Understanding spatio-temporal electricity demand at different urban scales: A data-driven approach

Voulis, Nina; Warnier, Martijn; Brazier, Frances

DOI
10.1016/j.apenergy.2018.08.121

Publication date
2018

Document Version
Final published version

Published in
Applied Energy

Applied Energy 230 (2018) 1157–1171

Contents lists available at ScienceDirect

Applied Energy

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

Understanding spatio-temporal electricity demand at different urban scales: A data-driven approach

Nina Voulis*, Martijn Warnier, Frances M.T. Brazier

Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands

HIGHLIGHTS

- Analysis of electricity demand in neighbourhoods, districts and municipalities.
- Data-driven classification of electricity demand profiles of nearly 15000 urban areas.
- Three statistically different urban area types found: residential, business and mixed.
- Urban area types differ in terms of demand profiles and energy user composition.
- Often-used residential-type demand represents only a minority of analysed areas.

ARTICLE INFO

Keywords:
Urban energy
Electricity demand profile
Spatio-temporal demand
Energy transition
Clustering

ABSTRACT

Cities and communities worldwide are seeking to become more sustainable by transitioning to renewable energy resources, and by introducing electric transportation and heating. The impact and suitability of such technologies for a given area heavily depend on local conditions, such as characteristics of local demand. In particular, the shape of a local demand profile is an important determinant for how much renewable energy can be used directly, and how charging of electric vehicles and use of electric heating affect a local grid. Unfortunately, a systematic understanding of local demand characteristics on different urban scales (neighbourhoods, districts and municipalities) is currently lacking in literature. Most energy transition studies simplify local demand to household demand only. This paper addresses this knowledge gap by providing a novel data-driven classification and analysis of demand profiles and energy user compositions in nearly 15000 neighbourhoods, districts and municipalities, based on data from the Netherlands. The results show that on all urban scales, three types of areas can be distinguished. In this paper, these area types are termed “residential”, “business” and “mixed”, based on the most prevalent energy users in each. Statistic analysis of the results shows that area types are pairwise significantly different, both in terms of their profiles and in terms of their energy user composition. Moreover, residential-type demand profiles are found only in a small number of areas. These results emphasise the importance of using local detailed spatio-temporal demand profiles to support the transition of urban areas to sustainable energy generation, transportation and heating. To facilitate the implementation of the obtained insights in other models, a spreadsheet modelling tool is provided in an addendum to this paper.

1. Introduction

The transition to renewable energy generation and sustainable electric transportation and heating is becoming a priority for many cities and communities worldwide [1,2]. To support the integration of sustainable technologies such as solar photovoltaics (PVs), electrical vehicles (EVs), and heat pumps (HPs), it is important to understand the local conditions in the areas in question. In particular, the class of energy users and their demand profiles determine to a considerable extent how much renewable energy can be used locally directly, and how the local grid is impacted by renewable energy resources and by new loads such as EVs and HPs. Unfortunately, currently a systematic understanding of the demand profiles of urban areas is lacking in literature. Energy transition studies focus on the behaviour of new technologies (such as PVs, EVs and HPs), while simplifying existing demand to household demand only and omitting non-household energy users [3–6]. This paper addresses this knowledge gap by modelling, classifying and analysing electricity demand of both household and non-

* Corresponding author.
E-mail address: n.voulis@tudelft.nl (N. Voulis).

https://doi.org/10.1016/j.apenergy.2018.08.121
Received 8 June 2018; Received in revised form 15 August 2018; Accepted 27 August 2018

household energy users at three different urban scales: neighbourhoods, districts and municipalities. The paper provides novel insights in the existing differences between demand profiles and energy user compositions of different urban areas, at these three urban scales. Insights from this paper can be used to improve current and future energy system models and support renewable resource integration and electrification of transportation and heating systems.

