Energy 263 (2023) 125939

Contents lists available at ScienceDirect

Energy

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

Understanding multi-scale spatiotemporal energy consumption data: A visual analysis approach

Junqi Wu [a], Zhibin Niu [a,*], Xiang Li [d], Lizhen Huang [c], Per Sieverts Nielsen [b], Xiufeng Liu [b,*]

[a] College of Intelligence and Computing, Tianjin University, China
[b] Department of Technology, Management and Economics, Technical University of Denmark, Denmark
[c] Department of Manufacturing and Civil Engineering, Norwegian University of Science and Technology, Norway
[d] University of California, Berkeley, United States of America

ARTICLE INFO

Keywords:
Demand-side management
Visual analysis
Spatiotemporal patterns
Consumption variability
Segmentation
Demand shift patterns

ABSTRACT

Understanding energy consumption patterns is crucial for energy demand-side management. Unlike traditional data mining or machine learning-based methods, this paper presents visual analysis methods for exploring energy consumption data from spatial, temporal, and spatiotemporal dimensions, including variability, segmentation, and energy demand shifts. To support the proposed methods, we develop a visual analysis tool that allows users to explore consumption data and validate their hypotheses based on visual results through human–client–server interactions. In particular, we propose a novel potential flow-based method to model energy demand shift patterns and have integrated it into the proposed analysis tool. We comprehensively evaluate the proposed methods and the tool using real-world electricity consumption data from the Shanghai Pudong district, and compare with traditional data mining methods. The results demonstrated the effectiveness and superiority of the proposed visual analysis methods, including its ability to discover the spatiotemporal variability of energy demand, customer groups, and demand shift patterns across different geographical areas and time horizons. All results can be well explained by knowledge of the energy consumption in the study region.

1. Introduction

We are currently witnessing rapid global warming, due in part to massive carbon emissions from energy use that contribute to the greenhouse effect. Governments and organizations around the world are developing carbon neutrality strategies. For example, the European Union aims to reduce its emissions by 55% by 2030 and achieve carbon neutrality by 2050 [1]. China aims to reduce its carbon emissions by 60%–65% from 2005 levels in 2030, and to reach peak emissions in 2030 [2]. This goal can be achieved by optimizing energy demand, improving energy efficiency and reducing carbon emissions, as well as increasing the use of renewable energy sources in energy production. In the global, the building sector accounts for a significant amount of energy use. For example, in the EU, residential and commercial buildings use about 40% of total energy and emit 30% of total carbon dioxide [3]. Thus, there is great potential for improvements in energy efficiency and carbon reduction in the building sector. As energy consumption is directly related to the living habits of customers, analysis of energy consumption data can help save energy by changing consumer behavior, which is relatively straightforward and cost-effective [4].

In the energy sector, with the widespread use of smart meter technology in recent years, utilities have accumulated a significant amount of granular data that can be used to improve energy management, including providing personalized services through customer segmentation, designing demand response programs, and planning energy supply [5,6]. In general, there are three types of energy data analysis methods. The first are statistical methods, which are often used by economists, i.e., calculating energy consumption using empirical numerical models. Although these methods are effective in understanding energy consumption data, such as trends or patterns, it is often difficult to discover the reasons behind them. As a result, the socioeconomic modeling approach lacks flexibility, intuitiveness, and interpretability. The second are data-driven methods, based on data mining or machine learning. These methods are effective and are becoming the most widely used today, but these methods are often difficult to understand and complex, requiring extensive domain knowledge and even some programming skills. The third are visual analysis methods, also used in this paper. Visual analysis has been used in bioinformatics and finance. For example, visual analysis is used to analyze genome structure [7,8]

* Corresponding authors.
E-mail addresses: zniu@tju.edu.cn (Z. Niu), xiuli@dtu.dk (X. Liu).

https://doi.org/10.1016/j.energy.2022.125939
Received 11 February 2022; Received in revised form 19 September 2022; Accepted 29 October 2022
Available online 4 November 2022

and investigate lesions [9], and in finance, it is used to detect financial risks [10], financial fraud [11], etc. Due to its intuitive nature and its ability to incorporate human cognitive ability into data analysis, visual analysis can be used to solve problems that are difficult or time-consuming to analyze. Currently, visual analysis remains in its infancy in the energy sector, but attracts increasing interest from researchers. Now it starts being used to solve certain problems in energy management, such as visualizing peaks and valleys to identify peak shaving potential and identifying customers with similar habits to provide personalized services [4]. The first academic conference on visualization and visual analysis of energy data (EnergyVis, https://energyvis.org) was held in 2020, and some interesting work was presented. Among others, Chetty et al. [12] performed an in-depth visual analysis to study the impact of the COVID-19 pandemic on the power grid and the likely future impact. Alshardy et al. [13] attempt to merge various visualization techniques to facilitate informed decision-making on electrification of remote areas in Indonesia. Mammen et al. [14] use the open-source visualization platform Grafana to study the relationship between indoor temperature and floor height in a building to provide suggestions for energy efficiency. We believe that visual analysis will receive more research attention in the future and will become a key component of smart energy systems.

In energy demand management, exploring crowd behavior and demand dynamics plays a critical role in achieving accurate energy supply and scheduling; however, utilities often lack sufficient understanding of the holistic and dynamic characteristics of energy demand. In this paper, we propose a visualization and analysis method using real-world electricity consumption data to analyze the crowd behavior of customers in both temporal and spatial dimensions. Specifically, we explore three aspects of the spatiotemporal characteristics of energy demand, including variability, segmentation, and demand shift dynamics. First, we analyze the spatiotemporal variability of energy demand, i.e., the demand variability of all customers in different geographic locations at a given time, as well as the demand variability over time for customers in the same geographic area. Demand variability can contribute to energy provisioning and supply coordination. Second, we segment customers based on their energy consumption patterns. Utilities can interactively identify customers with similar consumption patterns and provide accurate personalized services. Finally, we introduce the fluid dynamics concept, *potential flows* [15], to model the spatiotemporal dynamics of energy demand, through which users can discover the spatial, temporal, and spatiotemporal dynamics. To the best of our knowledge, this is the first attempt to model energy demand dynamics using the fluid dynamics concept. To facilitate the use of the proposed visualization and analysis methods, we developed a user-friendly web application that allows users to analyze consumption data interactively. The analysis process starts with formulating hypothesis about energy demand, then exploring the analytical results, and finally testing the hypothesis. This approach makes data analysis more flexible, intuitive, and convenient than data-driven methods. It only requires simple clicks or drags on the visual dashboard to complete the analysis process. In addition, we analyze global and local data structures and reveal their semantics to ensure that the results are sufficiently interpretable. Although the visualization and visual analysis method presented in this paper is for the analysis of three specific respects of electricity consumption data, the method is designed to be generic and is also applicable to other analytical tasks and energy types.

In summary, this paper makes the following contributions.

- We propose the visual analysis approach for energy demand-side management, allowing users to formulate and validate hypotheses, and obtain insights from data through a human–client–server interactive analysis loop.
- We propose visual analysis methods to model energy demand variability, segmentation, and shift patterns. In particular, we propose a novel potential flow-based method to model energy demand-shift patterns.
- We develop a visual analysis tool to support the proposed methods, enabling a user-friendly and efficient exploratory analysis of energy demand data from spatial, temporal and spatiotemporal dimensions.
- We evaluate the proposed visual analysis tool using real-world electricity consumption data, compare it with traditional data mining methods, and validate its effectiveness and superiority.

The rest of the paper is organized as follows: Section 2 describes the related work; Section 3 presents the data for the study and describes the research problem; Section 4 presents the visual analysis methods and the tool; Section 5 presents the results and analysis of case studies; Section 6 concludes the paper and presents the future research directions.

2. Related work

In this section, we will first review smart meter data analysis in general, followed by segmentation analysis; next, we will investigate visualization and visual analysis in the energy sector; and finally, we will examine spatiotemporal data analysis.

Smart meter data analysis. In recent years, smart meter data analytics has become increasingly popular in the energy sector due to widespread smart meter installations and scalable data collection [4]. Smart meter data can be used for a variety of purposes, among others including load prediction [16–18], load profile clustering [4,19] and household characteristics inferences [20,21]. Accurate forecasts not only help utilities plan resources and take control measures to balance energy supply and demand, but also help customers understand their energy consumption and future needs to better manage their usage costs [22]. Load profile clustering can be used to improve the accuracy of predictions at the individual or aggregate level [23,24]. Customer information can be derived from load profiles [21], which is the important information for utilities to provide personalized service or assist utilities in determining efficient rate structures and demand response [25]. Smart meter data disaggregation is another analysis approach, which helps customers understand the load profile of each appliance, e.g., [26,27]. This helps organize and optimize appliance usage schedules [28], as well as identify and eliminate low-efficient appliances [29]. In terms of the technologies for smart meter data analysis, they can be divided into three categories: memory-based, database-based, centralized, and distributed. A comparative evaluation of analysis technologies [30,31] shows that memory-based methods generally perform better on computing performance. In this paper, we use a visual analysis approach to smart meter data, and focus on studying the impact of crowd behavior on energy consumption, including demand variability, patterns, and demand shift across spatial and temporal dimensions, which provides an alternative perspective to other work.

Segmentation analysis and clustering. Customer segmentation analysis is one of the key components of energy demand-side management. A common approach to segmentation is to use cluster analysis, which divides load profiles into different groups or clusters based on similarity. A major problem when applying clustering is how to handle the curse of dimensionality, i.e., as the number of dimensions increases, cluster analysis becomes more difficult because the volume of space increases rapidly, requiring a significant amount of data to achieve the desired performance [32]. There are generally two approaches to dealing with high-dimensional data. The first is to reduce the dimensional space by extracting feature sets from high-dimensional data and then conduct the clustering on the resulting features. For example, Choksi et al. [33] combine SOM and feature-based dimensionality reduction method to extract features and then perform k-means clustering based on the extracted features. Haben et al. [34] perform k-means clustering based on seven features, including overnight, breakfast, daytime/evening, seasonality, weekday/weekend, and day. These studies revealed a significant improvement in feature-based clustering over clustering of raw hourly data alone. Another approach is to apply dimensionality reduction algorithms, such as principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE) and unified manifold approximation and projection (UMAP), directly to high-dimensional data to reduce dimensionality, then perform the clustering, e.g., [4,35,36]. This paper employs the latter approach, but is different from other works. The proposed customer segmentation is a two-step analysis method. The first step is to reduce high-dimensional time series to visible two-dimensional (2D) points, and then the obtained data points are placed on a 2D plane based on their similarity, i.e., if the more similar two points are, the closer they will be placed; the second step is to identify the customer groups with similar patterns, which is achieved through an interactive process between human and machine. This approach is more intuitive and allows one to identify more interesting customer groups.

