Energy 226 (2021) 120493

Contents lists available at ScienceDirect

Energy

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

Understanding energy demand behaviors through spatio-temporal smart meter data analysis

Zhibin Niu [a, b, 1], Junqi Wu [a, b, 1], Xiufeng Liu [b, *], Lizhen Huang [c], Per Sieverts Nielsen [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

A R T I C L E I N F O

Article history:
Received 2 October 2020
Received in revised form
8 March 2021
Accepted 25 March 2021
Available online 7 April 2021

Keywords:
Visual analysis
Pattern discovery
Energy consumption
Pattern explorer
Spatio-temporal patterns

A B S T R A C T

Energy demand-side management, especially empowered by the fine-grained smart meter data, plays a significant role in the rational allocation of energy, monitoring and supervision of energy consumption behaviors. Through the in-depth demand analysis including quantification of energy consumption dynamics and consumer preferences, energy decision-makers can develop reasonable and forethoughtful energy efficiency plans and demand-response programs. Previous work in energy-demand behavioral research relied primarily on ideal socio-economic models or data-driven approaches, both of which lack flexibility, intuition and interpretability. This paper proposes a novel spatio-temporal visual analysis approach for urban energy consumption pattern discovery in order to identify energy-saving potentials, plan energy supply and improve energy efficiency. In this approach, energy consumption time series are embeded into a two-dimensional scatterplot for coordinated visual exploration. Users can interactively explore and discover different patterns for decision-making purposes. In addition, we propose the method for modeling energy demand shift patterns based on a potential flow method and integrate it into a pattern exploration tool. The proposed approach is comprehensively evaluated through empirical studies using the real-world electricity consumption data from Pudong district, Shanghai. We identify five typical energy consumption patterns and demand shift patterns across different geographical locations, which can be well interpreted by the knowledge of energy consumption in the area of interest. The results demonstrate the effectiveness of the proposed approach and the tool. This tool can be integrated into smart energy systems for a better understanding of user energy consumption behaviors and preferences.

1. Introduction

Urban energy consumption accounted for 70% of the global energy supply in 2015. The International Energy Agency (IEA) also forecasts that 90% of global energy growth will be in cities, according to the official report [1]. The demand analysis of energy, including the form of electricity, heating, cooling, industry, buildings, and transportation to the identification of more achievable and affordable solutions to the transformation into future renewable and sustainable energy solutions is essential to the future energy systems [2–4]. Energy demand or energy consumption analysis lays the prerequisite for energy planning and policy-making in modern cities [5,6], thus also aligns and supports the energy strategy and industrial innovation target of the sustainable development goals (SDG) of United Nations [7]. Traditionally, energy consumers' preferences were reduced to an ideal socioeconomic metric, and conduct empirical studies, such as analysis energy data of a particular region or a period of specific citizens, businesses, and industries. In this paper, we develop a visual analysis based approach that allows users to investigate energy consumption patterns combining their expertise with observable/discoverable visual patterns. We propose an spatio-temporal analysis framework and develop a web-based user interface linking to energy data to support smart energy management.

Energy demand behavioral research can improve the understanding of consumption patterns, be constructive for demand-side

* Corresponding author.
E-mail addresses: mind3str@gmail.com (Z. Niu), wujunqi@tju.edu.cn (J. Wu), xiuli@dtu.dk (X. Liu), lizhen.huang@ntnu.no (L. Huang), pernn@dtu.dk (P.S. Nielsen).

[1] The authors contribute equally to this paper.

https://doi.org/10.1016/j.energy.2021.120493
0360-5442/

management, and impact relevant policy decisions [6]. With a thorough demand-side analysis, such as quantified energy consumption, understanding energy consumption behaviors and preferences, smart energy users or energy planners to make proper energy efficiency responses. Previous works in this field mainly use ideal socioeconomic models or data-driven models and has lacked the *flexibility, intuition and interpretability*. Economists prefer to employ several socioeconomic metrics computed from coarse-grained data, statistical measurements, and empirical numerical models to assist decisions. Such metrics are useful to understand trends or patterns, but have difficulty in revealing the reasons behind. The fast development of the Internet of Things in smart metering and remote sensing technologies in energy systems has accumulated massive fine-grained energy data. The data-driven approach utilizes machine learning technologies including frequent pattern discovery, clustering algorithms, and others, are borrowed to energy communities to discover, query, and understand the nontrivial, hidden, but potential useful patterns in the fine-grained energy data. However, such approaches usually require an in-depth understanding the energy domain problems, simultaneously grasping the machine learning algorithms and programming capability. This is challenging for the majority of energy domain researchers and industrial analysts.

We fill this gap by bringing a visual analytics based approach for the energy demand analysis. The energy consumers' location information, energy consumption fluctuations over time, and consumption patterns are visualized coordinately in the same web-based tool, Power Consumption Pattern Explorer (PCP). With the tool, smart system users can explore hidden patterns from overview to fine details through an *intuitive* way without requiring machine learning background knowledge. They can obtain insights into complex energy problems by amplifying domain expert's expertise with interactive knowledge discovery; this approach is more *flexible* than machine learning based approach, as the smart system users can interactively analyze crow energy demand behaviors by mouse brushing and observing spatial and temporal energy demand results coordinately; it has sufficient *interpretability* owing to the embedding algorithm can preserve global and local data structure and reveal semantics. As we are heading toward the era of Artificial Intelligence 2.0, hybrid intelligence with domain knowledge in the analysis can be an effective approach for shooting complex energy problems [8]. We believe that visual analysis will become an indispensable component for the next generation of smart energy systems, e.g., preventing, diagnosing and addressing emerging challenges.

In summary, this paper makes the following contributions:

- A visual pipeline for energy consumption pattern analysis is proposed, and a user-friendly tool is implemented to support knowledge discovery through user interactions.
- The methods for identifying typical patterns and for spatio-temporal pattern analysis are proposed. The typical patterns can help understand user consumption behaviors, while the shift patterns can help balance energy supply across different geographical locations. The typical patterns and demand dynamics can be explored intuitively and interactively.
- A visual analysis framework is proposed and the user-friendly tool is implemented. This paper evaluates the proposed method by an empirical case study using the real-world electricity consumption data from Shanghai Pudong district. Five typical consumption patterns are identified, as well as different spatio-temporal shift patterns. The case study validates the effectiveness of the proposed visual analysis framework in discovering spatio-temporal patterns at the urban scale.

The remainder of the paper is structured as follows: Section 2 reviews the related work. Section 3 presents the framework. Section 4 evaluates the framework through an empirical case study. Section 5 discusses the related issues. Section 6 concludes the paper and presents the future work.

2. Related work

2.1. Data-driven energy demand analysis

With the demand-side analysis, utilities can provide energy efficiency recommendations and personalized services (aka. energy demand side management). Since the seminal work of data-driven energy demand analysis in 1984 [6], the majority of the research have been focused on energy consumption pattern analysis and demand forecasting. Unsupervised learning based approaches have received an increasing attention in the energy consumption pattern analysis, due to the good interpretability. Algorithms, such as clustering, frequent pattern extraction, etc., help to uncover interesting patterns to better understand the underlying behaviors of the crowd. Whilst the energy demand forecasting relies heavily on supervised learning algorithms. It is estimated that more than half of the work (more exactly about 72%) in the energy sector utilize shallow learning algorithms, including artificial neural networks and support vector machine [9]. Recently, with the rapid advance of the deep neural network, some popular deep learning frameworks including LSTM, RNN and their derivatives [10] have been widely used for energy demand forecasting.

Pattern analysis was introduced to diverse energy consumer behaviorial analysis tasks. Hunt et al. create an energy demand model considering the trends and seasonal effects [11]. Gaussian distribution and the Kullback-Leibler divergence-based clustering method can be used to analyze household characteristics based on consumption patterns [12]. An association rule mining based quantitative approach is proposed to analyze residential electricity consumption patterns [13]. As occupant behavior is closely related to energy consumption, the frequent pattern mining is used to analyze variations of human behavior in Ref. [14], including variation in energy consumption, time and appliance use. Markov chains are extensively used to model occupant behavior and then to estimate energy demand and its fluctuations [15]. However, Markov chains have limitations in accurately capturing occupants' coordinated behavior and are prone to overfitting. Rich features related to the activities of coordinated occupants can be used to compare the behaviors between an occupant and its neighbourhood [16]. The customers with a similar load pattern can represent that they may have similar household structure or living habits. Therefore, load pattern analysis can be a crucial component for effective energy operation and management. Utilities can use segmentation analysis to improve their operations, design demand-response programs and provide personalized services. To date, clustering is one of the most used methods for customer segmentation analysis, by which time series are converted into reduced feature vectors and then grouped into different groups according to their distances [17].

2.2. Visual analysis for energy management

Visualization and visual analysis is an emerging interdisciplinary subject for analysis, reasoning, and decision-making through interactive visual interfaces [18]. Users can use visual analysis tools and technologies to obtain knowledge from massive, dynamic, uncertain or even conflicting data. Visual analysis enables users to detect expected information, explore unknown content, provide rapid, testable and understandable evaluations, and propose effective methods for evaluating communication. Visual visualization and visual analysis mainly include [19]: analytical reasoning that provides users with in-depth knowledge to directly support assessment, planning and decision-making; visual representations and interactions that use the high-bandwidth channel directly connecting the human visual system to the brain, and provide users with technical support to simultaneously observe, explore and understand a large amount of information; data representations and transformations that transform diverse data containing conflicting content and dynamic changes into data representations that can support visualization and analysis; supporting the production, presentation and dissemination of analysis results (Production, Presentation and Dissemination) - visual analysis results are transformed into communication information in the background to achieve an effective exchange of information with different audiences. The methodology has received an increasing research interest as it focuses on analytical reasoning facilitated by interactive user interfaces, and as a problem-driven data analysis technique, it helps users focus on the exciting aspects of data and improve data exploration efficiency, thus potentially extend human cognitive ability [20–22]. Using a variety of visualization techniques, humans can perform effective cognitive analysis, extract knowledge and reveal patterns from data. Visualization and visual analysis have been introduced in different application domains such as in public opinion analysis [23], research hotspot evolution [24], air pollution source analysis [25], financial risk management [26,27] and many others. Usually built on top of the statistical analysis or data mining layers, visual analysis adopts a human-machine interaction methodology to help target users such as decision makers, interactively gain insight into complex problems by combining them with domain knowledge. Since we consider energy demand as a problem in the analysis of spatio-temporal patterns, we briefly present the related work.