The novelty of this paper lies in (1) the focus on urban areas representing groups of energy users instead of on single energy users, (2) a systematic, data-driven approach based on a combination of energy and urban data sources, and (3) the scale of the analysis, covering 11 570 neighbourhoods, 2725 districts and 403 municipalities in the Netherlands. Methodologically, the paper relies on established approaches from data analysis and statistics. The results show that at each urban scale, three types of area demand profiles can be identified, which are further called "residential", "business", and "mixed", based on the most prevalent energy users in each of these areas. At each scale, these area types are pairwise statistically different from each other. Moreover, the results show that only a small share of the nearly 15000 urban areas analysed have a residential-type demand profile, emphasising the need to include both household and non-household energy users in energy transition studies. To facilitate the use of the results and insights obtained in this paper, a spreadsheet tool based on logistic regression is made available online in Appendix B. This tool makes it possible to determine the type of hourly demand profile for an area of interest based solely on cumulative annual demand data of local energy users. Such annual data are more widely available than hourly data, yet do not suffice to understand the temporal interactions between local demand and new technologies such as PVs, EVs and HPs.

The remainder of this paper is organised as follows. Section 2 addresses the relevant literature in the domains of urban energy systems modelling and data analysis. Section 3 describes the datasets, and the data analysis and statistical methods used. Section 4 gives an overview of the results, that are further discussed in Section 5. Finally, Section 6 summarises the conclusions.

2. Literature review

This paper provides insights in demand profiles and energy user composition on three different urban scales. It thus contributes to the improvement of urban energy system models. The first part of this literature review presented in this section discusses the state of the art in urban energy system modelling from the perspective of the demand profiles currently used. The second part of this literature review describes data analysis methods applied in this paper.

2.1. The need for urban-scale spatio-temporal demand profiles

Energy system models can support planning and management decisions, in particular in the transition to renewable generation [7]. Traditionally, energy system models have been developed on two scales: (1) the national scale, to guide utilities and authorities in energy planning [8,9], and (2) the single building scale, to inform building managers of the building energy consumption and energy savings possibilities [10,11]. The transition to renewable generation and the electrification of transportation and heating systems raise new challenges, such as increased load peaks and bidirectional electricity flows [3]. These challenges are local phenomena that occur at an intermediate scale for which traditional energy system models were not designed [9,12–14]. Assessing the local impact of PVs, EVs and HPs therefore requires new models at a scale that is smaller than the national [8,9,13], but goes beyond the individual buildings [10,11]. Such new models can provide the necessary insights in the temporal and spatial matching between supply of renewable generation and local demand [12–14] and are currently under development (see [10,15] for comprehensive reviews). The remainder of this literature review addresses demand data used in these new urban scale energy system models.

Demand profiles are key data input for energy system models as they describe the fluctuations of demand in the area of interest. For traditional models, ones that cover either large areas (national scale) or small areas (building scale), the most relevant fluctuations of demand are those over time [8–11]. As the intermediate (urban) scale becomes more important in energy system modelling for the energy transition, demand fluctuations over both time and space become relevant [12–14]. To underscore the importance of both dimensions, such demand profiles are explicitly called *spatio-temporal demand profiles*.

The demand profiles traditionally used by utility companies cover large areas, often the entire area serviced by each of the companies. These profiles therefore do not have specific spatial information [8]. In addition to these temporal demand profiles, utilities have since long had access to spatial information of their assets through so-called geospatial information systems (GIS) [16]. With the advance of smart meters, increasing amounts of detailed temporal metering data of energy users at known locations are collected. These data are currently primarily used for billing purposes [17]. Recently, there is a growing interest in the integration of GIS and metering systems to generate detailed spatio-temporal demand profiles [16].

Researchers generally do not have access to data of utility companies [14,18]. Realistically representing spatio-temporal variations in demand is therefore challenging, and requires data analysis and combination techniques of data at other – national or building – scales, and non-energy related sources. Two types of approaches can be distinguished: top-down and bottom-up. The top-down approach uses macro-economic and spatially or temporally aggregated data. This approach results in a coarse description of demand profiles and energy user composition. The bottom-up approach relies on historical data (such as building typology, age and occupancy), and engineering data (such as the physical descriptions of energy consumption at the appliance level) [8,12,19,20]. This approach is the preferred option for urban scale energy system models, as it results in detailed descriptions of spatio-temporal demand variations [8,19,20]. However, it requires large datasets, that are not often publicly available. In particular, demand profiles of non-household energy users [21] and the spatial distribution of energy users across urban areas are often lacking [14,15,22,23].