Visualization and visual analysis in the energy sector. Visualization and visual analysis can lead people to a better and more intuitive understanding of their data [37]. With the digitization of energy systems, data visualization and visual analysis are receiving ever-increasing attention in the energy sector. To better interpret energy consumption data, improve energy efficiency, and implement demand response programs, Borgeson et al. [38] proposed the Energy Visualization and Insight System (VISDOM), an integrated energy management system that combines various smart meter data analysis algorithms and visualization methods. Singh et al. [39] combine visual analysis and data mining methods to analyze and predict energy consumption time series, revealing various temporal patterns. Guo et al. [40] developed a visual analysis algorithm that can visually assess the efficiency of energy consumption in thermal comfort. Francisco et al. [41] propose a novel information representation that encodes different colors in a Building Information Model to reveal different spatial energy consumption level. A fully functional prototype for the thermal simulation of large-scale district energy systems was proposed based on the district heating network data. The results of dynamic simulation can be visualized in the form of animations or maps in QGIS [42]. Ellegarrd et al. [43] develop an event sequence visualization method that provides direct feedback and information on their residential energy use. Fan et al. transform the operational data into an undirected status graph for pattern analysis [44]. Goodwin et al. conclude that creative user-centered visualization may help energy stakeholders gain insight into data when datasets are largely unknown [45]. Liu et al. [46] develop a smart meter data dashboard that allows users to visualize disaggregate energy consumption, including baseload, activity load and load due to weather temperature. However, most of the existing work focuses on the visualization of using different statistical charts, while placing less emphasis on visual analysis. This paper fills this gap by introducing visual analysis that allows users to analyze data through interactions while incorporating their cognitive ability.

Spatiotemporal data visualization. Spatiotemporal data visualization has attracted great interest in various fields. Most of the current spatiotemporal research is related to trajectory data analysis, e.g., [47–49], where trajectory patterns are extracted to help improve mobility. Attempts at spatiotemporal visualization of energy data are rather limited. The following are some efforts we found. Luo et al. [50] analyze the energy consumption and emissions of taxis and their spatiotemporal distribution in Shanghai by applying big data analysis on GPS data from taxis. Liu et al. [51] model multidimensional spatiotemporal data as tensors and propose a new tensor decomposition to extract latent patterns; and implement a visual analysis framework for spatiotemporal data exploration. Liu et al. [52] employ a flow map to track crowd behavior, which can accurately identify the location of events. Kim et al. [47] implement a gravity-based flow extraction model that can effectively separate human movement from spatiotemporal data. Ahmed et al. [53] proposed ContourDiff, a vector-based visualization tool capable of visualizing trends in spatial and temporal domains in contour maps. Inspired by the different visualization techniques, in this paper we apply the potential flow map to visualize the spatiotemporal shift pattern of the energy demand.

3. Data and problem statement

Datasets. The study uses the electricity consumption dataset over the period from July 2015 to June 2018, which was collected in Pudong District by Shanghai National Grid Company. The original dataset consists of more than one million customers, which are overloaded for the entire analysis process. Without losing the representativeness and generality, we performed uniform sampling of the whole dataset, which obtains 10,000 customers for this study. Table 1 shows some time series readings of a customer's electricity consumption and its geographical location (Latitude and Longitude). All time series have a daily resolution, with total energy demand, *pap_r* (positive active power), peak period demand (6:00–22:00), *pap_r*1 (positive active power at peak hour), off-peak period demand (22:00–6:00), *pap_r*2, (positive active power at trough hour). The total demand should be equal to the sum of peak and off-peak demands, i.e., *pap_r* = *pap_r*1+*pap_r*2, but it may be slightly different due to some instrument measurement, floating-point rounding, or data processing errors.

Table 1
The electrical household consumption data.

CustomerID	pap_r	pap_r1	pap_r2	Date	Latitude	Longitude
224653173	1.67	0.82	0.85	01/01/2017	121.62492	31.18847324
224653173	2.80	1.30	1.50	02/01/2017	121.62492	31.18847324
224653173	2.24	1.20	1.04	03/01/2017	121.62492	31.18847324

Problem statement. For demand-side management, utilities are often interested in quantifying energy consumption, understanding the spatial distribution and variation of consumption over time, and discovering consumer groups to plan energy distribution, customize services, and design demand response programs. However, most existing work only studies consumption statistics or forecasts over time in a local area [54], which does not possess comprehensive and holistic information for the above decision makings.

To address these issues, we will conduct a spatial and temporal analysis of energy consumption data, and our study focuses on the following three aspects: spatial–temporal variability analysis, customer segmentation analysis, and spatial–temporal shift pattern analysis of energy demand. The main reason is that the proposed studies are the most relevant and essential information for demand-side management. The variability analysis will assess the variation of energy demand at a specific geographic scale, such as a city, over time. Segmentation analysis will use a visual analysis approach that allows users to interact and discover different customer groups based on their domain and cognitive ability. The shift pattern analysis will illustrate the distribution of energy demand and the shifts of energy demand in different geospatial regions at a specific point in time and overtime. This shift pattern analysis has not yet been sufficiently explored, to the best of our knowledge.

We will use a data-driven approach combined with visual analysis and visualization to explore spatiotemporal patterns of demand variability, segmentation, and shifts in the following study, using Pudong electricity data as an example. Although the dataset used is location- and type-specific, the proposed method can be easily applied to other locations and other types of energy demand, such as water, gas, and heating. The proposed method can be integrated into a smart energy system to support demand-side management.

4. Methodology

This section will describe overview, data processing, visual analysis methods, and implementation of the tool.

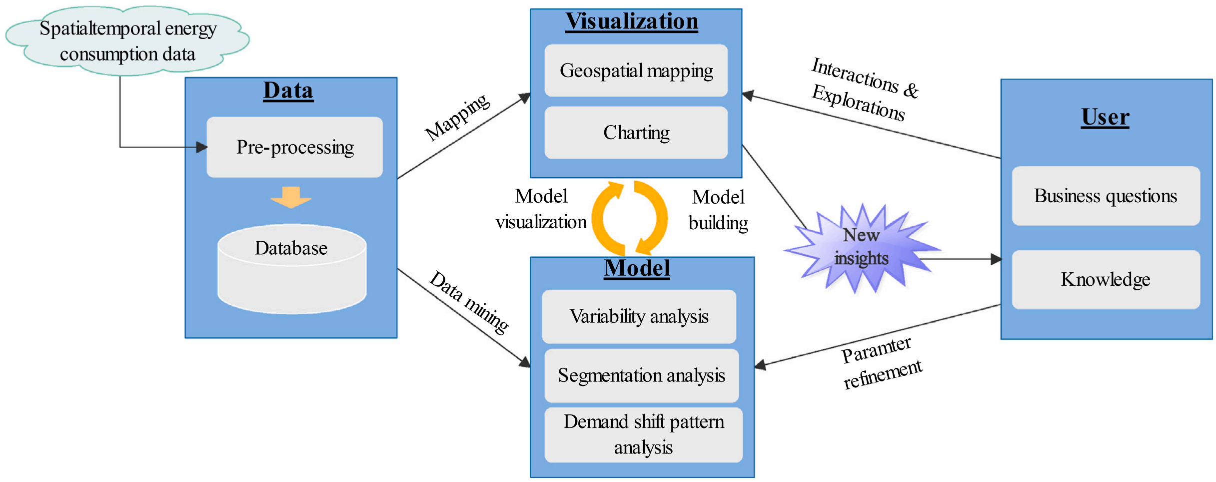

Fig. 1. Overview of the visual analysis based method.

4.1. Overview

This paper employs a visual analysis approach to explore energy consumption time series from multiple spatiotemporal scales. Unlike common data-driven approaches, visual analysis approaches provide visual charts that explain the data and allow users to interact with the charts for better insights. Fig. 1 shows the proposed visual analysis framework, which consists of the following components: data, model, visualization, and user. The data are time series of energy consumption with spatiotemporal information, cleaned and stored in the underlying relational database management system. We develop three analytic models to help users gain knowledge from the data, including the variability model, interactive segmentation model, and demand shift pattern model. With this framework, users can easily add new models by implementing standardized application programming interfaces (API). The visualization component is a web application with a user-friendly interface that visually presents data and communicates with users through interactions. The user interface supports geospatial mapping and data visualization in spatial and temporal dimensions. The user component is the subject who acquires knowledge from the data through interactions. The knowledge discovery process typically begins by formulating a business question, then exploring analytical views to answer the question, and refining model parameters for deeper insights. This is an iterative process that involves users in a loop.

4.2. Data processing

The data preprocessing mainly consists of two tasks: (1) ensuring data quality and (2) transforming data into suitable formats for visual analysis. The two tasks are crucial for the subsequent analysis.

For the first task, we improve data quality, including fixing missing values and removing outliers. In the original dataset, missing values accounted for about 1% of the total data, due to device failure or transmission errors. First, we fill the missing values using a linear interpolation approach because the energy consumption of a household is unlikely to change dramatically within consequential days. Then, we smooth the outliers in the energy consumption time series. Outliers are reflected in an abrupt abnormal increase in values, which can be identified by the Tukey method [55]. Values that lie far away from the central value are classified as outliers. As there are different smoothing methods available, it should be based on the time series characteristics to select an appropriate method. In this paper, we choose the window-based convolution approach [56] to smooth the consumption time series. Given a time series $X = \langle x_1, \dots, x_n \rangle$, we use the following smoothing function to generate a new value:

$$x_j^* = \frac{\sum_{i=-m}^{i=m} c_i x_{j+i}}{N} \tag{1}$$

where x is the original value, x^* is the resulting value, c_i is the coefficient for the ith value within a smoothing window, and N is the number of convolution integers, equal to the size of the smoothing window $(2m + 1)$. The index j is the running index of the original ordinate data array. The smoothing array consists of $2m + 1$ points, where m is the half-width of the smoothing window size. The Savitzky–Golay filter coefficients (c_i) can be obtained directly from [57]. In this study, the half-width m of the smoothing window is specified as 4.

The second preprocessing task is mainly concerned with data normalization and dimension reduction. We applied the Z-score [58] to normalize the consumption time series. The Z-score describes how many standard deviations a given measure is above or below a population mean, which is defined as the following.

$$z = \frac{(x - \mu)}{\sigma} \tag{2}$$

where x is the observed measurement, μ is the expected measurement (population mean), and σ is the population standard deviation. Observations above the mean have positive standard scores and observations below the mean have negative standard scores. Once the standard score has been calculated, we plot the normalized energy consumption in a uniform normal distribution to minimize the impact of abnormal values on the trend of time series. Another transformation task is dimension reduction for visualizing high-dimensional time series in a visual 2D view. Dimension reduction, also known as the embedding technique, projects high-dimensional data into low-dimensional data while maintaining the global and local relational structure of the data. The most commonly used embedding techniques are PCA, t-SNE and UMAP. PCA is a linear dimension reduction method generally used for dimension reduction with dimensions less than 10, while both t-SNE and UMAP are nonlinear dimension reduction methods. In this paper, based on our previous study [4], we chose t-SNE because of its better performance.

All well-prepared data were stored in a database management system, which are used for exploratory analysis in our visual analysis system.

4.3. Spatiotemporal variability analysis of energy demand

For utilities, it is crucial to analyze the variability of the energy demand in spatial and temporal dimensions. Variability can effectively reflect not only the characteristics of customers geographically located at different places but also the changes in customer behavior and other factors that may affect mass behaviors, such as climate.

To calculate the spatiotemporal variability (stability) of energy demand, we calculate the average difference of the demands of all households within a given area over time. Here we quantify the variability by considering all the abundances of individual energy demand relative to each other, then calculate the average variability of the popularization. The process can be formalized as follows.