The visualization of spatio-temporal data has been extensively researched and applied in different fields [28,29]. A majority of the studies are trajectory analysis [30–32]. Pattern extraction can be applied to obtain significant latent patterns from the movement data. Space-time cube representation is an information visualization technique where spatio-temporal data points are mapped into a cube [33,34]. AirVis is designed to assist domain experts to efficiently capture and interpret the uncertain propagation patterns of air pollution based on graph visualizations [35]. Multidimensional spatio-temporal data are modeled as tensors and then decomposed to extract the latent patterns for comparison and visual summarization [36]. Flow maps are used to track clustering behavior, and direction maps draw on the orientation of vectors, are used to precisely identify the location of events [37]. Kim et al. proposed a gravity-based flow extraction model by extending a density difference model, which can effectively separate human movement from spatio-temporal data without using trajectory information [38]. A population-based vector field was proposed to visualize the dynamics of temporal and geographical demand. By representing transportation systems as vector fields that share the same spatio-temporal domain, demand can be projected onto the systems to visualize the relationships between them [39]. Miller et al. introduced the DayFilter process for building performance evaluation, which uses a set of temporal data mining techniques including SAX, clustering and visual analysis [40]. Within the system, discrepancies, or irregular daily patterns are filtered and marked for in-depth and detailed analysis for potential energy saving opportunities. Motifs are detected and grouped using k-means clustering algorithm.

In the energy sector, the visualizations such as graphs and bar charts have been extensively used to compare energy consumption over time. The seminal literature presents several power system visualization techniques to help analyze the relationships between network power flows using animation, contouring of bus and transmission line flow values, and interactive 3D visualization [41]. Coincidence factor-based heatmap is the visualization method used to identify peak demand changes and avoid power outage [42]. Calendar-type pixel visualizations, with color enhancement of anomaly scores, integrated with the spatial visualization, line graph and trees, are designed to detect anomalies of energy consumption data [43]. FigureEnergy is an interactive visualization tool that allows users to annotate and manipulate a graphical representation of their electricity consumption data; and annotate their past energy consumption by understanding when and how. To do so, a certain amount of energy was used in Ref. [44]. Operational performance is integrated with building information modeling (BIM) as a visualization dashboard to support the building energy management [45]. Ambient and artistic visualization for residential energy use feedback is explored, where Phyllotaxis design, Hive design and Pinwheel design in energy use are discussed [46]. Matches, Mismatches, and Methods for Multiple-View workflows for energy portfolio analysis are discussed [47]. Mosaic groups mapping encoded by household energy use combines with geodemographics to enable a better understanding energy user types in the UK [33]. GreenGrid is designed to explore the planning and monitoring of the Electricity Infrastructure. Geographic layout coming with a weighted network interface is designed to quickly identify where the system would be most likely to separate if an uncontrolled islanding event were to occur [48]. Liu et al. developed a data pipeline [49,50] and a dashboard, SMAS [51], to streamline the whole process of smart meter data analysis, including data pre-processing, cluster scaling, segmentation and visualization on a map.

This paper aims to analyze and discover energy demand shift patterns through our proposed novel interactive visual analysis framework. The discovery of shift patterns focuses on investigating the effects of the crowds' spatial mobility on energy consumption over time. The shift pattern analysis has received much less research focus than typical consumption pattern analysis. Most of the current work focuses on the spatial variability of energy consumption, sources and aggregation. Our method differs substantially from these existing works as ours can discover the spatio-temporal shift patterns of energy demand. This involves analyzing the dynamics of energy demand shifts on both spatial and temporal dimensions.

3. Problem statement and methods

This section first provides an overview of the visual analysis framework, then describes the pattern recognition methods, and finally introduces the visual analysis tool.

3.1. Overview

The proposed visual analysis framework has a three-layer architecture, consisting of a data layer, a visualization layer, and an exploration analysis layer, as shown in Fig. 1.

In the data layer (left in Fig. 1), the energy consumption data are collected and pre-processed before stored in the database. The raw data are usually not ready for subsequent analysis, e.g., with abnormal and missing values. Thus, data pre-processing is crucial in ensuring data quality for the subsequent pattern detection. The well-prepared data are then forwarded to the next layer, where visual analysis is performed.

The visualization layer consists of temporal and spatio-temporal pattern analysis components (middle in Fig. 1). Clustering high dimensional data is usually time-consuming and difficult to achieve good results [52]. Therefore, in this paper, t-SNE [53] is used to

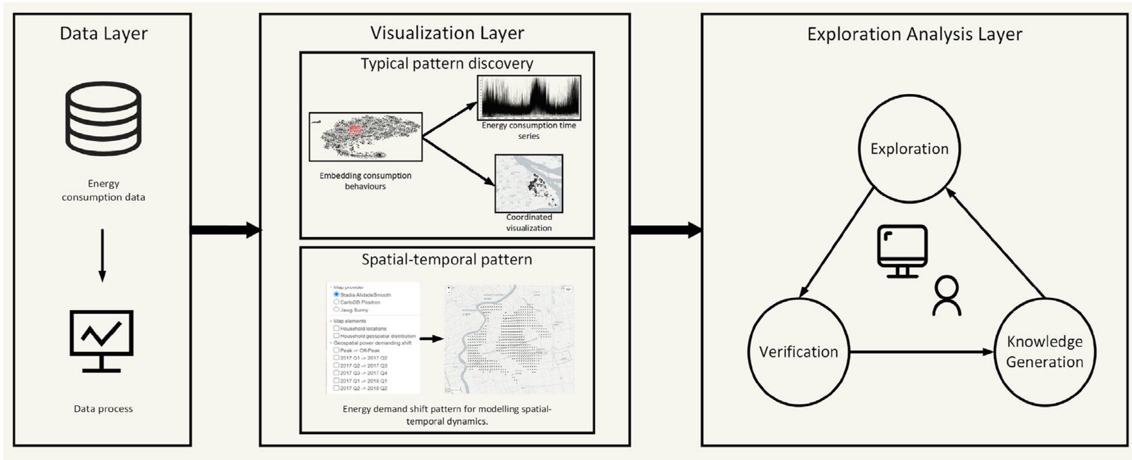


Fig. 1. Overview of the proposed visual analysis framework.

reduce the time series dimensionality before pattern analysis. It segments households according to the similarities of energy consumption patterns. In comparison, the potential flow-based modeling algorithm can capture the spatial shift patterns over time, including spatial shifts of energy demand and peak load. According to the shift pattern, the supply-side energy flexibility potential can be determined, and utilities can better plan energy supply at different temporal and spatial scales.

The exploration analysis layer is to gain knowledge from the visualization layer (right in Fig. 1). This layer provides a visual analysis interface where users can get insight into the data by interactions. Users can explore data in various ways, ask different questions, and answer their questions through visual analysis results on the view. The visual analysis tool, PCP, is designed to answer the following two classic questions, *“What is the consumption trend or pattern over time?”* and *“Does the crowd mobility affect energy demand?“*. The analysis process consists of three steps: data exploration through user interactions, idea verification and knowledge acquisition.

3.2. Data

The data used in this paper are daily resolution electricity consumption data from Pudong District in Shanghai, China. We collaborated with the Shanghai National Grid Company and collected the data from 2015 to 2018. The data comprise approximate 10,000 household energy customers, which were obtained by uniformly sampling from the full set of the customers, which is over one million. Notably, the data we obtained is all residential. Geographical coordination is also available to these customers. Table 1 shows a sample of the data consisting of customer identity (customerID), time series attributes (pap_r (total demand of the day), pap_r1 (peak period demand of the day, ranging from 6:00 to 22:00) and pap_r2 (off-peak period demand of the day, ranging from 22:00 to second day 6:00)), the measuring date (Date) and the geographical coordinate (Latitude and Longitude).

Table 1
Sample rows of the daily-grained energy demand data.

CustomerID	pap_r	pap_r1	pap_r2	Date	Latitude	Longitude
1100216777	6.03	5.17	0.86	2017-11-25	31.121.05	121.55181
1100216777	32.48	21.31	11.16	2017-07-17	31.121.05	121.55181
1100216777	5.77	4.90	0.87	2017-09-15	31.121.05	121.55181
1100216777	22.16	13.90	8.25	2017-08-25	31.121.05	121.55181
1100216777	13.81	6.67	7.14	2017-07-16	31.121.05	121.55181
1100216777	6.74	6.00	0.75	2017-11-12	31.121.05	121.55181

3.3. Data preprocessing

The spatio-temporal analysis of energy consumption data enables a better understanding of the environment’s effects and consumer behaviors over time. The first step of data analysis is data preprocessing, as the data may have quality issues, e.g., with noise, irregularities and missing values. Data quality can affect the results of visual analysis. However, data preprocessing is a non-trivial task that involves several cleansing steps. In this paper, these steps are performed to make the data ready for subsequent visual analysis, including anomaly removal, normalization, and dimensionality reduction, as described below.

Window-based convolution smoothing is used to smooth time series before further analysis. The smoothing operation creates an approximation function by removing noisy data and reconstructing the curve through interpolation to follow the trend of time series and capture the significant pattern. Smoothing can reduce random variations of a time series and provide a more accurate and intuitive view for potential patterns.

Z-Score is used to normalize energy consumption time series [54]. The Z-score is a signed fraction of a standard deviation by which the value of an observation or data point is higher than the average of the observed or measured values. The observations above the mean have a positive standard score, while the observations below the mean have a negative standard score. After calculating the standard score, the normalized energy consumption is plot into a uniform normal distribution, by which the impact of abnormal variations on the trend of a time series can be minimized.

3.4. Temporal pattern discovery for energy consumption

This section presents the embedding method to visualize energy consumption patterns that can be used, for example, to investigate consumer behaviors, segment customer groups, and design targeting demand-response programs.

Embedding technologies usually project high dimensional data into a lower space while keeping the global and local data relative structures. The most widely used embedding technologies are Principal Component Analysis (PCA) [48], a linear dimensionality reduction technology due to its efficiency and convenience. However, it usually deals well with data samples with lower than ten attributes and has difficulty processing real high-dimensional data sample features. Nonlinear dimensionality reduction techniques are also widely used. The representative nonlinear dimensionality reduction techniques include t-distributed Stochastic Neighbor Embedding (t-SNE) [53] and Uniform Manifold Approximation and Projection (UMAP) [55]. In this paper, t-SNE has an on par performance with UMAP on the energy consumption data (see Experiments), thus we choose t-SNE as the representative method and report the results.