The degree to which detailed spatio-temporal demand profiles are used in energy system models varies. Four different approaches to demand data handling can be distinguished in recent literature:

- The first approach leaves **demand data out of scope**, focusing instead on model development. For instance, Brownsword et al. [24] have developed a model that can simulate spatial and temporal variations in urban demand, yet the authors do not address the question of obtaining suitable demand profile data [24]. Ramirez Camargo et al. [6] propose a procedure that can assess the local impact of PVs on urban scales, introducing a detailed representation of solar power generation and its local impact, however, without taking local demand into account [6].
- The second approach accounts for **temporal demand variations**, typically assuming household demand only, and thus neglecting spatial demand variations that exist in urban areas due to local differences in the number and the type of household and non-household energy users. Most energy system modelling studies follow this approach. Only a few examples are briefly discussed here. For instance, Morvaj et al. [3] explore the impact of power system decarbonisation in urban districts, assuming a district of 55 exclusively residential buildings in their case study [3]. Paevere et al. [4] and Neaimeh et al. [5] study EV charging profiles and their local impact, also considering household demand only, although Neaimeh et al. explicitly include charging at non-household charging points [5].

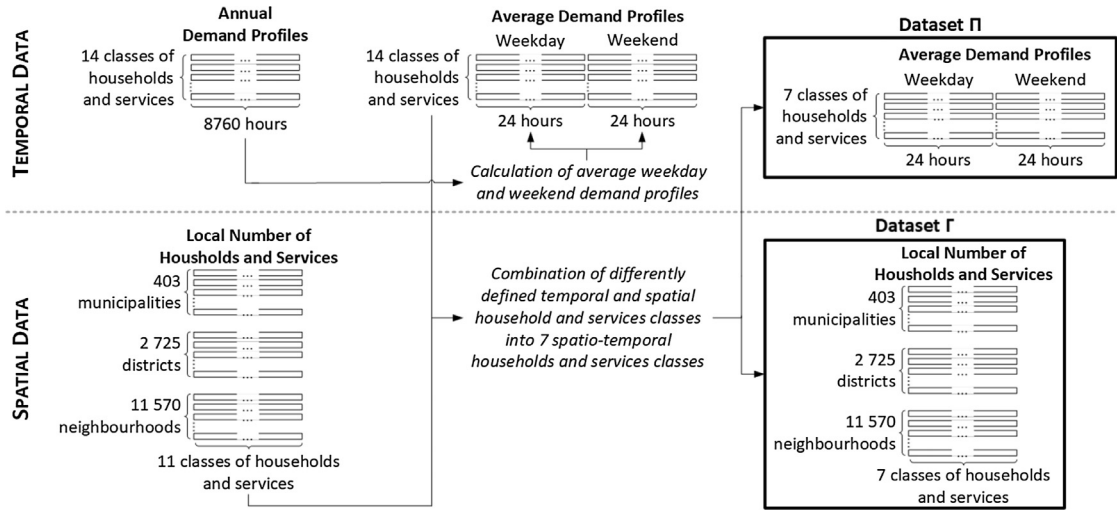

Fig. 1. Flow diagram of the data gathering and processing phase. The gathered data contain two datasets: (1) *temporal* annual demand profiles of households [36,37] and services [38], and (2) *spatial* local registration data of the number of households and services in 14698 urban areas (Dutch municipalities, districts and neighbourhoods) [39,40]. Two data processing steps are depicted. First, for the temporal data, average weekday and weekend demand profiles are calculated from the annual demand profiles. Second, the temporal and spatial datasets are made compatible. Both datasets pertain to households and services, however, the definition of service classes differs between the datasets. A single set of mutually compatible household and service sector classes is created by combining and scaling energy user classes of the two original datasets (for details, see Table A1 in the Appendix). The result of this phase are two mutually compatible datasets, called Π and Γ. Dataset Π has a dimension of 7 energy user classes × 2 day types × 24 h. Dataset Γ has a dimension of 14698 urban areas × 7 energy user classes.

- The third approach addresses **spatial demand variations**, while using cumulative annual data and thus not taking temporal demand variations into account. For instance, Yamaguchi et al. classify urban districts according to building type to gain a better understanding of spatial variations in energy use and CO_2-emissions [25]. Howard et al. have built a model that estimates spatial heating and cooling end-use intensity [13]. Chow et al. have proposed a method for spatial demand forecasting [26]. Alhamwi et al. have developed an open-source GIS-based platform for spatial demand modelling, leaving the temporal component for future research [27].
- The fourth approach develops and uses **spatio-temporal demand profiles**, striving to represent both spatial and temporal variations in urban demand in detail. The models using these spatio-temporal demand profiles are developed for different purposes. For instance, Mikkola and Lund, Fonseca et al. and Best et al. characterise local energy consumption [14,20,28]. Pitt and Kirschen, and Gerbec et al. aim to devise better tariff plans for utilities [29,30]. Andersen et al. have published several papers describing forecasting of future demand [22,23,31]. Davila et al. have developed targeted emission reductions plans for cities [12].