First, the number of all possible combinations of comparison at the time t within a given geographic spatial area is calculated as follows:

$$N_t = \frac{n_t(n_t - 1)}{2} \tag{3}$$

where n_t represents the number of households at t within a given area. Then, we define c_t as the list of possible pairwise comparisons, $c_t = 1, \dots, N_t$. Therefore, c_t represents a pair of energy demands of two households $h_{t,i}$ and $h_{t,j}$ at time t. The proportional difference of each pair c_t can be defined as:

$$H(c_t) = \begin{cases} 0 & \text{if } h_{t,i} = h_{t,j} \\ \frac{|h_{t,i} - h_{t,j}|}{\max(h_{t,i}, h_{t,j})} & \text{otherwise} \end{cases} \tag{4}$$

where $|\cdot|$ represents the absolute difference between the demand for two households and $\max(\cdot)$ represents the maximum demand for them. The above equation can algebraically be transformed into the following:

$$H(c_t) = 1 - \frac{\min(h_{t,i}, h_{t,j})}{\max(h_{t,i}, h_{t,j})} \tag{5}$$

which indicates that the spatiotemporal difference is derived from a ratio comparison of each household within a region over time. The difference in every time step will be compared with every other in a geographic area, resulting in a distribution of proportional differences $H(c_t)$. As a result, the population variability can be derived from the frequency distributions of the $H(c_t)$ scores, and an average will provide the sufficient summary for it, that is,

$$V(t) = \frac{\sum^{c_t} H(c_t)}{N_t} \tag{6}$$

Therefore, we can calculate the spatiotemporal variability of energy demand based on a simple but thorough comparison of all households over time. In Eq. (6), $V \in [0, 1]$, the zero score represents that all households have the same demand, complete stability, while the value of 1 is reached when the difference in population size is infinite.

4.4. Segmentation of energy consumption patterns

Segmentation is used to group customers with similar consumption patterns, which can be used to provide personalized services. With the visual analysis approach, consumption patterns will be mapped into a lower-dimensional space for visualization.

Let $x_1, \dots, x_N$ be the time series data for urban energy consumption and N be the number of households. The probabilities p_{ij} that are proportional to the similarity of two objects x_i and x_j can be calculated as follows:

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2N} \tag{7}$$

in which $p_{i|j}$ is defined as:

$$p_{i|j} = \frac{exp(-\|x_i - x_j\|^2)}{\sum_{k \neq l} exp(-\|x_k - x_l\|^2)} \tag{8}$$

where $p_{ij} = p_{ji}$, $p_{ii} = 0$, $\sum_{ij} p_{ij} = 1$.

To map the similarity of patterns from high- to low-dimensional space, we arrange low-dimensional points $y_1, \dots, y_N$ according to their similarity computed by the t-distribution method, so that the more similar the two points are, the closer they are placed on a map. The similarity in the low-dimensional space is defined as follows:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{9}$$

The mapping should minimize this difference in point distributions between high- and low-dimensional spaces. Therefore, the location of a point y_i in the low-dimensional map can be determined by minimizing the Kullback–Leibler (KL) divergence between the two distributions P and Q:

$$KL(P\,||\,Q) = \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}} \tag{10}$$

which can be obtained by applying the gradient descent method, defined as follows:

$$\frac{\delta KL}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1} \tag{11}$$

4.5. Spatiotemporal demand-shift pattern analysis

Understanding the spatiotemporal patterns of energy demand shifts is crucial for utilities to improve their operations, energy supply, and distribution. A visual analytic model helps uncover demand-shifting patterns in different geographic locations over time.

In this paper, we introduce the potential flow [15] to model the spatiotemporal dynamics of energy demand. Potential flow has been used to model fluid dynamics in fluid engineering, for example, water waves, electro-osmotic flow, and groundwater flow. It models the velocity field as the gradient of a scalar function: the velocity potential. We observe that geospatial energy demands continuously change over time. Thus, the demand can be viewed as a continuum occupying a region simply connected to at a point in time but with irrotational characteristics. Inspired by fluid dynamics and continuum mechanics, we therefore model the continuum of energy demand shifts as potential flows [15], and visualize them on a map. The visual analysis tool created in Section 4.6 will help to understand the dynamics of energy demand through the proposed potential flows. This can significantly help utilities make decisions regarding energy distribution and supply.

According to fluid dynamics, the angular velocity of a flow can be defined as

$$\omega = \frac{1}{2} \cdot \nabla \times \vec{V} \tag{12}$$

If the angular velocity is zero, then

$$\nabla \times \vec{V} = 0 \tag{13}$$

For the irrotational flow, the potential function φ is defined as $\vec{V} = -\nabla\varphi$, where the minus sign represents that φ decreases in the flow direction. Therefore,

$$\nabla \times \vec{V} = -\nabla \times \nabla\varphi = 0 \tag{14}$$

which means that for all φ, $-\nabla \times \nabla\varphi = 0$. Therefore, the velocity potential exists only for the irrotational flow, with the same as the potential flow. The potential function φ is a continuous scalar function $\varphi(x, y, z, t)$ satisfying the irrotationality condition, where x, y, z represents the dimensions in a 3-D spatial space at the time at t.

In this paper, we model the dynamics of energy demand shift of households distributed on a 2D map, each of which has the GPS coordinate (x, y). As households are spatially discrete and distributed on an irregular grid, we employ kernel density estimation (KDE) to encode their energy consumption into a continuous representation, defined as follows:

$$\hat{f}_t(x)\Big|_t = \frac{1}{n} \sum_{i=1}^{n} c_i K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^{n} c_i K\left(\frac{x - x_i}{h}\right) \tag{15}$$

where h is the bandwidth; $x_i = (lon_{i1}, lat_{i2})^T$, is the coordinate of a household i; K is the kernel function, which is a symmetric multivariate density; and c_i is a normalized value of average energy consumption used to reweight demand strength with respect to geographic distribution, which is defined as follows:

$$c_i = \lfloor \gamma E \rfloor \tag{16}$$

where E represents the energy consumption of x_i and γ is the filter coefficient defined by the users. We select the Gaussian kernel to estimate the demand strength because it can provide a reasonable estimate even for a small dataset, which is defined as follows:

$$K_h\left(x - x_i\right) = e^{-\frac{\|x-x_i\|^2}{2h^2}} \tag{17}$$

With the kernel density matrix (strength map), the temporal dynamics of the energy demand over time from t_1 to t_2 can be obtained by Eq. (18), which calculates the gradient of the strength difference in demand.

$$Shift|_{t_1,t_2} = \nabla(\hat{f}_{t_2} - \hat{f}_{t_1}) \tag{18}$$

The vector flow fields (arrows) represent shifts in energy demand, where the arrow represents the direction of the shift while the length represents the strength of the demand; the longer the arrow, the greater the demand shift.

4.6. Visual analysis tool for pattern discovery

In this section, we present the requirements of the visual analysis tool,[1] its implementation and the description of the interactive visual analysis to discover patterns of energy consumption and patterns of consumption shifts in spatiotemporal dimensions.

The visual analysis system will address the following three requirements:

- *Visual based consumption pattern discovery by user interactions.* Clustering is one of the most widely used methods for pattern discovery. For example, the k-means clustering algorithm segments energy consumption load shapes based on similarity. However, the challenge is that users must specify the number of clusters, k, prior to clustering, which is often not feasible or optimal, for example, when integrating online data. Instead of manually specifying the number of clusters, the proposed visual method leaves the cluster segmentation tasks to users, who can discover different patterns interactively by selecting data points in a two-dimensional scatterplot generated by the dimensionality reduction method presented in Section 4.4.
- *Holistic analysis of the spatiotemporal energy-demand dynamics.* Spatiotemporal analysis of energy consumption data is challenging because the entanglement and dynamics of spatial and temporal distributions make it difficult to verify assumptions made by users in visual analysis. Based on previous literature on spatiotemporal analysis, a holistic approach that combines both temporal and spatial information will help discover new insights.
- *Mine mass consumption behaviors through visual analysis.* The study of the behavior of mass energy consumption in different geospatial locations and over time plays a vital role in better energy provision and planning. Traditional methods use data mining or mathematical modeling, but they are often difficult to use for nonexpert users. Therefore, we introduce a visual analysis method, which allows users to better mine the mass consumption behaviors.

Based on the above requirements, we designed and implemented a visual analysis system to support the exploratory analysis of energy consumption. Fig. 2 provides an overview. This system employs a layer architecture consisting of: (1) a backend server which stores the data and performs demand-shift modeling based on the user setting; (2) a web-based front-end visual analysis user interface where users can perform exploratory analysis including consumption pattern discovery and demand-shift pattern discovery; and (3) user interaction and knowledge acquisition, by which users can initiate asking questions, explore the response charts, and acquire knowledge in a cognitive process.

The front-end visual analysis user interface is the key component of this visual analysis system. Panel 1 is the user interface for exploring energy consumption and segmentation through user interactions. It consists of three views, View $A1$, $B1$ and $C1$. View $A1$ is the entry point to explore consumption patterns. The points in the point cloud were generated from consumption time series using the dimensionality reduction method presented in Section 4.4. The points are organized based on the similarity of load shapes; the more similar, the closer. Therefore, users can interactively select closely placed points, and discover the load shapes that are shown in View $B1$. The corresponding geographic locations of these customers are shown in View $C1$. Pattern discovery based on visual analysis incorporates human cognitive ability, which can be illustrated by the following example. Suppose that we want to discover customers with air conditioners. We first assume that these customers consume more energy in winter and summer due to heating and cooling; we then select different closely located points in the point cloud and observe the corresponding load shapes until we obtain a satisfactory result based on the judgment with domain knowledge. This can be an iterative process with multiple attempts and observations.

Panel 2 is for the discovery of the energy demand-shift pattern. View $A2$ is the control panel and the entry point to explore the shift in energy demand. The shift of energy demand over two discrete time periods can be represented by the potential flows showing on the spatial dimension, e.g., on a map. With this analysis system, users can interactively select periods of interest through user interactions with the buttons on View $A2$. Currently, the system supports three types of demand-shift analysis, including a peak–valley period, a regular-split period, and multiple periods. In the underlying, the spatial shift of energy demand over time is modeled by the fluid dynamics approach, where potential flows will be calculated by the back-end engine (see Section 4.5). Then, the demand shift will be displayed as potential flows on a map (View $C2$). Take the current result in Panel 2 as the example. The results indicate the demand shift pattern of a regular-split period from April to August 2017, and the shift patterns of the regular-split period are indexed visually in View $D2$. When an indexed view is selected in View $D2$, the details of the selected shift pattern will be shown in View $C2$. View $B2$ is the metering view, which gives a statistical overview of energy demand and demand distributions on a daily and household basis. The design of the metering view is consistent with the charting technologies that are widely used in the energy sector in order to give an overview for total, peak, valley, and average daily demand.