Formally, t-SNE can be described as follows: given a set of n high dimensional data objects $x_1, \ldots, x_n$, the probability of similarity for two data objects, x_i and x_j, is represented as P_{ij}. According to [553, the similarity of x_j to x_i is a conditional probability, $P_{i|j}$. Whether x_j will be picked as x_i's neighbor or not is determined by the probability density under the Gaussian distribution centered at x_i. The conditional probability $P_{i|j}$ is defined as follows:

$$P_{i|j} = \frac{\exp\left(-||x_i - x_j||^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-||x_i - x_k||^2 / 2\sigma_i^2\right)} \tag{1}$$

where σ_i is the bandwidth of the Gaussian kernel, $\| \cdot \|$ is the distance. P_{ij} can then be calculated by the following formula:

$$P_{ij} = \frac{P_{j|i} + P_{i|j}}{2n} \tag{2}$$

Besides, the similarity probability of a data object to itself is set to 0. The purpose of t-SNE is to obtain a low dimensional spatial distribution reflecting the similarity P_{ij} as much as possible through iterative learning. To achieve this, it uses the method that is similar to obtain P_{ij} to calculate the similarity probability of the low dimensional data objects y_i and y_j, which is defined as follows:

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

where $y_1, \ldots, y_N$ are the data objects in a low dimensional space, $\| \cdot \|$ is the distance, k and l are between 1 and N. Here, the heavy-tailed Student t-distribution [56] (with one degree of freedom as same as the Cauchy distribution) is used to calculate the similarity between low-dimensional points, so that different objects can be placed further apart in a low dimensional space. The similarity probability of a data object to itself is set to zero.

Last, the position of the data point in the low dimensional space is minimized by Kullback-Leibler divergence between the probability distribution P of the high-dimensional space and the probability distribution Q of the low-dimensional space, which is defined as follows:

$$KL(P||Q) = \sum_{i \neq j} P_{ij} \log \frac{P_{ij}}{Q_{ij}} \tag{4}$$

The Kullback-Leibler divergence is minimized using the Stochastic Gradient Descent (SGD) method [57].

In this paper, Euclidean distance is used as the distance metric to measure the similarity of data points in t-SNE [58]. The distance between two points in Euclidean space is the length of a line segment between them, defined as

$$d(p,q) = \sqrt{(q.x - p.x)^2 + (q.y - p.y)^2} \tag{5}$$

where p, q are two data points shaped like (x, y).

With a visual analysis tool, users can interactively select nearby points and highlight time series patterns by superimposing them in a line graph. The position of scattered location icons (or points) on the map can help users understand the geographic distribution of objects with similar patterns. In this way, customers with similar consumption behavior can be discovered interactively using the tool.

3.5. Spatio-temporal demand shift modeling for energy consumption

Understanding the spatio-temporal patterns of energy consumption can help utilities improve operations, develop energy strategies and offer personalized services. For electricity consumption data, we found that there exist demand shifts across different geographical locations over time. This raises an interesting question about how to visualize these demand shifts in a user-friendly way to help utilities balance energy supply and improve flexibility. This section will describe the potential flow-based approach proposed to model energy demand dynamics across spatial and temporal dimensions.

In fluid dynamics, the potential flow describes the velocity field as the gradient of a scalar function: the velocity potential. As a result, a potential flow is characterized by an irrotational velocity field, which is a valid approximation for several applications. Therefore, we can observe that the geospatial energy demand changes continuously over time, and thus the demand is a continuum occupying a simply connected region in the time dimension with an irrotational characteristic. Inspired by the advancement of fluid dynamics and continuum mechanics, the continuum can be modeled as a potential flow [59].

This potential flow-based modeling algorithm is explained by the schematic diagram in Fig. 2. First, the energy consumption data are collected and sampled using the weighted sampling method [60]. This sampling method is used because it can effectively minimize bias [60]. Weighted sampling is defined as adding weights to the original data to measure its significance. The higher the weight of a data point, the more critical it is in the data set. What's more, the weighted sampled data can make the results of the kernel density estimation more accurate. In this step, the data points are defined as a ternary vector containing the latitude and the longitude of each user as well as the average power consumption as weights. The dynamics of the energy demand over time is visualized as a scatter plot at different times, e.g. t_1 and t_2 in Fig. 2a.

Then, the weighted sampled data are fed into the kernel density estimation algorithm, and the kernel density estimation matrix is calculated for different moments, and each value in the matrix corresponds to the strength of energy demand with a different GPS coordinate. We encode the matrix values as gradient colors to show the different energy demand strengths, as shown in Fig. 2b. In the end, generated by the difference obtained by subtracting the kernel density estimation matrix at different times, a graph of energy demand variation encodes the energy demand dynamics in both spatial and temporal dimensions, as shown in Fig. 2c.

By calculating the gradient of this matrix, we obtain the direction of the energy demand shift, which is the offset (dx, dy) for the original GPS coordinates. Original GPS coordinates (x, y) correspond to the arrow's tail coordinates, and the head coordinates of the arrow are $(x + dx, y + dy)$. The length of the arrow $\sqrt{(dx)^2 + (dy)^2}$, is the rate of energy demand shift. The longer the length of the arrow, the more rapid the demand changes. Such a flow map describes how energy demand shifts in space and time, quantitatively and qualitatively.

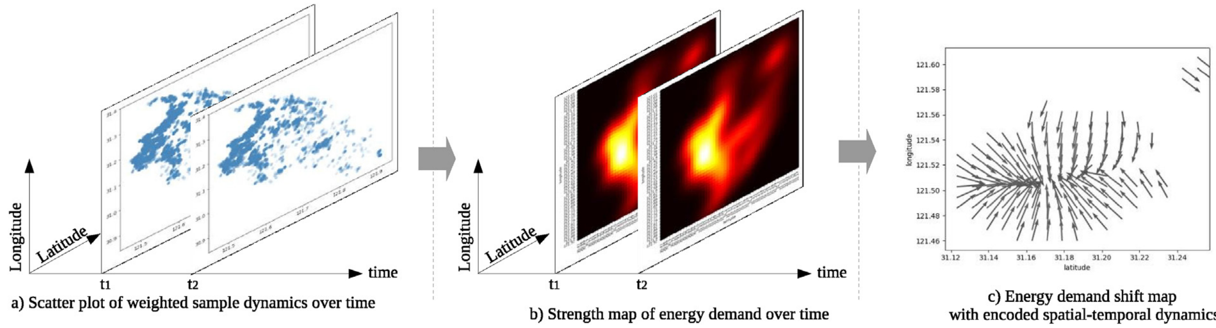


Fig. 2. Schematic diagram of energy demand shift modeling. The procedure for modeling the energy demand shift can be divided into three steps: 1) The locality data is weighted by the energy consumption at each moment; 2) Strength maps of the energy demand for each moments are calculated through kernel density estimation; 3) Modeling and visualization of the energy demand shift based on potential-flow (see Section 3.5).

To model the spatial energy demand, a kernel density estimation based approach is proposed to encode discrete household energy consumption into a continuous representation, which can efficiently generate a smoother vector field. In detail, let $x_1, \ldots, x_n$ be the samples of energy customers who are the GPS locations in this study. Each sample is a $2d$-variable random vector drawn from a common distribution described by a force function S. The following formula is used to estimate the energy demand strength of the population.

$$\widehat{S}_h(x) = \sum_{i=1}^{n} c_i K_H(x - x_i) \tag{6}$$

where $x = (lon, lat)^T$, $x_i = (lon_{i1}, lat_{i2})^T$, $i = 1, \ldots, n$, are 2D vectors; c_i is a normalized value of average energy consumption used to reweight demand strength with respect to geographic distribution; H is bandwidth (or smoothing), a $d \times d$ matrix, which is symmetrical and positive definite; K is the kernel function, which is a symmetrical multivariate density. The kernel function is defined as follows:

$$K_H(x) = \frac{1}{n}\Big|H|^{-1/2} K_H\left(H^{-1/2}x\right) \tag{7}$$

In this paper, the Gaussian kernel is chosen to estimate the strength of demand because it can provide a reasonable estimate for a small data set. However, for a large or medium data set, the Epanechnikov kernel can be a better option for its lower computational complexity [61].

With the kernel density matrix (strength map), the temporal dynamics of the energy demand over the time from t_1 to t_2 can be modeled by Equation (8). The temporal dynamics can be obtained by calculating the gradient of the demand strength difference.

$$Shift_{t_1,t_2} = \nabla(S_{t_2} - S_{t_1}) \tag{8}$$

The vector flow fields (the arrows) represent the shifts in energy demand, which are visualized and analyzed by the visual analysis tool developed in this paper.

3.6. Visualization and visual analysis user interface

This section presents the visual analysis tool, *Power Consumption Pattern Explorer (PCP)*.[2] The user interface design follows Shneiderman's mantra [20] of *"Overview first, zoom and filter, then details-on-demand"*. Fig. 3 shows the user interface, which consists of the following three coordinated views:

– *View A.* This view displays the spatial information of the sampled customers on a map. It supports users in selecting different map types, displaying the customers' geographic locations via markers, and visualizing the spatial distribution density via a heatmap and the energy demand shift via vector flow fields.

– *View B.* This view supports to display the energy consumption time series for the customers selected in View C.
– *View C.* This view is the temporal behavioral navigator that allows users to examine different energy consumption patterns or demand shift patterns. The closer the points are to each other, the more similar patterns they have.

The implementation of PCP follows a three-tier architecture, *data layer, analysis layer* and *visualization layer* (note that this is a software architecture that is different from the layered architecture of the visual pipeline presented in Section 3.1). In the data layer, PCP currently uses CSV files as the underlying database, but it would be preferable to use a data management system, which will be future work. Time series in the data layer is read and visualized in the web-based user interface. In the analysis layer, all algorithms are implemented in Python, including dimensionality reduction, consumption pattern, and shift pattern discovery algorithms. The user-interactive analysis can generate intermediate data, such as the flow field arrows (Scalable Vector Graphics (SVG)) and the associated latitude and longitude coordinates. These intermediate data are saved as a GeoJSON file in the data layer. In the visualization layer, HTML5, CSS, and JavaScript are used to implement the user interface. The JavaScript library, Leaflet.js [62] is used for map visualization and d3.js [63] is used to visualize time series and scatter plots as well as for interaction design.