The papers that follow the last approach typically cover one or only a small number of areas and therefore do not provide a systematic understanding of spatio-temporal demand variations. This paper serves this exact purpose by conducting a systematic analysis of spatio-temporal demand variations on different urban scales based on nearly 15000 urban areas in the Netherlands.

2.2. Clustering techniques for urban demand modelling

The analysis in this paper covers a large dataset of thousands of Dutch urban areas: hundreds of municipalities, and thousands of districts and neighbourhoods constituting those municipalities. The scale of the dataset and the goal of the analysis, classification and comparison of "archetype" urban scale spatio-temporal demand profiles, mandates the use of data analysis techniques. *Clustering* is an established approach for this type of problem, it is a collective term for a range of unsupervised algorithms that classify patterns into groups (clusters) [32]. In the context of this research, patterns are spatio-temporal demand profiles on different urban scales, and groups are archetypes of similar urban scale spatio-temporal demand profiles.

Clustering is a technique that has been frequently used for classification of electricity demand profiles. It has been mostly applied to gain insights in large databases of *single* energy users demand profiles for the purpose of improving utilities' understanding of their energy users demand and for subsequent adjustment of billing tariffs. A number of authors [29,30,33,34] have proposed and implemented different clustering algorithms. Gerbec et al. [30] use the fuzzy c-means clustering algorithm to allocate energy users demand profiles to service sector energy users with specific economic activities [30]. Pitt and Kirschen choose a combination of four algorithms: one-pass clustering, binary splitting algorithm, iterative join-two algorithm and exhaustive binary search for database knowledge discovery purposes [29]. Figueiredo et al. apply solely the C5.0 clustering algorithm for the same purposes [33]. Räsänen et al. use k-means clustering for load forecasting [34]. Lamedica et al. have developed a tool based on clustering techniques for the identification of outliers in demand profiles at the high to medium voltage substation level. The tool offers both traditional and hierarchical algorithms [35]. In contrast to the authors above, Yamaguchi et al. apply clustering to *districts*, with Osaka, Japan, as a case study. However, the authors based their clusters on commercial floor space instead of on demand profiles. In their paper, the authors compare clustering based on single buildings and on districts as clustering units [25]. Finally, Chicco provides a comprehensive review of clustering techniques used for electrical load pattern grouping, comparing the most commonly used clustering techniques and cluster validity indicators [17]. The following section follows the general methodology described by Chicco, describing the details of the clustering technique and data used in this paper.

3. Methods

The approach in this paper follows existing procedures for classification and analysis of individual energy users demand profiles [17]. This approach consists of four phases: (1) data gathering and processing, (2) pre-clustering, (3) clustering, and (4) post-clustering. The novel aspect of this paper is the application of this approach to demand profiles of urban areas.

3.1. Data gathering and processing phase

Creating and analysing detailed spatio-temporal demand profiles on urban scales is data intensive. This section describes the datasets used. The Netherlands is used as the geographic area of study with the year 2014 as the reference period. Fig. 1 shows the flow diagram of this phase.

3.1.1. Data gathering

This paper uses a bottom-up approach, and combines datasets from different sources as complete energy demand datasets at urban scales are not (publicly) available. The temporal dimension is based on temporal *demand profiles of buildings*. The spatial dimension is based on *administrative registration data* of households and services in municipalities, districts and neighbourhoods.

3.1.1.1. Temporal dimension data. The temporal dimension is described by hourly demand[1] profiles of buildings. The profiles span the entire year 2014. Data for households and services[2] are obtained from two different sources. Households demand data are obtained from [36], with the average yearly household consumption assumed to be 3500 kWh [37]. Service sector demand data are based on reference building data of the United States Department of Energy (U.S. DOE) [38]. This is, to the best of our knowledge, the only publicly available resource that provides hourly demand profiles for non-household energy users. The use of U.S. service sector demand profiles in the Dutch context is the subject of previous work [21], upon which the present paper builds further.