Fig. 3 shows the index view for the period 2017-04-14 to 2017-05-14 (left) and its shift pattern after zooming in (right). In this visual analysis system, we introduce four visual elements to represent the demand shift:

- *Demand shift:* It is represented by a flow map where the length of the arrow codes the strength of the demand change (v);
- *Demand-shift window:* It gives the coverage of the analysis, and its border color encodes the spatial demand change φ. The window-shape design will not obscure the map, but give the necessary quantitative information for the demand shift;
- *Demand-shift color legend:* It gives the corresponding absolute value for the spatial energy demand shift in the grid area. Quantitative results for the demand shift can be calculated over spatial locations and time horizons;

[1] https://youtu.be/swVUcveIWng.

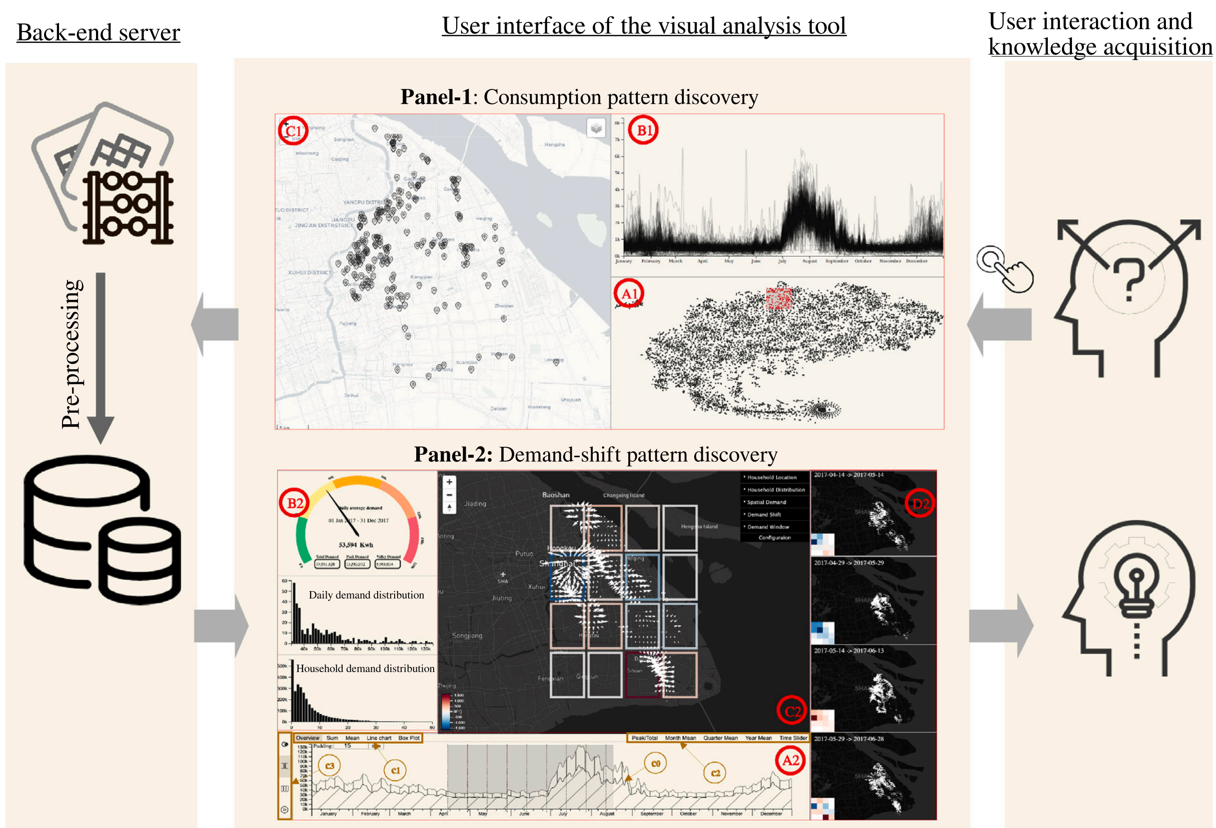

Fig. 2. Overview of the visual analysis tool.

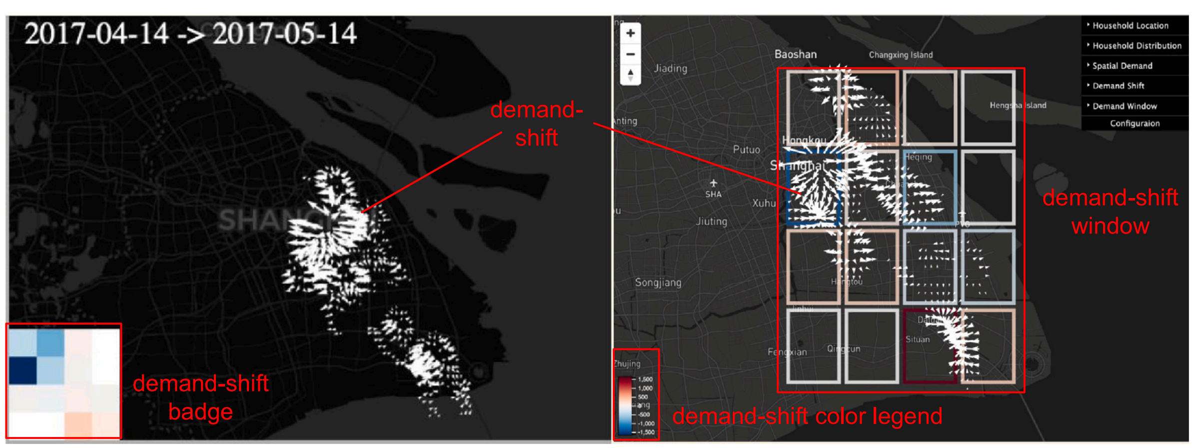

Fig. 3. Visual design for energy demand shift patterns.

- *Demand-shift badge:* It also encodes the spatial energy demand change φ, which has the same meaning as the demand-shift window but gives a visual summary of demand change in the grid area in the visual index view. We use a solid grid, instead of a frame, because it is much smaller on the visual index and more prominent.

Such a visual design assists users in quickly finding the interest areas with more information. View $D2$ gives a thumbnail of the "overview" of energy demand, which is represented by a small demand-shift badge. At the same time, users may want to view details, so we provide to use the Demand-shift window for users to visualize the details. The vector arrows visualize the flow directions, while the color of demand-shift windows visualizes increasing or decreasing energy demand. The design follows the Schniederman Mantra: first the overview, zoom, and filter, then the details on demand [59].

4.7. User interaction and knowledge acquisition process

This section describes how users interact with the system for exploratory analysis and knowledge acquisition. As discussed in the previous subsection, panels 1 and 2 are for exploring customer segmentation based on consumption patterns and energy demand shift patterns over spatial and temporal dimensions. The corresponding starting point for the two types of exploratory analysis is View $A1$ and $A2$, respectively. The following describes two typical user interaction scenarios and the acknowledge acquisition process.

S1: Typical patterns discovery. In this scenario, a user can interact with the system by discovering typical consumption patterns and identifying the spatial distribution of customers in the study area. First, the user can start by asking a question, e.g., *who are the early birds with a morning peak between 5:00 and 7:00?*. The user selects points at different locations in view $A1$ of panel 1, observes their patterns in view $B1$ and the geographic distribution in view $C1$. Then, the user can discover the change in consumption patterns based on the similarity of points (or the spacing of points) in a two-dimensional space. This is done by successively selecting points close to each other and then observing the change in the patterns in the spatial space. Third, users can select scatterplots generated by different dimension reduction methods, including t-SNE and MDS, to observe the differences and compare their ability to discover typical patterns.

S2: Spatiotemporal shift pattern discovery. In this scenario, a user can examine the patterns of energy demand interactively. View $A2$ in Fig. 2 displays the temporal energy demand control ($c0$) and provides the main functional interaction controls ($c2$ and 3) to support spatiotemporal demand-shift analysis. In greater detail, the temporal energy demand control ($c0$) uses a stream graph to visualize energy demand during peak and non-peak (valley) periods. The user can toggle the auxiliary analysis line (yearly, quarterly, monthly average demand and peak-to-valley ratio) in the $c2$ to select the period of interest for further analysis. The demand-shift analysis starts by pressing the functional buttons of $c3$. If the user has decided the period of interest to analyze through the previous step, (s)he can define the exploration task in one of the following temporal types: peak–valley period, regularly split period, or multiple periods. The user can toggle the corresponding button, select the period(s) of interest by brush operation on $c0$, and finally toggle the compute button on $c3$ to generate the results listed as the demand-shift visual index in View $D2$.

5. Results and analysis

In this section, we first perform an exploratory analysis of the data, analyze the viability of the energy demand, and finally explore demand segmentation and shift patterns through user interaction with the implemented visual analytic tool. Since visual analysis methods cannot always quantify performance, we use case studies and expert views to demonstrate the effectiveness of our approach.

5.1. Exploratory analysis for the datasets

As described in Section 3, the data in the study are the electricity consumption of the Pudong district, Shanghai. Daily consumption was divided into peak and off-peak hour consumption. To better understand the data, we illustrate daily consumption using the histogram in Fig. 4. As shown, daily consumption varies from 0 to 100 kWh, while most daily consumption is located at 0–8 kWh, accounting for approximately 75%.

Fig. 5 shows the physical distribution on the map of residential households in the Pudong district by the heat map and the spatial points. As shown in this figure, the densest area is in the western area of the Huangpu River, the traditional center of the city in Pudong district. The other dense area is near Zhangjiang HiTech Park (not shown in this figure for space reasons), a new area, which has been developed since 1992.

Fig. 6 displays the distribution of daily energy consumption over different months between 2016 and 2017 with a boxplot. To evaluate the correlation between energy consumption and weather temperature, we provide the temperature plot of Shanghai in 2016–2017 (see Fig. 7). The typical weather in Shanghai is that summer is hot and winter is cold and wet. According to the results in Fig. 6, July and August were the months with the highest demand for energy consumption, followed by the months from December to February. Therefore, the energy demand visually correlates with the temperature, that is, the high demand in both summer and winter. It may be due to air conditioning for heating and cooling. The summer of 2017 is warmer than 2016 and, as a result, the energy demand for 2017 is visualized as higher than 2016 (see Fig. 6).

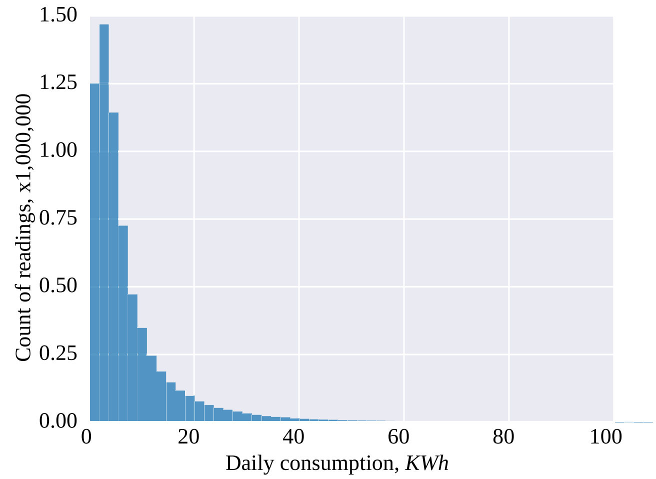

Fig. 4. The histogram of daily readings.