4. Experiments

This section will describe the used data and report the empirical studies of energy consumption patterns and shift pattern discovery using the proposed visual analysis method.

4.1. Empirical study and exploratory data analysis

This subsection reports the data statistics of the household electricity consumption data. Fig. 4 shows the household spatial distribution by a heatmap of energy consumption density. The figure shows that the northern part of the Pudong district has a higher household density than the southern part. The following number confirms this visual analysis result. According to the census

[2] https://pcp.scicloud.site.

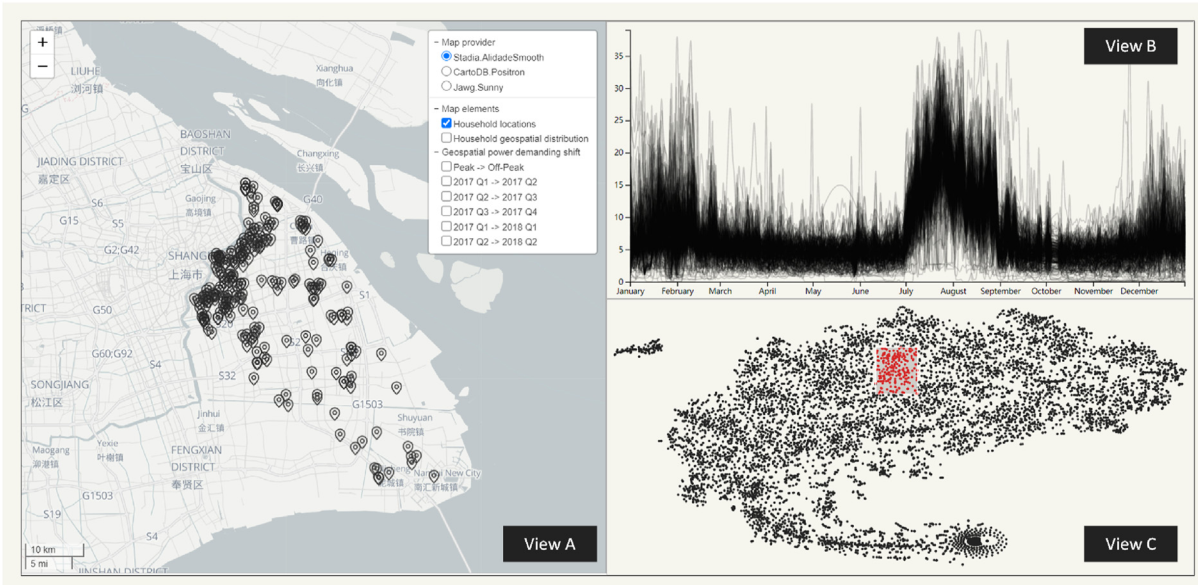


Fig. 3. The user interface of PCP. A user can select different clusters using the interactive scatter plot function of View C, then observe energy consumption patterns (View B) and geographical distribution (View A) to find the user's clusters of interest.

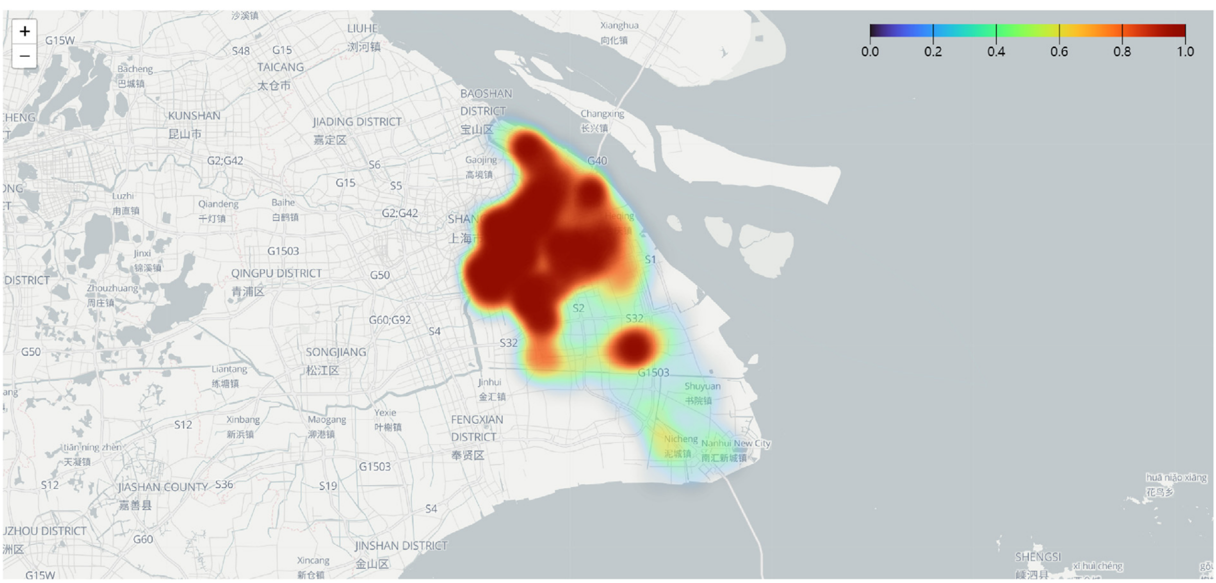


Fig. 4. Visualization of household spatial distribution by heatmap.

data collected in 2010, 2/3 of the population lives in the northern area, close to the city center, while 1/3 of the population lives in the Southern area. The visualization result validates that it is plausible to obtain the data using the uniform sampling method.

The distribution of electricity consumption is presented by *frequency histogram*. Figs. 5 and 6 show the distribution of the daily and annual consumption of households, respectively. Note that the abnormal hourly consumption values were removed by Tukey's range test [64] before this visual analysis is performed ($\alpha = 0.05$), which results in 1.2% of the rejected values, i.e., anomalies. The hourly and yearly consumption values both show a positive skewed distribution with a long tail, which means that most customers' consumption is to the left of the average. From the results, 1/4 of households consume less than 2.40 kWh per day and 3/4 less than 8.67 kWh per day; 1/4 of households consume less than 1145 kWh per year and 3/4 less than 3016 kWh per year. The zero value frequency is high in both figures, representing that no energy was consumed on a particular day or household. It may be for the reason that people are away or the home is not occupied.

Table 2 summarizes the yearly and daily consumption data statistics, which include min-max values, mean value, quantities, variance, standard deviation, skewness and tailedness (Kurtosis). The last two measures quantify the skewness of the values visually confirmed by the distribution figures.

Next, the aggregated daily electricity consumption of all customers and the weather temperature time series (high and low) are plotted on the same figure (see Fig. 7) to investigate their correlation. The electricity consumption has a bimodal pattern, suggesting that more energy was consumed in summer and winter. This pattern can result from high-temperature days with air

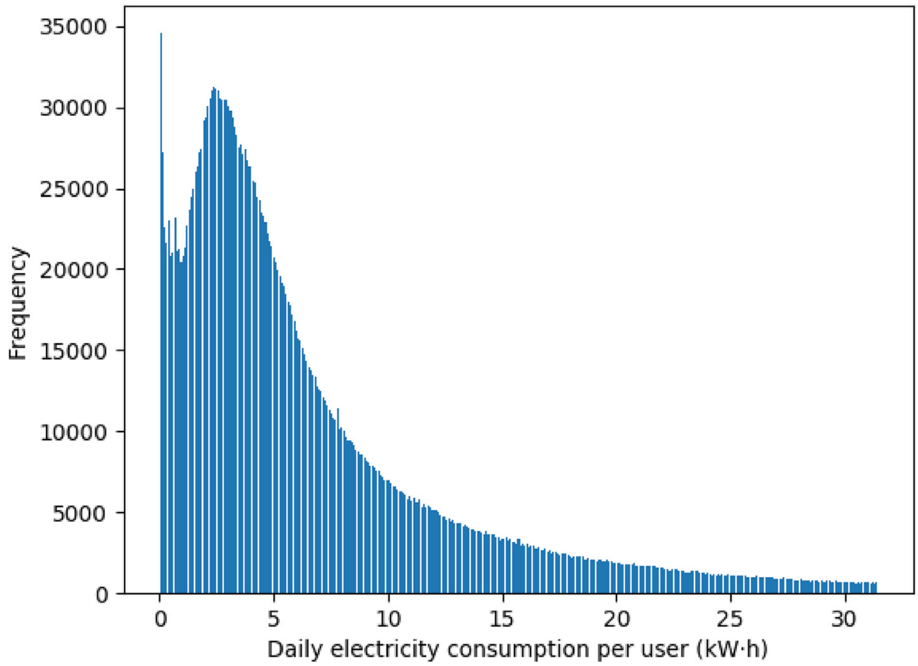


Fig. 5. The distribution of daily electricity consumption of households.

