021;80(17):26145–59.
[37] Mariano-Hernández D, Hernández-Callejo L, Zorita-Lamadrid A, Duque-Pérez O, García FS. A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis. J Build Eng 2021;33:101692.