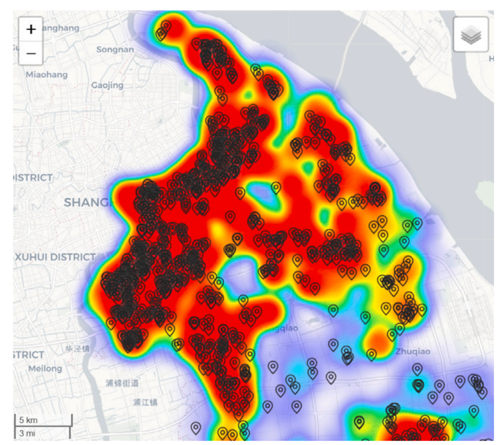

Fig. 5. Distribution of households.

5.2. Spatiotemporal variability analysis

We study the spatiotempral variability of energy demand for the data using the method proposed in Section 4.3. Although the boxplot in Fig. 6 summarizes the spread and skewness of energy consumption through the quartiles and the variability outside the upper and lower quartiles through its whisker lines, it can be interesting to further quantify the variability in spatiotemporal dimensions with the proposed method. We take the consumption of 2017 as sample data (with higher variability according to the boxplot due to a warmer summer), and plot the quantitative values of variability over different months in Fig. 8. The results show that the variability of winter and summer is higher than that of the other two seasons, which means that air conditioners can increase the variability of energy demand. In terms of variability values, July is the highest, while October is the lowest. July is the hottest month of 2017, with the highest temperature reaching 43 °C (see Fig. 7). The weather in autumn and spring in Shanghai is generally very mild. For example, the weather temperature in October 2017 is 15–20 °C, which assumes that no air conditioner is used.

To further explore the difference in variability, we plot histograms of households distributed in different consumption intervals for July

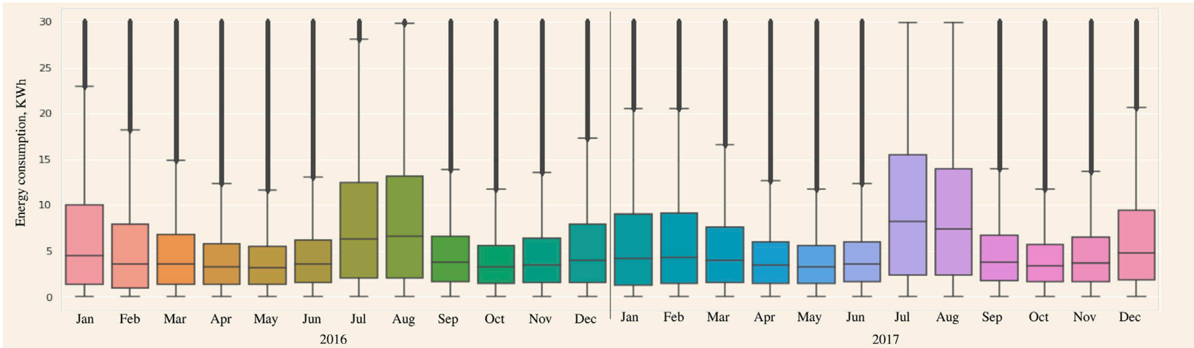

Fig. 6. Electricity consumption in 2016–2017.

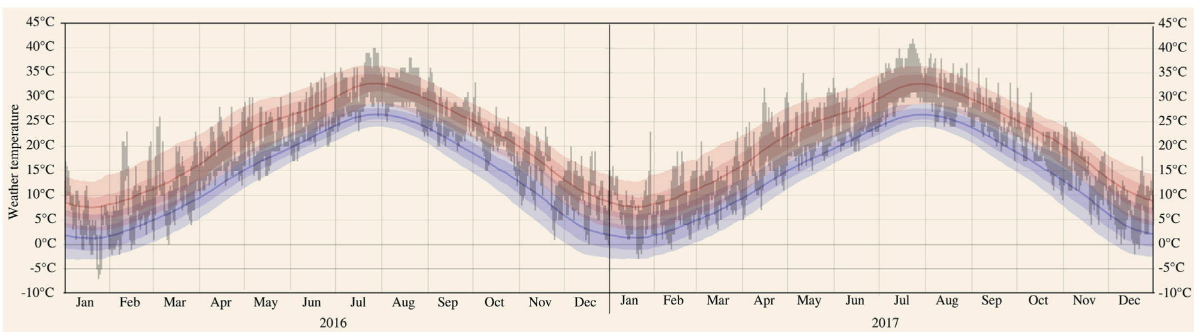

Fig. 7. Shanghai weather temperature data in 2016–2017 [60]. The daily range of reported temperatures (gray bars) and the 24-h maxima (red ticks) and minima (blue ticks), placed above the daily average of maximum (light red line) and minimum (light blue line) temperatures, with bands from the 25th to 75th and 10th to 90th percentile.

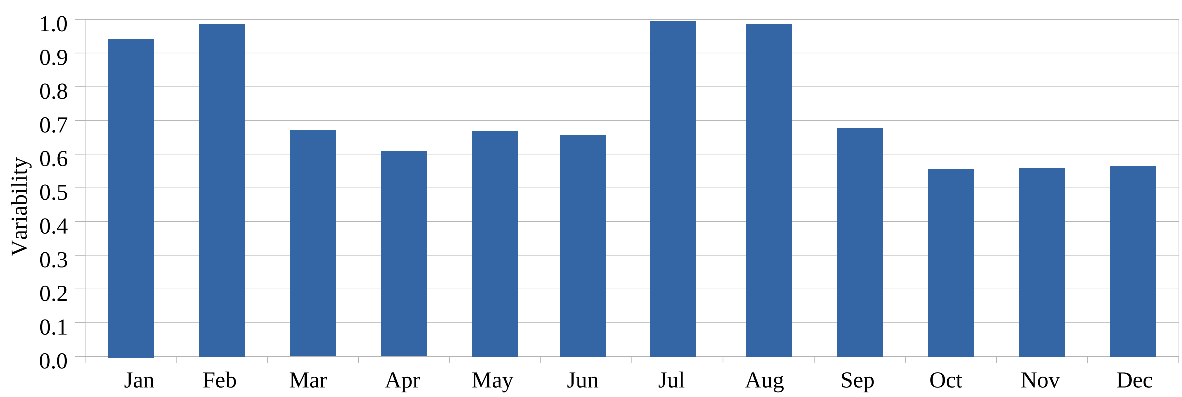

Fig. 8. Variability of energy demand of 2017.

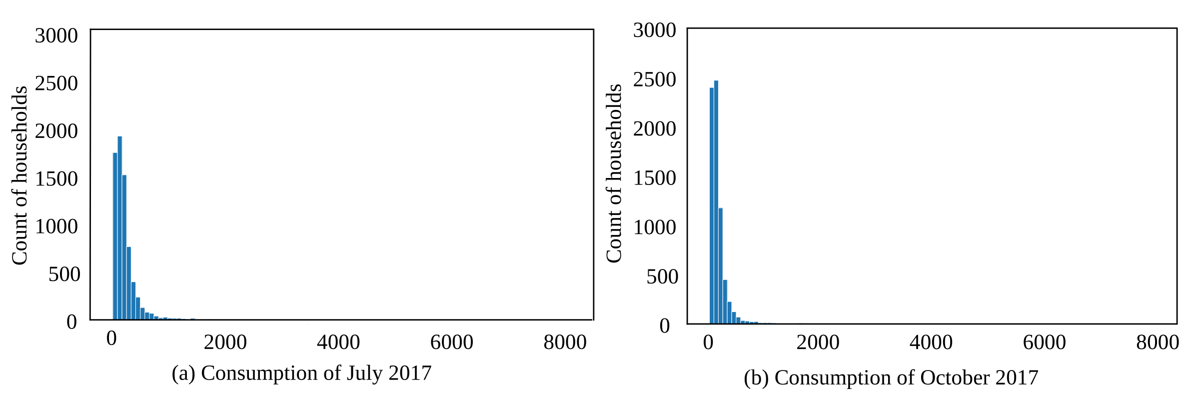

Fig. 9. Comparison of the household distribution between the months with maximum and minimum variability.

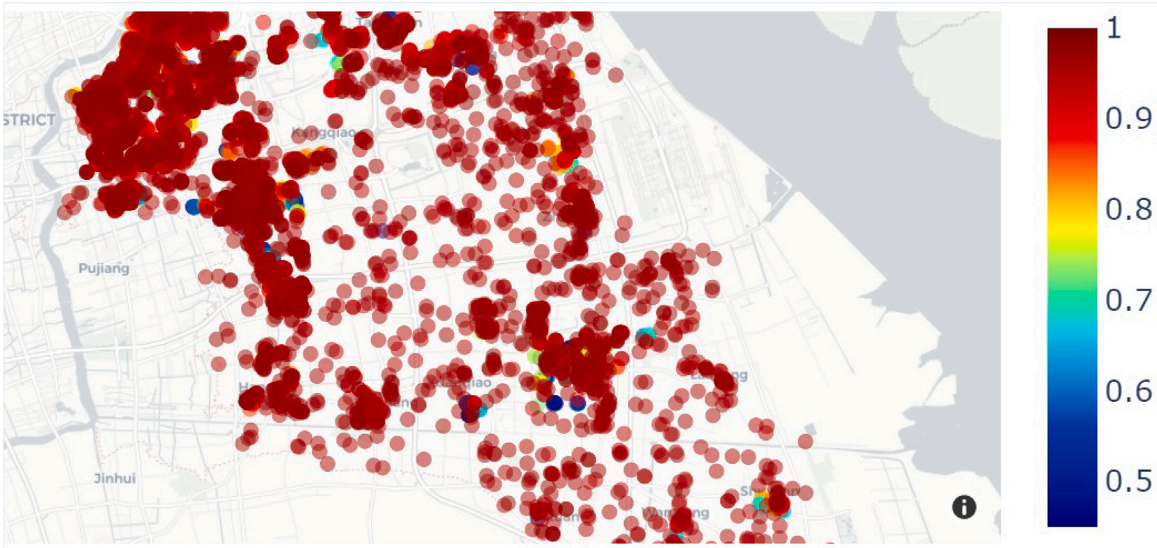

Fig. 10. Spatiotemporal variability of energy demand of July 2017.

and October 2017 (see Fig. 9). We see that daily consumption in July is more dispersed than in October (with higher bars). In other words, the daily consumption of most households is at a low level in October, for example less than 10 kWh. Therefore, its variability is not as high as in July.

We now assess the variability of energy consumption from the spatial dimension. Fig. 10 shows the variability of household demand in July, which is the largest difference from the previous study. As can be seen, most households show high variability above 0.8, while only a few households show lower variability. The results may not be very surprising because the variability is determined by the proportional difference between households, while household consumption may show extremes during the warmest month. For example, consumption may be much higher than usual if residents were at home due to the use of air conditioning, or it may remain low if residents were away for vacations. The upper left corner is the most dense area close to the Huangpu River, the traditional city center.

In fact, with the proposed visual analysis tool, users can interactively explore the variability of energy demand in both temporal and spatial dimensions. It can be interesting to observe the transition of variability on both dimensions, which can provide insight to utilities for energy supply, e.g., the provision for the bust of energy demand.