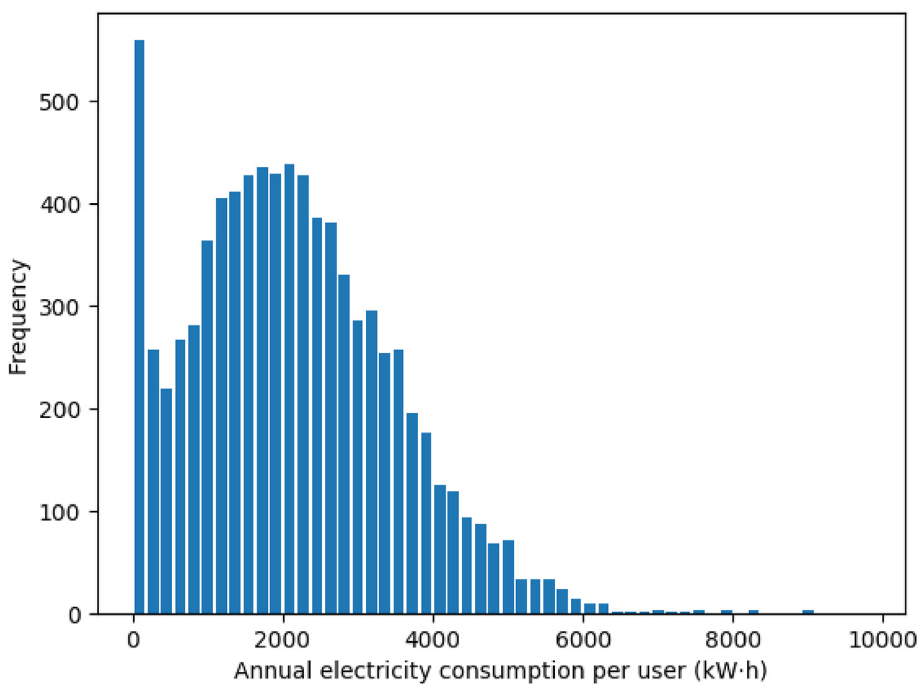


Fig. 6. The distribution of annual electricity consumption of households.

conditioning for cooling, and low-temperature days for heating. Nevertheless, there is a comfort zone between 15 °C and 25 °C, where there is neither heating nor cooling. In addition, the figure indicates that electricity consumption decreased during the first week of October. It is because this week is an extended holiday, China's national holiday. Many people were traveling and only baseload was consumed, for example, by refrigerators.

With the visual analysis tool, PCP, users can interactively explore hidden, implicit and useful information. The next subsection will describe how to discover typical energy consumption patterns with this tool.

4.2. Method selection: a comparison of embedding algorithms for energy data

The embedding algorithms t-SNE and UMAP have become the two most popular methods for data dimensionality reduction. Both methods work similarly to minimize the loss function by using

Table 2
Summary of the electricity consumption data statistics.

Statistics	Yearly (kWh)	Daily (kWh)
Minimum	0.00	0.00
Mean Value	2142	6.60
Maximum	9825	31.45
Q1 (Quantile 25%)	1145	2.40
Q2 (Median value)	2027	4.55
Q3 (Quantile 75%)	3016	8.67
Variance	1.83e+06	38.06
Standard deviation	1353	6.17
Mean absolute deviation	1084	4.60
Skewness	0.62	1.65
Kurtosis	0.56	2.54

gradient descent, which similar data points get closer, while different data points get further apart [65]. However, if the algorithm choice can make the computation results optimal when the data are different remains questionable.

To investigate suitable dimensionality reduction algorithms for our data, we use Euclidean distance as the similarity measure of dimensionality reduction algorithms in our experiments to compare and analyze the results of t-SNE and UMAP with different initialization (see Fig. 8). From the figure, it can be seen that both algorithms can identify similar energy consumption patterns. The algorithm can identify several different clusters in the scatterplots. However, the UMAP results are more compact and compressed in terms of distribution and have a larger empty area. As a result, the clusters of different patterns may overlap and the boundaries are unclear, making visual interaction difficult. In addition, this affects the aesthetics of the graph. In our data, the effect of the different initialization of the algorithms on the results can be seen in the difference of the scatter layout, but this does not affect the ability of pattern discovery.

4.3. Semantic analysis for energy consumption patterns through t-SNE projection

The visual analysis method for discovering typical temporal patterns is presented in the following. The dimensions of time series are reduced to a 2D space using the t-SNE algorithm, which is visualized as a scatter point in the scatter plot view. The time series of electricity consumption with similar patterns are tightly bundled. Therefore, typical patterns can be discovered interactively by selecting the points placed closely in the view. Fig. 9 shows the interactive view of the visual analysis system, in which five typical patterns were discovered, which are described below.

(i) **Idle pattern.** The points representing the idle patterns are aggregated at the bottom of the view like an ellipse (see Fig. 9). The idle pattern is characterized by year-round electricity consumption of almost zero. However, after zooming in with the exploration tool, at a fixed date each month, some slight variations can be observed for some households, meaning that these homes are possibility unoccupied but regularly inspected. In contrast, the homes with zero consumption maybe unoccupied throughout the year, e.g., new apartments.

(ii) **Bimodal consumption pattern.** It is a major pattern with two surge energy consumption periods in winter and summer. The bimodal pattern is most closely related to weather temperature changes (see the correlation in Fig. 7). This is a classic universal power pattern resulting from seasonal temperature changes, as residential electricity consumption varies with temperature. This pattern implies that more electricity is used in summer and winter, while less in spring and autumn. This is mainly due to the use of air conditioning for cooling in summer and heating in winter. The temperature in Pudong district is usually above 30∘C in summer. According to our investigation, we found that more people tend to turn on air conditioning in summer than in winter, which explains why the peak of consumption in summer is higher than in winter.

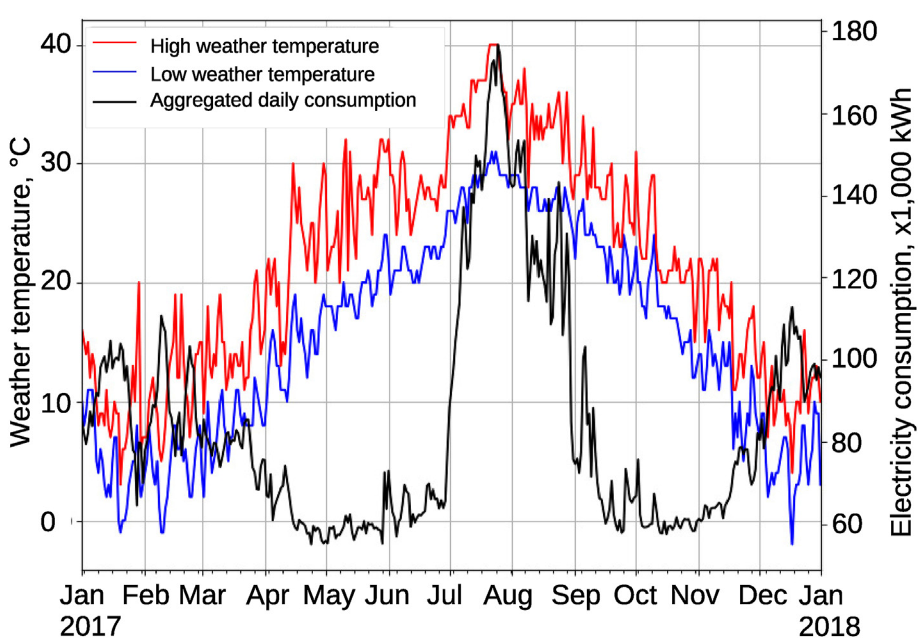


Fig. 7. The correlation between total daily electricity consumption and weather temperature. As indicated, the dashed line of energy consumption is, to some extent, correlated with the weather temperature. For example, in summer, energy consumption peaks when the temperature rises.

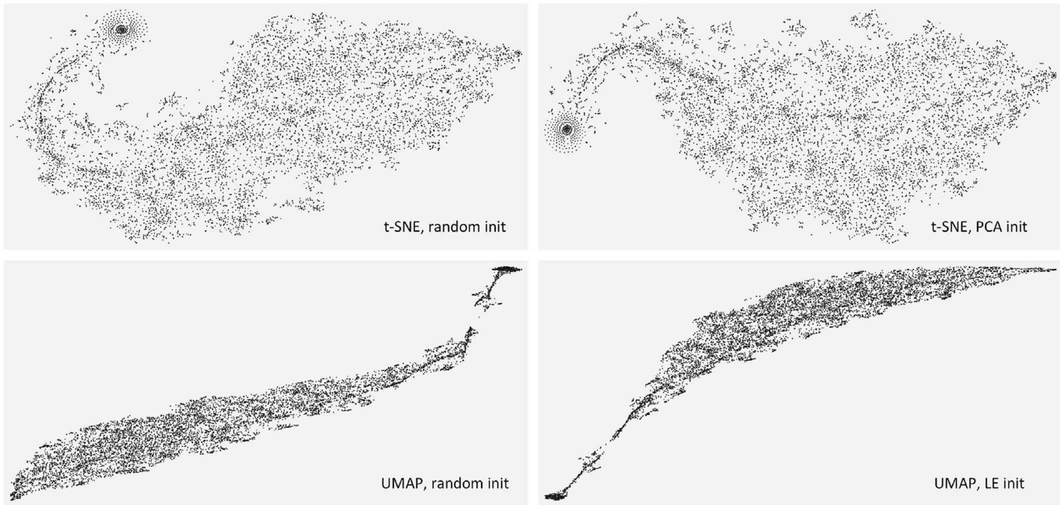


Fig. 8. Comparison of dimensionality reduction algorithms. We used random and PCA initialization for t-SNE (sklearn v0.24.0); and random and LE initialization for UMAP (umap-learn v0.5.0). As for UMAP, all other parameters were kept as default, except the number of iterations ($n\_epochs = 1000$). For t-SNE, all parameters were kept as default, except the early exaggeration ($early\_exaggeration = 4$), the learning rate ($learning\_rate = 1000$) and the number of iterations ($n\_iters = 1000$).

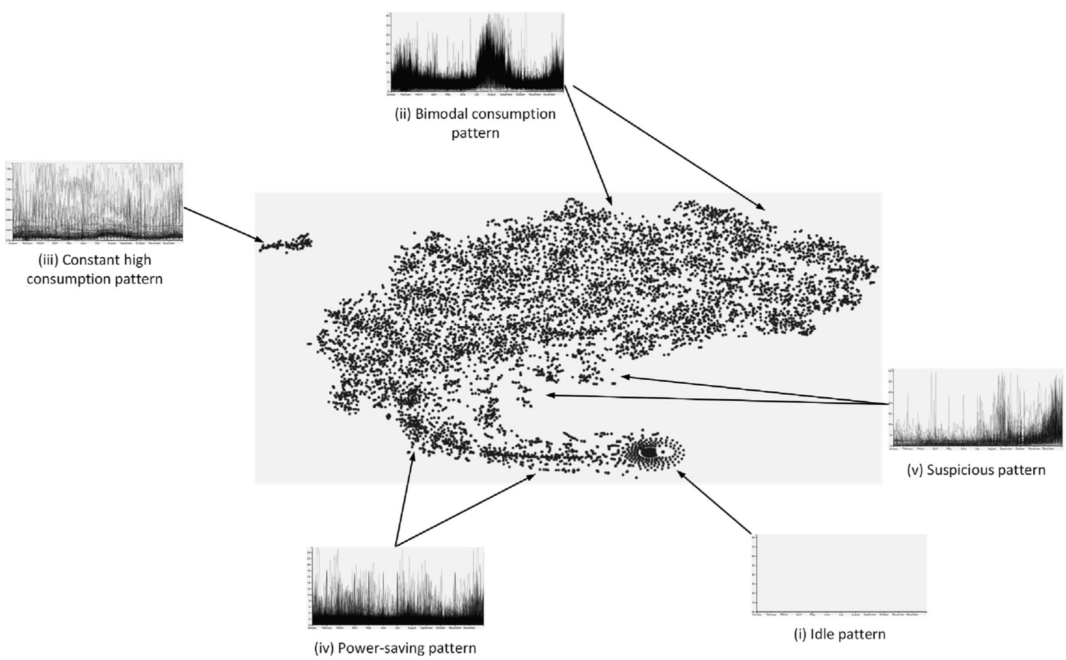


Fig. 9. Typical temporal power demanding patterns by tSNE.