3.1.1.2. Spatial dimension data. To determine the geographical energy user distribution, local registration data of households and services are used. These data are obtained from Statistics Netherlands [39,40]. The data for the municipality scale are publicly available [40], the datasets for district and neighbourhood scales are protected by privacy laws and are not publicly available [39], but could be accessed by the authors through a research agreement. The data describe the number of households and services in 14698 areas in the Netherlands (11570 neighbourhoods, 2725 districts and 403 municipalities).

The urban scales are defined as follows. Municipalities are the largest scale, and are defined by Dutch law. A municipality is the third level of government in the Netherlands, after the central and provincial government [41]. Districts are subdivisions of municipalities and are typically defined by a single prevailing land use or building type. A district consists of one or multiple neighbourhoods. A neighbourhood is defined either based on socio-economic or historic resemblance, *i.e.* a neighbourhood has a strong socio-economic coherence, or has been developed as a single area. Districts pertain to a single municipality, and neighbourhoods to a single district.[3] Both districts and neighbourhoods are defined by the local municipality authorities. Note that defining factors can therefore differ between municipalities. The size of urban scales in terms of area and annual (modelled) electricity demand is summarised in Table 1.

3.1.2. Data processing

In the data processing phase, temporal and spatial dimension data are adjusted to make the temporal and spatial datasets compatible and usable by the subsequent clustering phase.

Table 1
Summary of urban scale sizes in the Netherlands in terms of land area and annual (modelled) electricity demand.

	Municipality	District	Neighbourhood
Mean area (ha)	8680	1097	255
Median area (ha)	6613	635	64
Maximum area (ha)	46005	24748	12821
Mean demand (GWh/y)	102	15	3.5
Median demand (GWh/y)	57	7.2	1.6
Maximum demand (GWh/y)	1972	428	97

3.1.2.1. Temporal dimension data. The single-year temporal demand profiles of households and services are divided into working days (Monday through Friday) and weekends (Saturday, Sunday and holidays),[4] see upper part of Fig. 1. This conversion is similar to [17,30,31,43] and serves here the purpose of reducing the problem size and thus solution time for the subsequent clustering phase [33,34].

3.1.2.2. Spatial dimension data. The main data processing action for the spatial dimension data is data cleaning, for instance, removing duplicates.

3.1.2.3. Making temporal and spatial datasets compatible. Four issues are solved to make the temporal and spatial datasets compatible:

- For some service sector subsectors, the temporal dimension dataset has a higher granularity than the spatial dimension dataset. For instance, temporal demand profiles are available for "Small Offices", "Medium Offices" and "Large Offices" separately, while the spatial service sector registration data combine all offices in a single category "Office". In such cases, the more detailed temporal demand profiles are combined into a single profile using weighting factors determined in [21] (*e.g.*, demand profiles of "Small Offices", "Medium Offices" and "Large Offices" are combined into a single demand profile for "Offices"). Table A1 in the Appendix provides an overview of the relationships between the 14 subsectors from the temporal dataset, the 11 subsectors from the spatial dataset and how they are combined into 7 subsectors in the joint dataset.
- Only partial data are available for some service sector subdivisions. For the subsectors "Prison", "Sports Facility" and "Industry", spatial data are available, but corresponding temporal demand data are lacking. These subsectors are therefore not considered in the analysis. For the reference building "Hospital", temporal demand data are available, while spatial data are available for the subsector "Healthcare". However metadata that can be used to match U.S. DOE "Hospitals" to the "Healthcare" subsector in the Netherlands with a satisfactory degree of confidence are lacking [44]. Therefore the "Hospital" reference building cannot be used to model "Healthcare", and this subsector is further omitted in the analysis.
- The temporal and spatial datasets refer to different "bases", respectively *buildings* and *registered entities* (households and services). However, buildings and registered entities are not necessarily equal. For instance, a single building can accommodate multiple registered entities, *e.g.*, a single office building can be used by multiple companies, that are thus all registered at the same address. To account for this difference in bases between the two datasets, the temporal demand profiles are scaled. This scaling is carried out using so-

[1] The demand modelled does not include new technologies such as PVs, EVs and HPs.

[2] The industrial sector is left out of scope because industrial energy users require a case-by-case instead of a statistical approach due to their large size.

[3] Although municipalities, districts and neighbourhoods are relatively stable, combinations or divisions occasionally occur [42]. As the reference year in this paper is 2014, the areas as defined in 2014 are used.