5.3. Segmentation analysis

Segmentation analysis provides a better understanding of households, which can help improve energy efficiency, provide personalized services, and implement specific energy use policies. In this section, we will first use the proposed visual analysis approach for segmentation, then compare it with the traditional k-means clustering method, and finally make a discussion. This study uses data from 2017 as the sample dataset.

Fig. 11 presents the results of using the method proposed in Section 4.4. The scatterplot in panel A represents all households after dimension reduction, which were organized according to the similarity of their consumption patterns. The more similar they are, the closer they are. To discover typical customer groups (i.e., clusters), users can simply drag with the mouse to select the points, then the corresponding consumption time series will be displayed on panel B and the geographical distribution of households will be shown in panel C. Through several interactions, we have identified seven typical patterns for our discussion, which are shown in subfigures (i) to (vii).

The first is the high-energy consumption pattern, as shown in subfigure (i). The characteristic is that it has consumption throughout the year and with high fluctuations. The reasons may be due to low-efficiency appliances, large households, or other unusual activities, such as hosting a party. The other extreme of energy use is the energy-saving pattern shown in subfigure (vi), which has relatively low consumption throughout the year. These may be small households, such as single-family households or low-income families that are very conservative in their energy use. Bimodal patterns are found to be the most common pattern, which has consumption peaks in both summer and winter (see subfigure (ii) and (iii)). The two subfigures show a slight difference in the summer and winter consumption variances. The bimodal pattern can be explained by the electricity consumption for cooling and heating in summer and winter. In fact, we have discovered that many households have a bimodal pattern, but with lower consumption. These may be single-family households or households with high-efficiency air conditioners. The balance pattern in subfigure (iv) shows a slight variation in electricity consumption throughout the year, i.e., it is not affected by the change in season, which may be those with good thermal insulation. The outlier pattern in subfigure (v) shows some abnormally high in the overall low consumption, which may be caused by human activity, switching errors, power leakage, or theft. The last is the idle pattern shown in subfigure (vii), characterized by randomly distributed consumption throughout the year. These may be new apartments or vacation apartments that have been randomly inspected and consume energy, such as turning on and off lights. Notably, households with inactive lifestyles also account for a considerable proportion in Shanghai.

We now use the traditional k-means clustering algorithm for segmentation to identify different customer groups, and compare it with the proposed visual analysis method. The k-means clustering is one of the most widely used algorithms for segmentation. However, users have to specify the number of clusters, k, before running the algorithm, which is often difficult. In this study, we first calculate the Sikhousette scores by increasing k from 2 to 16 and using the elbow method to identify the optimal number of clusters. The result in Fig. 12 indicates that $k = 6$ should be chosen for the algorithm. The corresponding clustering results are shown in Fig. 13. According to the results, the k-means algorithm successfully identified high-consumption patterns in subfigure (i) and five bimodal patterns arranged according to their intensity of consumption from high to low.

Therefore, comparing the proposed visual analysis and the k-means algorithm, we can observe that the proposed visual analysis method

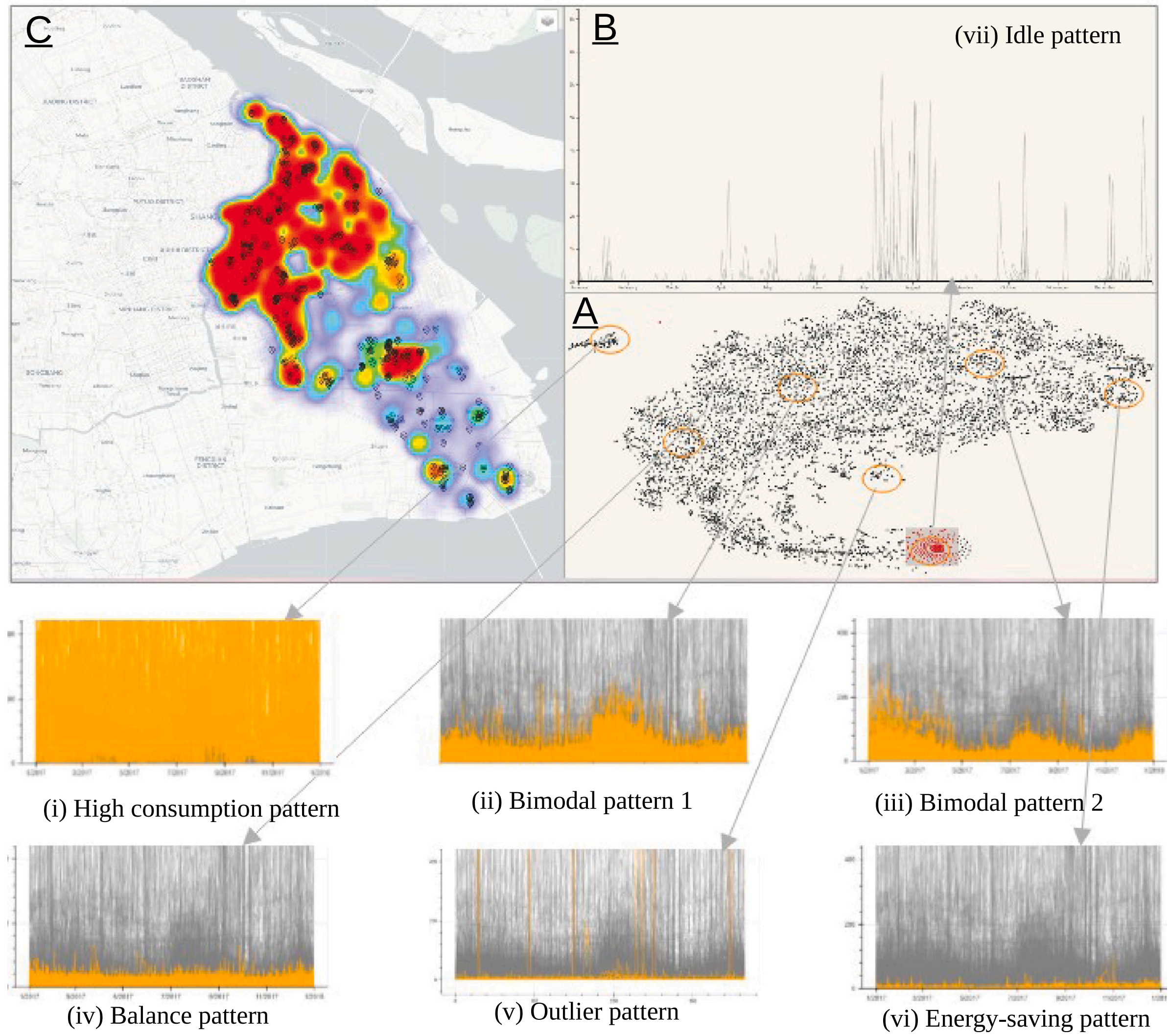

Fig. 11. Segmentation analysis based on interactive visual analysis approach.

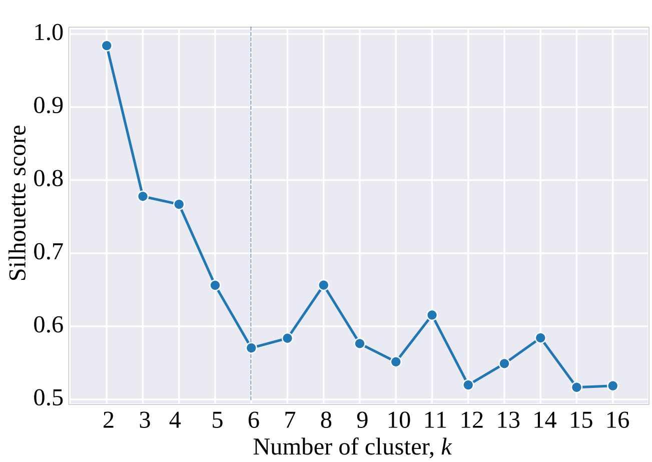

Fig. 12. Determination of the number of clusters.

can not only successfully identify all patterns using the k-means algorithm (recall that the proposed method can identify more bimodal patterns with slight variations), but also uncover other patterns that were not identified using the k-means algorithm. The explanation of the clustering results is similar to what we discussed in Fig. 11. The high-consumption pattern has high consumption with high fluctuations or constant high consumption but with low fluctuations. In the first case, it may be the use of low-energy efficient appliances, such as on/off, that causes large fluctuations; in the second case, it may be a large household that uses a lot of energy. For bimodal models, these may be households with air conditioners for heating in winter and cooling in summer. The difference in use may be due to the number of air conditioners.

Therefore, through the comparison, we can draw the following conclusions for the proposed method: (i) it is much more user-friendly and does not require specifying the number of clusters and rerunning the algorithm; (ii) human cognitive ability helps improve the results, which is achieved through multiple interactions; (iii) Although user-friendly, the identification process requires multiple trials and judgments, which requires some domain knowledge.

5.4. Demand shift analysis

Inspired by fluid dynamics, we analyze the spatiotemporal shift pattern of energy demand using the potential flow presented in Section 4.5.

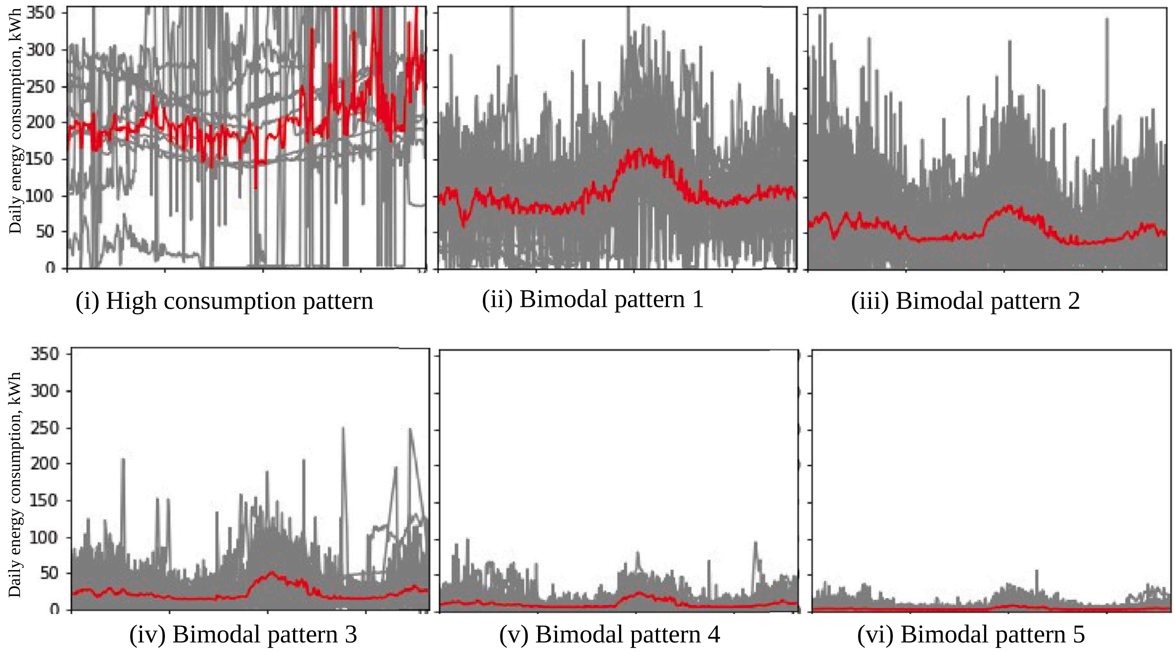

Fig. 13. Segmentation analysis based on interactive visual analysis approach.