(iii) **Constant high consumption pattern.** There is a cluster with a continuous high consumption mode separated from the main body, located in the upper left corner of the view. This mode is characterized by constant high consumption pattern throughout the year and small fluctuations within a fixed range. In contrast to the electricity consumption with a bimodal model, the consumption with this pattern is higher than the average daily consumption on most days. This pattern can occur for several reasons, for example, a household with low-efficiency equipment or a big apartment.

(iv) **Power-saving pattern.** The clusters with this pattern are in the bottom corner of the view, like a fish tail. The power-saving pattern has approximately the same shape as the bimodal pattern, but its consumption is much lower than the bimodal pattern. The peak in winter is almost the same as in spring and fall, and the summer peak season is relatively short, mainly in August. This pattern may be that these families live in new apartments equipped with energy-efficient appliances, or that they are low-income families who are very cautious about using too much energy.

(v) **Suspicious pattern.** The clusters with this pattern can be found in the lower central part of the view. This pattern is characterized by a low stable consumption before the fall. However, the consumption after September becomes high and fluctuates irregularly. The daily consumption is much higher than average, which can be considered abnormal. There are many reasons for this, for example, it can be caused by irregular living habits. Although it is difficult to determine the real reason, the pattern can provide users the tip for a tracking purpose.

4.4. Discovery of energy-demand shift pattern by potential-flow

This section will examine the shift pattern of electricity demand in Pudong District using the quiver plot method. For better visualization, the quiver plots in the following subsections show only those changes where the difference in demand strength is greater than 55% (see definition in Equation (8)).

4.4.1. Demand shift pattern at the daily scale

The Pudong district's power consumption shift pattern is visualized by the quiver plot method shown in Fig. 10. The result only shows the shifts on the daily temporal scale, but the method can also support the analysis of other temporal scales if a more acceptable resolution of the data is used.

The quiver plot shows that the high energy demand changes from commercial to residential areas when people go home after work. The arrows' tails point to the high demand area during the day, while the arrows' heads point to the area with a high demand after the shifts. The length of arrows encodes the shifts' demand strength-the shorter length of an arrow, the slower the shift rate.

The two residential areas are marked with light red belts. The left red belt covers several popular residential areas, including Sanlin Town, Zhoupu Town, Weifangxin Town, and Lujiazui Town. Both sides of the left belt are the main commercial areas, including Huangpu River district to the left and Zhangjiang HiTech park to the right. The Huangpu River district homes many industrial companies, while the Zhangjiang HiTech Park homes many office buildings. The energy demand in both commercial areas is shifting to the residential area after work. The right red belt is also a residential area, and most of the people living there work in the Zhangjian HiTech Park. Therefore, a similar shift pattern can be seen in Fig. 10.

4.4.2. Demand shift pattern at the quarterly or annual scale

This subsection examines the shift pattern of high energy demand on a quarterly or annual scale. Fig. 11 shows the shift pattern of the energy demand over the quarters of 2017 using the full set of the sampled data, *d*. The length of an arrow is proportional to the strength of the electricity demand shift. Accordingly, the shifts of the electricity demand first deviate from Q1 to Q2 from the commercial area (arrow tails) to the residential area (arrow heads) (see Fig. 11a). The shifts on the right-hand side converge from the residential area (arrow tails) back to the commercial area (arrow heads) from Q2 to Q3 (see Fig. 11b). However, the shifts of the electricity demand deviate from the commercial area (arrow tails) to the residential area (arrow heads) again from Q3 to Q4, like from Q1 to Q2 with little difference (see Fig. 11c).

The following explains these shift patterns. First, the shift pattern depicted in Fig. 11a may be caused by the traditional Chinese holidays in Q1, including the long Chinese Spring Festival (about two weeks), when most people stay at home and consume more energy than usual. Second, the shift pattern shown in Fig. 11b may be caused by the opening of Shanghai Disneyland in Q3 at the location near the residential area. Third, Fig. 11c may be due to the fact that Shanghai gradually entered the winter season. When the weather gets colder, more people will prefer to stay indoors, which consume more energy. This is especially the case in the residential areas.

Fig. 12 shows the shift pattern over a yearly interval, i.e., from Q1 2017 to Q1 2018 or from Q2 2017 to Q2 2018. Interestingly, although the shift strength difference threshold is set far below 55%, no significant shift in demand is observed, which means that the shift caused by the crowd behavior in a high granular time interval is negligible. It might also be dominated by the high-consumption customers, which will be discussed in the next subsection.

4.4.3. Sensitivity analysis

The shift in demand is mainly caused by the difference in energy consumption between different spatial areas. Since customers can

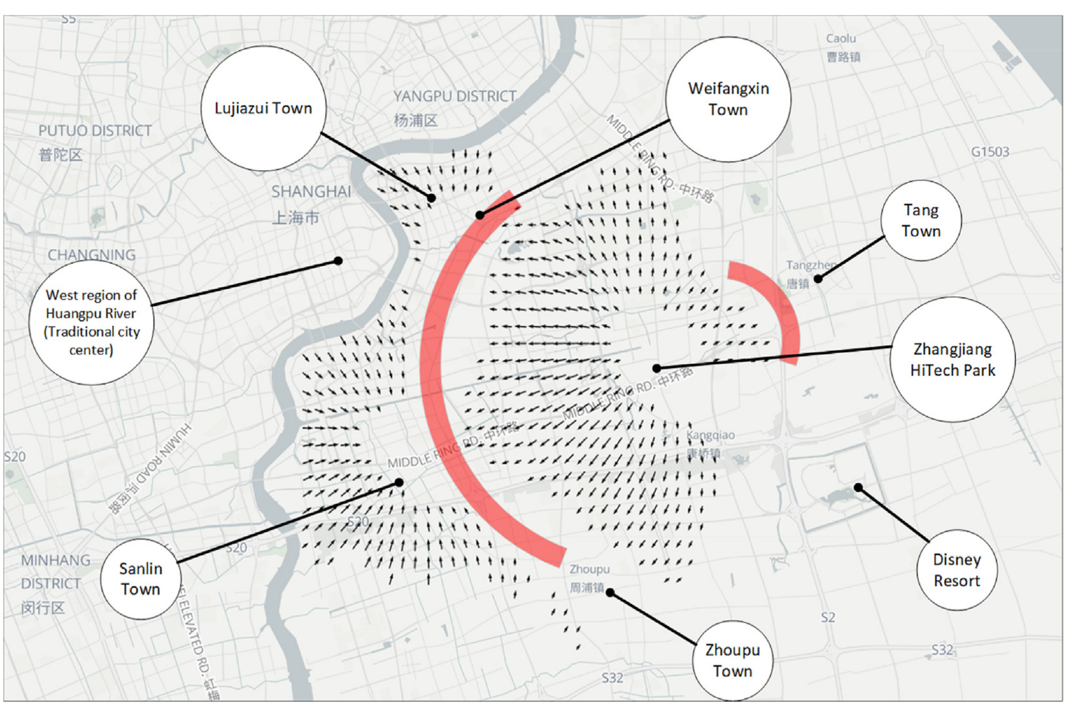
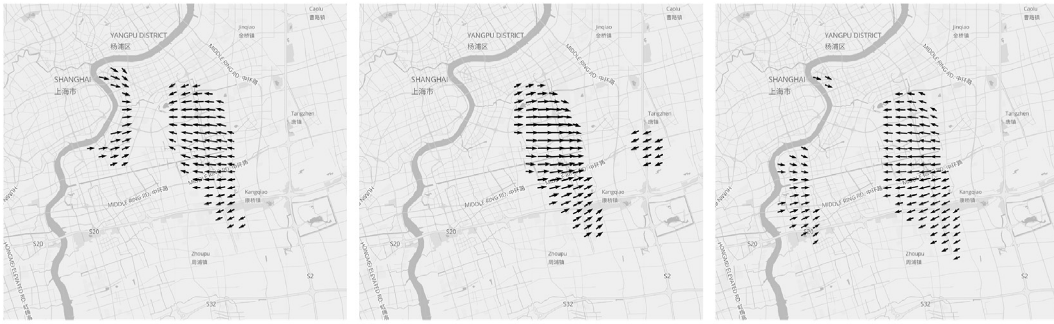


Fig. 10. The demand shift pattern at the daily scale.



a) From Q1 to Q2, 2017 b) From Q2 to Q3, 2017 a) From Q3 to Q4, 2017

Fig. 11. The demand shift pattern at the quarterly scale.

Fig. 12. The demand shift pattern at the yearly scale.

be classified according to their consumption intensity, it is interesting to examine the shift pattern sensitivity of different customer groups according to the consumption intensity. This study will divide the customers into several groups by percentile concerning their annual consumption. For simplicity, the full set of sampled data d is split into d' and d'' with respect to i-th percentile (see Fig. 13). In the following, the 90-th percentile is used for partitioning the data.

Fig. 14 shows the shift pattern of the two divided data sets, corresponding to a high and a low consumption group, respectively. The results indicate that the density maps of electricity demand strength have a very similar shape for both data sets, only with a subtle difference of density distribution. This makes it difficult to distinguish and interpret them plausibly. However, the dynamics of the density maps can be interpreted by the flow fields generated for them (see the third sub-figure on the left). Although both flow fields show a convergent trend to the central blank area, the flow map with anomalies shows an obvious divergence center and a convergence center on the map; while the flow map without anomalies has several convergent regions of the demand shifts, including a big cross-region shift and five small local shifts.

For further investigation, the quarterly shift patterns for the data sets d, d' and d'' are shown in Figs. 11, 15 and 16, respectively. By comparison, the number of local spatial shifts in Figs. 15 and 16 have little difference in scale, different from Fig. 11. However, the number of local spatial shifts in Fig. 16 is closer to the spatial shifts depicted in Fig. 11. After reviewing the data, it was found that the customers with high consumption can easily dominate the trend of spatial demand shifts. Therefore, the consumption variation over time for high-consumption customers is not as obvious as for low-consumption customers, but their consumption has a significant impact on the results of the potential flow-based modeling

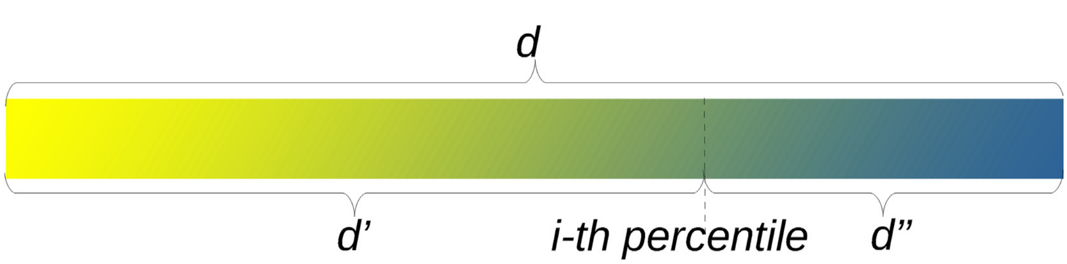


Fig. 13. The data sets partitioned according to percentile.

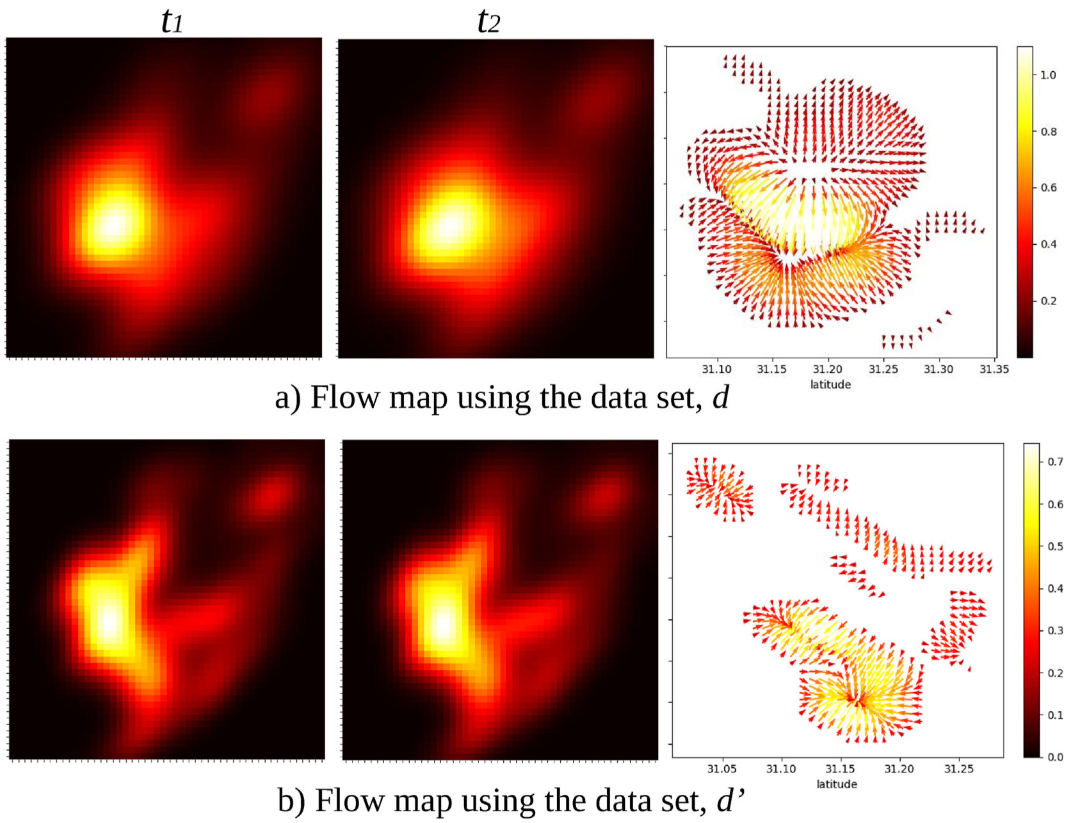


Fig. 14. The impact of consumption values on shift pattern.

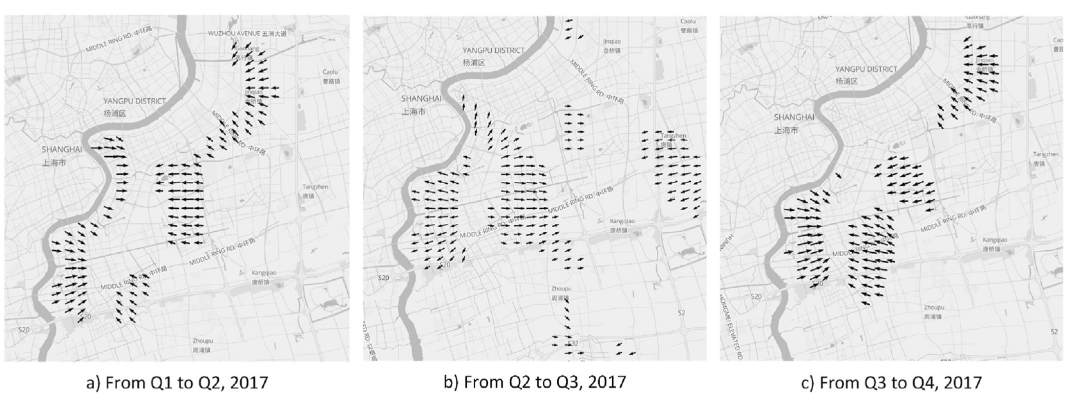


Fig. 15. Quarterly demand shift pattern using the data set.d'

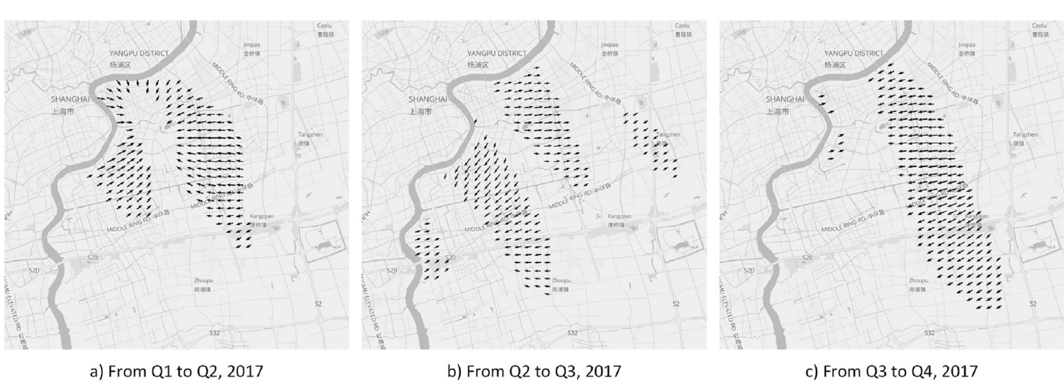


Fig. 16. Quarterly demand shift pattern using the data set.d''

algorithm. Perhaps by removing the data for high-consumption customers, we get a general pattern of daily energy consumption.

5. Discussion

The above section evaluated the proposed visual analysis framework and the PCP tool using a real-world case study from Pudong District in Shanghai. Some interesting points remain to be discussed. Although visual analysis or visual data mining has been around for a decade, it is mainly used in physics, bioinformatics and network security. The application to the energy sector is still in its infancy, but it has great potential, especially for understanding consumer behavior, fault diagnosis, interrelationships, etc.

First, this paper attempts visual analysis application in the energy domain and demonstrates its ability to detect typical energy consumption patterns and analyze spatio-temporal shift patterns of energy demand. Second, user interaction is the core of visual analysis, and therefore a user-friendly interface becomes an indispensable component for performing effective visual analysis. This paper introduces the visual analysis tool, PCP, which allows users to examine different patterns guided by their cognition and knowledge. In this empirical study, five typical patterns were discovered representing different consumption habits of customers behind. Domain knowledge helps visual analysis, such as asking relevant questions, performing useful analysis, and discovering meaningful results through user interactions. Third, there is room for the improvement of the proposed methods and the tool listed in future work. Although this tool supports the exploration of different time-granular data, as this study uses the daily resolution data, the study can only demonstrate the capabilities of discovering patterns in high resolution.

Ideally, pattern dynamics should be visualized by regular updates when finer granular data are available, and real-time changes can be detected. Last, the spatio-temporal shift pattern can be an essential tool for utilities to balance energy supply between different spatial locations. According to the sensitivity analysis presented in section 4.4.3, it is necessary to identify different customer groups according to their consumption intensity to capture local shift patterns better.

6. Conclusions and future work

Digitization of energy systems requires novel tools and methods that can help energy demand-side management. This paper presented a visual analysis framework supporting both spatial and temporal pattern analysis using energy consumption data. The paper first described the technique of dimensionality reduction, t-SNE, and discussed how to reduce high dimensional data to a low dimensional space and visualize them. The paper then presented the process for discovering typical consumption patterns, making it possible to recognize different customer groups with different consumption behaviors or living habits. The paper proposed the demand-shift pattern discovery method that supports the detection of demand changes according to spatial and temporal dimensions. To facilitate the use, the paper also implemented a web-based tool to support interactive visual analysis by users. In the end, the paper evaluated the proposed visual analysis framework and the tool by a real-world case study of pattern discovery using the electricity consumption data from Pudong district in Shanghai. The empirical study successfully identified five typical consumption patterns, and the demand-shift patterns across time and space. The experimental results validated the plausibility of the proposed method and its robustness.