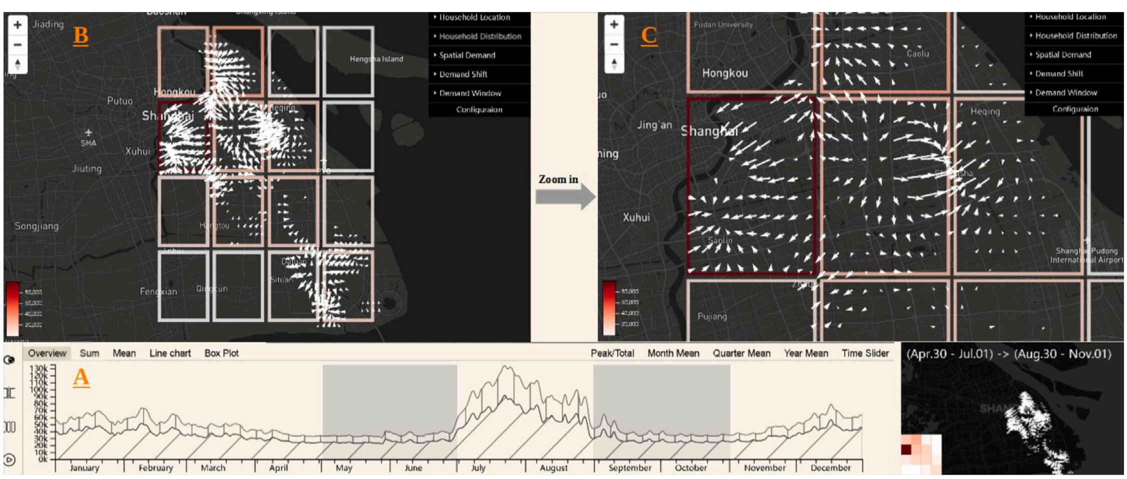

Fig. 14. An example of exploring spatiotemporal shift pattern of energy demand.

Fig. 14 shows an example of using the visual analysis tool to explore the spatiotemporal shift pattern of energy demand. With this tool, we can select the discrete time horizons in control panel A, and visualize the spatiotemporal demand shift on the flow map in panel B. In this example, we discretize energy consumption in the ranges, including April 30–July 1 and August 30–November 1, 2017, and observe shift patterns from both spatial and temporal dimensions. The flow map visualizes an interesting pattern of crowd energy demand behavior. According to the demand-shift badge, we can observe that energy demands are unbalanced in different regions. As shown, the arrows converge and diverge in the upper right area in Panel B, which is a highly dense residential area in Sandlin town. This shift pattern means a considerable change in the load in the area, exceeding 80,000 kWh. We zoom in on this area to obtain the flow map in Panel C, and we observe that the change rate in demand (through the length of the arrow) and direction are more clearly visible. Armed with this information, utilities can be advised to, for example, implement effective demand response programs, optimize their operations, or properly plan energy supply.

We now evaluate the spatiotemporal shift pattern between peak and off-peak hours within a day using the proposed potential flow method (see Fig. 15). The light red area is the traditional residential area of the Pudong district, while the areas on either side of the residential area are the commercial areas, including the traditional commercial area along the Huangpu River and the Zhangjiang HiTech Park. Most of the people who work in the two commercial areas live in this residential area. Thus, the flow arrows indicate that the high energy demand areas will shift to residential areas when people return home from work. The length of the arrow can roughly quantify the strength of the difference, i.e., the demand difference between peak and off-peak hours.

We now evaluate the demand shift model over different time horizons. Fig. 16 shows the shift of energy demand every two consecutive quarters in 2017. We can observe an interesting pattern from the three subfigures (i)–(iii). Clearly, the demand shifting patterns in subfigures (i) and (iii) are very similar, while the potential flows in subfigure (ii) show an opposite direction, which point to the light red area. The main reason is that Shanghai Disneyland Park and Zhangjiang HiTech Park are located in this area, and during the summer months in July and August, this area becomes one of the densest areas for tourists, including students and their parents, who visit Disneyland park and live in hotels in this area. Fig. 17 shows the corresponding quantitative

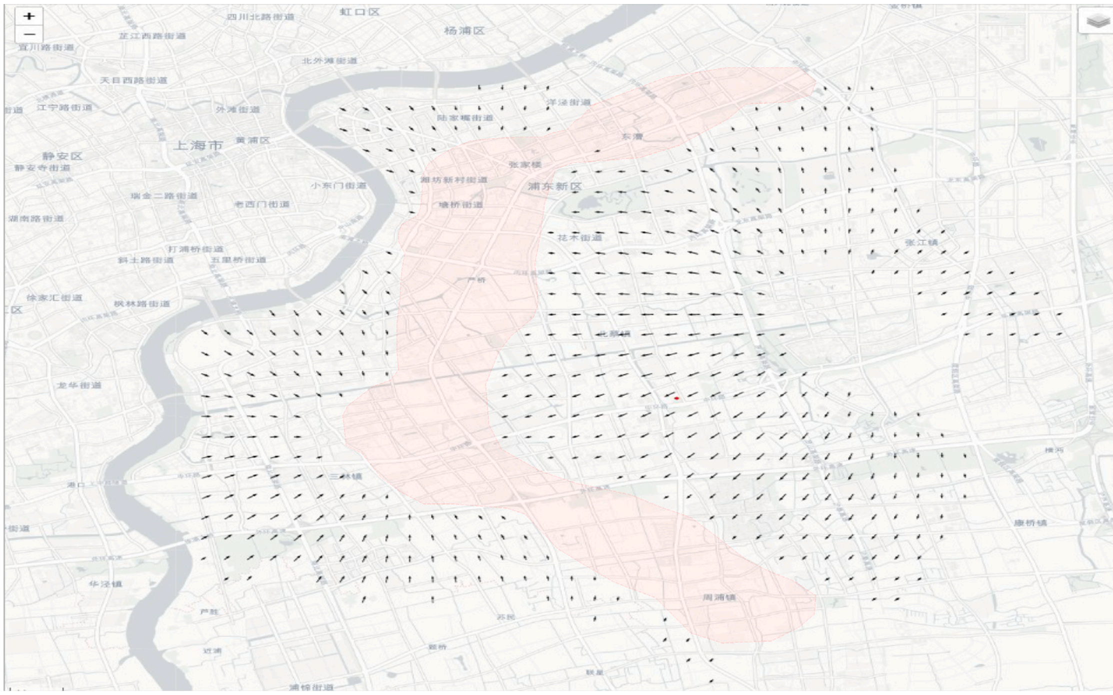

Fig. 15. An example of exploring spatiotemporal shift pattern of energy demand.

(i) Q1/2017 → Q2/2017 (ii) Q2/2017 → Q3/2017 (iii) Q3/2017 → Q4/2017

Fig. 16. Demand shift pattern across different time horizons (quarterly).

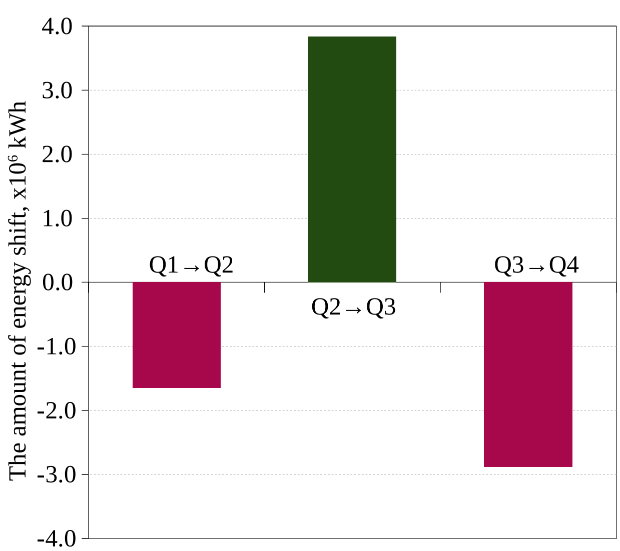

Fig. 17. The amount of energy shifting between two consequential quarters in 2017.

shift amount for every two consecutive quarters. As can be seen, the demand shifts from Q1 to Q2 and Q3 to Q3 are both negative, i.e., the mass energy demand was reduced, while only the shift from Q2 to Q3 is positive, which represents the high demand during summer.

In summary, with the proposed potential flow method, users can easily visualize the demand shift in the different spatial and temporal dimensions and quantify the difference in the shift. Users can also easily analyze the shifting pattern at multiple scale levels with the visual analysis tool. As in this study, the smart meter data used is at the daily resolution, but the proposed approach can be applied at a finer level, for example, hourly. Although demand shift can be obtained by analyzing feeder and substation data, the proposed method provides another option for utilities to plan energy supply. Our tool does not require station-level data but only the demand data, and is very user-friendly, as users can analyze the data by interactions.

6. Conclusion and future work

Smart meter data analysis is essential to improve energy management and services. In this paper, we proposed visualization and visual analysis methods to analyze energy consumption data for the three aspects of energy demand-side management, including consumption variability, customer segmentation, and energy demand shift. First, we modeled the dynamics of energy demand by introducing the concept of fluid dynamics *potential flows*, which enables one to visualize energy data from spatial, temporal, and spatiotemporal dimensions. Second, to better utilize the proposed method, we developed a visualization and visual analysis system that allows users to interactively analyze energy consumption data and incorporate their cognitive capability and domain knowledge into the analysis process. Finally, we comprehensively evaluated the proposed methods and the tool using real-world electricity consumption data from the Pudong district in Shanghai and compared the visual-based method with the traditional k-means algorithm in customer segmentation. Experimental results have shown that (1) customers with high energy consumption have higher variability; (2) the visual analysis-based segmentation method is capable of discovering more customer groups than the traditional method; and (3) the spatiotemporal dynamics of energy demand can be visually represented by a potential flow map, which is typically challenging for data-driven methods. All results of the proposed visual analysis method can be reasonably interpreted using energy domain knowledge, which validates its effectiveness in energy management.

There are several directions for future work. First, we will develop a powerful potential flow algorithm to achieve a better visualization effect of spatiotemporal energy demands. Second, we will refine the tool to support multisource and multigranularity utility data. Last, we will evaluate the system with more empirical data and apply the prototype to related applications in the energy sector.

CRediT authorship contribution statement

Junqi Wu: Methodology, Implementation, Writing - drafting. **Zhibin Niu:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Lizhen Huang:** Conceptualization, Writing – review & editing. **Per Sieverts Nielsen:** Conceptualization, Writing – review & editing. **Xiufeng Liu:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Project administration.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

Acknowledgments

The research was supported by the China Natural Science Foundation (61802278), the Flexibility for Smart Urban Energy Systems (FlexSUS) project funded by Horizon 2020 Research and Innovation Program of the European Union under the Marie Sklodowska-Curie Grant Agreement (754462), and the Flexible Energy Denmark (FED) project (8090-00069B) funded by Innovationsfonden.