There are several directions for future work. First, the framework for spatio-temporal analysis of energy demand can be combined more seamlessly. The current system supports an exploratory analysis of both temporal analysis and geospatial shifts of demand, but the analysis is separate and the coupled information has been inadvertently lost. It will be preferable to support the analysis for a unified model of analysis of both spatial and temporal dimensional information. Second, the computation of the geospatial shift of power demand is essentially based on the strength of local distribution changes, which mainly reflects partial dynamics, whereas in practice there is not only local dynamics but also a global trend. We seek to improve the flow generation algorithm and to provide a more accurate geospatial shift of demand. Third, we plan to apply our approach to analyze the data of more energy types and to improve the approach accordingly.

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.

Acknowledgements

The research was supported by the Natural Science Foundation of China (61802278), the European Union's Horizon 2020 Research and Innovation Program under the Marie Sklodowska-Curie Grant Agreement (754462), the Flexible Energy Denmark project (FED) funded by Innovationsfonden (8090-00069B) and the Flexibility for Smart Urban Energy Systems project (FlexSUS) (91352) funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 775970.

References

[1] Capuano L. International energy outlook 2018 (IEO2018). In: US energy information administration (EIA): Washington, DC, USA 2018; 2018. p. 21.
[2] Lund H. Renewable energy systems: a smart energy systems approach to the choice and modeling of 100% renewable solutions. Academic Press; 2014.
[3] Lund H, et al. Energy storage and smart energy systems. International Journal of Sustainable Energy Planning and Management 2016;11:3−14.
[4] Lund H, et al. Smart energy and smart energy systems. Energy 2017;137:556−65.
[5] Bhatia R. Energy demand analysis in developing countries: a review. Energy J 1987;8. Special Issue.
[6] R C N. (P. on Energy Demand Analysis. Improving energy demand analysis. National Academies; 1984.
[7] Griggs D, et al. Sustainable development goals for people and planet. Nature 2013;495:305−7. 7441.
[8] Pan Y. Heading toward artificial intelligence 2.0. Engineering 2016;2(4):409−13.
[9] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018;81:1192−205.
[10] Kim T-Y, Cho S-B. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019;182:72−81.
[11] Hunt LC, Judge G, Ninomiya Y. Underlying trends and seasonality in UK energy demand: a sectoral analysis. Energy Econ 2003;25(1):93−118.
[12] Hino H, et al. A versatile clustering method for electricity consumption pattern analysis in households. IEEE Transactions on Smart Grid 2013;4(2):1048−57.
[13] Wang F, et al. Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. Energy Convers Manag 2018;171:839−54.
[14] Singh S, Yassine A. Mining energy consumption behavior patterns for households in smart grid. IEEE Transactions on Emerging Topics in Computing 2017;7(3):404−19.
[15] Kavgic M, et al. A review of bottom-up building stock models for energy consumption in the residential sector. Build Environ 2010;45(7):1683−97.
[16] Baptista ML, et al. Accurate household occupant behavior modeling based on data mining techniques. In: Twenty-eighth AAAI conference on artificial intelligence; 2014.
[17] Kwac J, et al. Utility customer segmentation based on smart meter data: empirical study. In: 2013 IEEE international conference on smart grid communications (SmartGridComm). IEEE.; 2013. p. 720−5.
[18] Keim D, et al. Visual analytics: definition, process, and challenges. In: Information visualization. Springer; 2008. p. 154−75.
[19] Keim D, et al. Mastering the information Age - Solving problems with visual

analytics. Eurographics Association; 2010. ISBN: 9783905673777, https://books.google.com/books?id=rKxOMQAACAAJ.
[20] Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations". The craft of information visualization. Elsevier; 2003. p. 364–71.
[21] Card M. Readings in information visualization: using vision to think. Morgan Kaufmann; 1999.
[22] Ware C. Information visualization: perception for design. Elsevier; 2012.
[23] Cui W, et al. Textflow: towards better understanding of evolving topics in text. IEEE Trans Visual Comput Graph 2011;17(12):2412–21.
[24] Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci Technol 2006;57(3):359–77.
[25] Deng Z, et al. Airvis: visual analytics of air pollution propagation. IEEE Trans Visual Comput Graph 2019;26(1):800–10.
[26] Niu Z, et al. "regvis. Net–A visual bibliography of regulatory visualization". arXiv preprint arXiv:2007.03573 2020.
[27] Niu Z, et al. iConViz: interactive visual exploration of the default contagion risk of networked-guarantee loans. In: 2020 IEEE conference on visual analytics science and technology (VAST). IEEE; 2020. p. 84–94.
[28] Andrienko N, Andrienko G, Gatalsky P. Exploratory spatio-temporal visualization: an analytical review. J Vis Lang Comput 2003;14(6):503–41.
[29] Demšar U, et al. Analysis and visualisation of movement: an interdisciplinary review. Movement ecology 2015;3(1):5.
[30] Ferreira N, et al. Visual exploration of big spatio-temporal urban data: a study of New York city taxi trips. In: IEEE transactions on visualization and computer Graphics, vol. 19; 2013. p. 2149–58.
[31] Lin W, et al. A tube-and-droplet-based approach for representing and analyzing motion trajectories. In: IEEE transactions on pattern analysis and machine intelligence, vol. 39; 2017. p. 1489–503.
[32] Tominski C, et al. Stacking-based visualization of trajectory attribute data. In: IEEE transactions on visualization and computer Graphics, vol. 18; 2012. p. 2565–74.
[33] Goodwin S, Dykes J. Visualising variations in household energy consumption". In: 2012 IEEE conference on visual analytics science and technology (VAST). IEEE.; 2012. p. 217–8.
[34] Andrienko NV, et al. Space transformation for understanding group movement. IEEE transactions on visualization and computer Graphics, vol. 19; 2013. p. 2169–78.
[35] Deng Z, et al. AirVis: visual analytics of air pollution propagation. IEEE transactions on visualization and computer Graphics, vol. 26; 2019. p. 800–10.
[36] Liu D, Xu P, Ren L. TPFlow: progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis. IEEE Trans Visual Comput Graph 2018;25(1):1–11.
[37] Liu K. "Pattern exploration from citizen geospatial data". PhD thesis. Purdue University Graduate School; 2019.
[38] Kim S, et al. Data flow analysis and visualization for spatiotemporal statistical data without trajectory information. IEEE Trans Visual Comput Graph 2017;24(3):1287–300.
[39] Liu X, Yan WY, Chow JY. Time-geographic relationships between vector fields of activity patterns and transport systems. J Transport Geogr 2015;42:22–33.
[40] Miller C, Nagy Z, Schlueter A. Automated daily pattern filtering of measured building performance data. Automation in construction, vol. 49; 2015. p. 1–17.
[41] Overbye TJ, Weber JD. New methods for the visualization of electric power system information. In: IEEE symposium on information visualization 2000. INFOVIS 2000. Proceedings. IEEE; 2000. 131–16c.
[42] Yarbrough I, et al. Visualizing building energy demand for building peak energy analysis. Energy Build 2015;91:10–5.
[43] Janetzko H, et al. Anomaly detection for visual analytics of power consumption data. Comput Graph 2014;38:27–37.
[44] Costanza E, Ramchurn SD, Jennings NR. Understanding domestic energy consumption through interactive visualisation: a field study. Proceedings of the 2012 ACM Conference on Ubiquitous computing. 2012. p. 216–25.
[45] Gerrish T, et al. BIM application to building energy performance visualisation and management: challenges and potential. Energy and buildings, vol. 144; 2017. p. 218–28.
[46] Rodgers J, Bartram L. Exploring ambient and artistic visualization for residential energy use feedback. IEEE Trans Visual Comput Graph 2011;17(12): 2489–97.
[47] Brehmer M, et al. Matches, mismatches, and methods: Multiple-view workflows for energy portfolio analysis. IEEE Trans Visual Comput Graph 2015;22(1):449–58.
[48] Wong PC, et al. A novel visualization technique for electric power grid analytics. IEEE Trans Visual Comput Graph 2009;15(3):410–23.
[49] Liu X, Nielsen PS. Scalable prediction-based online anomaly detection for smart meter data. Inf Syst 2018;77:34–47.
[50] Liu X, Nielsen PS. A hybrid ICT-solution for smart meter data analytics. Energy 2016;115:1710–22.
[51] Liu X, et al. Benchmarking smart meter data analytics. In: EDBT 2015:385–96.
[52] Berndt DJ, Clifford J. Using dynamic time warping to find patterns in time series". In: Proceedings of the 3rd international conference on knowledge discovery and data mining. AAAIWS'94. Seattle, WA: AAAI Press; 1994. p. 359–70.
[53] van der Maaten L, Hinton GE. Visualizing Data using t-SNE". J Mach Learn Res 2008;9:2579–605. https://academic.microsoft.com/paper/2187089797.
[54] Patro S, Sahu KK. Normalization: A preprocessing stage". arXiv preprint arXiv: 1503.06462 2015.
[55] Becht E, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol 2019;37(1):38–44.
[56] Foss S, Korshunov D, Zachary S, et al. An introduction to heavy-tailed and subexponential distributions, vol. 6. Springer; 2011.
[57] Newton D, Yousefian F, Pasupathy R. Stochastic gradient descent: recent trends. In: Recent advances in optimization and modeling of contemporary problems; 2018. p. 193–220.
[58] Lee Rodgers J, Nicewander WA. Thirteen ways to look at the correlation coefficient. Am Statistician 1988;42(1):59–66.
[59] Batchelor CK, Batchelor G. An introduction to fluid dynamics. Cambridge university press; 2000.
[60] Fleuret F, Geman D. Stationary features and cat detection. In: J Mach Learn Res 2008;9:11.
[61] Epanechnikov VA. Non-parametric estimation of a multivariate probability density. Theor Probab Appl 1969;14(1):153–8.
[62] Agafonkin V. Leaflet: an open-source JavaScript library for mobile-friendly interactive maps. 2014. p. 2020. September 21.
[63] Bostock M, Ogievetsky V, Heer J. D3 data-driven documents". IEEE Trans Visual Comput Graph 2011;17(12):2301–9.
[64] Abdi H, Williams LJ. "Tukey's honestly significant difference (HSD) test". In: Encyclopedia of research design. Thousand Oaks, CA: Sage; 2010. p. 1–5.
[65] Kobak D, Linderman GC. Initialization is critical for preserving global data structure in both t-SNE and UMAP. In: Nature biotechnology; 2021. p. 1–2.