References

[1] Bogdanov D, Ram M, Aghahosseini A, Gulagi A, Oyewo AS, Child M, Caldera U, Sadovskaia K, Farfan J, Barbosa LDSNS, et al. Low-cost renewable electricity as the key driver of the global energy transition towards sustainability. Energy 2021;227:120467.
[2] Duan H, Mo J, Fan Y, Wang S. Achieving China's energy and climate policy targets in 2030 under multiple uncertainties. Energy Econ 2018;70:45–60.
[3] Nematchoua MK, Sadeghi M, Reiter S. Strategies and scenarios to reduce energy consumption and CO2 emission in the urban, rural and sustainable neighbourhoods. Sustainable Cities Soc 2021;72:103053.
[4] Niu Z, Wu J, Liu X, Huang L, Nielsen PS. Understanding energy demand behaviors through spatio-temporal smart meter data analysis. Energy 2021;226:120493.
[5] Yazdanie M, Orehounig K. Advancing urban energy system planning and modeling approaches: Gaps and solutions in perspective. Renew Sustain Energy Rev 2021;137:110607.
[6] Gianniou P, Liu X, Heller A, Nielsen PS, Rode C. Clustering-based analysis for residential district heating data. Energy Convers Manage 2018;165:840–50.
[7] Zhou G, Soufan O, Ewald J, Hancock RE, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res 2019;47(W1):W234–41.
[8] Kalesinskas L, Cudone E, Fofanov Y, Putonti C. S-Plot2: rapid visual and statistical analysis of genomic sequences. Evol Bioinform 2018;14:1176934318797354.
[9] Fatula S. Detection of cancerous lesions in skeletal remains using visual methods and radiographs. Int J Osteoarchaeol 2021;31(5):916–25.
[10] Liu Q, Li Q, Zhu Z, Ye T, Ma X. Inspecting the process of bank credit rating via visual analytics. In: 2021 IEEE visualization conference (VIS). IEEE; 2021, p. 136–40.
[11] A. Leite R, Gschwandtner T, Miksch S, Gstrein E, Kuntner J. Neva: Visual analytics to identify fraudulent networks. In: Computer graphics forum, Vol. 39. Wiley Online Library; 2020, p. 344–59.
[12] Chetty S, Wang H, Goodwin S. Visualising the effect of COVID-19 on electricity consumption in victoria, Australia. In: Proceedings of the twelfth ACM international conference on future energy systems. 2021, p. 367–71.
[13] Alshardy A, Goodwin S, Liebman A. Data visualisation for remote area electrification decision-making: A case study of Indonesia. In: Proceedings of the twelfth ACM international conference on future energy systems. 2021, p. 382–8.
[14] Mammen PM, Mehrotra S, Kumar H, Ramamritham K. Want to reduce energy consumption, which floor should I prefer? In: Proceedings of the eleventh ACM international conference on future energy systems. 2020, p. 534–8.
[15] Batchelor CK, Batchelor G. An introduction to fluid dynamics. Cambridge University Press; 2000.
[16] Khan ZA, Ullah A, Haq IU, Hamdy M, Maurod GM, Muhammad K, Hijji M, Baik SW. Efficient short-term electricity load forecasting for effective energy management. Sustain Energy Technol Assess 2022;53:102337.
[17] Gao Z, Yu J, Zhao A, Hu Q, Yang S. A hybrid method of cooling load forecasting for large commercial building based on extreme learning machine. Energy 2022;238:122073.
[18] 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.
[19] Wang Y, Jia M, Gao N, Von Krannichfeldt L, Sun M, Hug G. Federated clustering for electricity consumption pattern extraction. IEEE Trans Smart Grid 2022;13(3):2425–39.
[20] Li J, Chen Z, Cheng L, Liu X. Energy data generation with wasserstein deep convolutional generative adversarial networks. Energy 2022;257:124694.
[21] Pekey M, Çelebi YD, Anıl C, Levi A. Private information inference of households from electricity consumption data. In: 2021 international balkan conference on communications and networking (BalkanCom). IEEE; 2021, p. 166–70.
[22] Gajowniczek K, Zkbkowski T. Short term electricity forecasting using individual smart meter data. Procedia Comput Sci 2014;35:589–97.
[23] Sharma A, Jain SK. A novel seasonal segmentation approach for day-ahead load forecasting. Energy 2022;257:124752.
[24] Barbour E, González M. Enhancing household-level load forecasts using daily load profile clustering. In: Proceedings of the 5th conference on systems for built environments. 2018, p. 107–15.
[25] Ruzbahani HM, Rahimnejad A, Karimipour H. Smart households demand response management with micro grid. In: 2019 IEEE power & energy society innovative smart grid technologies conference (ISGT). IEEE; 2019, p. 1–5.
[26] Langevin A, Carbonneau M-A, Cheriet M, Gagnon G. Energy disaggregation using variational autoencoders. Energy Build 2022;254:111623.
[27] Liu X, Bolwig S, Nielsen PS. SmartM: A non-intrusive load monitoring platform. In: International conference on business information systems. Springer; 2019, p. 424–34.
[28] Pamulapati T, Mallipeddi R, Lee M. Multi-objective home appliance scheduling with implicit and interactive user satisfaction modelling. Appl Energy 2020;267:114690.
[29] Armel KC, Gupta A, Shrimali G, Albert A. Is disaggregation the holy grail of energy efficiency? The case of electricity. Energy Policy 2013;52:213–34.
[30] Liu X, Golab L, Golab WM, Ilyas IF. Benchmarking smart meter data analytics. In: EDBT. 2015, p. 385–96.
[31] Liu X, Golab L, Golab W, Ilyas IF, Jin S. Smart meter data analytics: Systems, algorithms, and benchmarking. ACM Trans Database Syst 2016;42(1):1–39.
[32] Houle ME, Kriegel H-P, Kröger P, Schubert E, Zimek A. Can shared-neighbor distances defeat the curse of dimensionality? In: International conference on scientific and statistical database management. Springer; 2010, p. 482–500.
[33] Choksi KA, Jain S, Pindoriya NM. Feature based clustering technique for investigation of domestic load profiles and probabilistic variation assessment: Smart meter dataset. Sustain Energy Grids Netw 2020;22:100346.
[34] Haben S, Singleton C, Grindrod P. Analysis and clustering of residential customers energy behavioral demand using smart meter data. IEEE Trans Smart Grid 2015;7(1):136–44.
[35] Azizi E, Kharrati H, Mohammadi-Ivatloo B, Mohammadpour Shotorbani A. Clustering electricity big data for consumption modeling using comparative strainer method for high accuracy attainment and dimensionality reduction. J Energy Manage Technol 2019;3(1):1–7.
[36] Eskandarnia E, Al-Ammal HM, Ksantini R. An embedded deep-clustering-based load profiling framework. Sustainable Cities Soc 2022;78:103618.

[37] Wang J, Hazarika S, Li C, Shen H-W. Visualization and visual analysis of ensemble data: A survey. IEEE Trans Vis Comput Graphics 2018;25(9):2853–72.
[38] Borgeson S, Flora JA, Kwac J, Tan C-W, Rajagopal R. Learning from hourly household energy consumption: extracting, visualizing and interpreting household smart meter data. In: International conference of design, user experience, and usability. Springer; 2015, p. 337–45.
[39] Singh S, Yassine A. Big data mining of energy time series for behavioral analytics and energy consumption forecasting. Energies 2018;11(2):452.
[40] Guo H, Meggers F. Visualizing the exergy destructed in exergy delivery chain in relation to human thermal comfort with exflow. 2018.
[41] Francisco A, Truong H, Khosrowpour A, Taylor JE, Mohammadi N. Occupant perceptions of building information model-based energy visualizations in eco-feedback systems. Appl Energy 2018;221:220–8.
[42] Nageler P, Heimrath R, Mach T, Hochenauer C. Prototype of a simulation framework for georeferenced large-scale dynamic simulations of district energy systems. Appl Energy 2019;252:113469.
[43] Ellegård K, Palm J. Visualizing energy consumption activities as a tool for making everyday life more sustainable. Appl Energy 2011;88(5):1920–6.
[44] Fan C, Xiao F, Song M, Wang J. A graph mining-based methodology for discovering and visualizing high-level knowledge for building energy management. Appl Energy 2019;251:113395.
[45] Goodwin S, Dykes J, Jones S, Dillingham I, Dove G, Duffy A, Kachkaev A, Slingsby A, Wood J. Creative user-centered visualization design for energy analysts and modelers. IEEE Trans Vis Comput Graphics 2013;19(12):2516–25.
[46] Liu X, Golab L, Ilyas IF. SMAS: A smart meter data analytics system. In: 2015 IEEE 31st international conference on data engineering. IEEE; 2015, p. 1476–9.
[47] Kim S, Jeong S, Woo I, Jang Y, Maciejewski R, Ebert DS. Data flow analysis and visualization for spatiotemporal statistical data without trajectory information. IEEE Trans Vis Comput Graphics 2017;24(3):1287–300.
[48] Wang H, Huang H, Ni X, Zeng W. Revealing spatial-temporal characteristics and patterns of urban travel: A large-scale analysis and visualization study with taxi GPS data. ISPRS Int J Geo-Inf 2019;8(6):257.
[49] Yang Y, Cai J, Yang H, Zhang J, Zhao X. TAD: A trajectory clustering algorithm based on spatial-temporal density analysis. Expert Syst Appl 2020;139:112846.
[50] Luo X, Dong L, Dou Y, Zhang N, Ren J, Li Y, Sun L, Yao S. Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of Shanghai, China. J Clean Prod 2017;142:926–35.
[51] Liu D, Xu P, Ren L. TPFlow: Progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis. IEEE Trans Vis Comput Graphics 2019;25(1):1–11.
[52] Liu K. Pattern exploration from citizen geospatial data (Ph.D. thesis), Purdue University Graduate School; 2019.
[53] Ahmed Z, Beyene M, Mondal D, Roy CK, Dutchyn C, Schneider KA. ContourDiff: Revealing differential trends in spatiotemporal data. In: 2021 25th international conference information visualisation (IV). IEEE; 2021, p. 35–41.
[54] Suganthi L, Samuel AA. Energy models for demand forecasting—A review. Renew Sustain Energy Rev 2012;16(2):1223–40.
[55] Kannan KS, Manoj K, Arumugam S. Labeling methods for identifying outliers. Int J Statist Syst 2015;10(2):231–8.
[56] Rubinstein RY, Marcus R. Efficiency of multivariate control variates in Monte Carlo simulation. Oper Res 1985;33(3):661–77.
[57] Steinier J, Termonia Y, Deltour J. Smoothing and differentiation of data by simplified least square procedure. Anal Chem 1972;44(11):1906–9.
[58] Curtis AE, Smith TA, Ziganshin BA, Elefteriades JA. The mystery of the Z-score. Aorta 2016;4(04):124–30.
[59] Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. In: The craft of information visualization. Elsevier; 2003, p. 364–71.
[60] 2016–2017 Weather History in Shanghai China. 2022, https://weatherspark.com/ [Online; accessed 19-September-2022].