[4] Seasonal demand variation is not taken into account for the following reason. The average deviation between seasonal and annual demand profiles for both weekdays and weekends is limited to 10–20%. Splitting the dataset in different seasons decreases the number of datapoints available in each data subset, thus decreasing the power of subsequent statistical tests and increasing the likelihood of finding false positives.

called *scaling factors* obtained through linear regression, as described in detail in [44].

- The average size of buildings differs between the U.S. and the Netherlands [21]. This issue is solved together with the previous issue through scaling of the reference building demand profiles using scaling factors.

The result is a combination of two mutually compatible datasets that describe seven energy users classes: "Households", "Restaurants", "Offices", "Hotels", "Schools", "Shops" and "Warehouses". The temporal dimension is described by a 7 × 2 × 24 dataset: average daily electricity demand of 7 energy user classes for 2 day types (weekday and weekend) and for 24 h. This dataset is further referred to as Π. The spatial dimension is described by a 14698 x 7 dataset: local registration data of the same 7 energy user classes in 14698 areas in the Netherlands (11 570 neighbourhoods, 2725 districts and 403 municipalities). This dataset is further referred to as Γ. The combined, area-specific spatio-temporal demand profiles are constructed as described in the next section, where they are called *feature vectors*, following the nomenclature in clustering literature.

3.2. Pre-clustering phase

In the pre-clustering phase, the datasets are prepared to be used as input for the clustering algorithm. This phase is depicted in Fig. 2. The clustering algorithm classifies so-called *patterns*. Thus, the available data need to be defined in terms of such patterns. For the purpose of clustering, each pattern is typically described by a so-called *feature vector*. The individual scalar components of the feature vector are called *features* [32].

The adapted spatial and temporal datasets are combined into feature vectors that each describe the demand profile of a single area (a neighbourhood, district or municipality) for one of the two day types (weekday or weekend). Let S be the set of the three urban scales {Municipality, District, Neighbourhood} and T the set of the two day types {Weekday, Weekend}. Let A_s denote the set of areas on urban scale s. For each of the six combinations of urban scale and day type (s, t), the 24-h spatio-temporal demand profile of area a_s is described by the feature vector $D_{a_s,t} = \{d^1_{a_s,t}, \ldots, d^h_{a_s,t}, \ldots, d^{24}_{a_s,t}\}$ with $d^h_{a_s,t}$ the individual features. This formulation is an extension of the approach in [33].

The feature vector $D_{a_s,t}$ is calculated from datasets Π and Γ as follows. Let C be the set of the seven modelled energy user classes {Households, Restaurants, Offices, Hotels, Schools, Shops, Warehouses}. Let $D_{c,t}$ be the 24-h demand profile of energy user class c on day type t. This profile is a row vector from dataset Π as shown in Fig. 2. Let n_{c,a_s} be the number of energy users of class c in area a_s on urban scale s, and N_{C,a_s} the vector containing $n_{c,a_s}\ \forall\ c \in C$. The vector N_{C,a_s} is a row vector from dataset Γ, see also Fig. 2. The feature vector $D_{a_s,t}$ is the area profile of urban area a_s on scale s for day type t. This area-specific demand profile equals the sum of the demand profiles of the different energy user classes multiplied by the respective number of energy users in the area in question:

$$D_{a_s,t} = \sum_{c \in C} D_{c,t} \cdot n_{c,a_s} \tag{1}$$

The aim of clustering is classification of spatio-temporal demand profiles in terms of their shape. For this purpose, profiles need to be normalised using a suitable normalising factor. The maximal electricity demand of the spatio-temporal demand profiles $D_{a_s,t}$ is used as the normalising factor (similarly to [17,30,33]), yielding per-unit profiles $\widetilde{D}_{a_s,t}$, which are called "representative load patterns" (RLPs):

$$\widetilde{D}_{a_s,t} = \frac{D_{a_s,t}}{\max(D_{a_s,t})} \tag{2}$$

In this paper, RLPs are used as clustering feature vectors.

3.3. Clustering phase

Clustering is an established unsupervised data analysis technique used for classification of patterns (in this case, spatio-temporal demand profiles) into groups called "clusters" [32]. The input and output datasets of the clustering phase are shown in Fig. 3. Several clustering algorithms exist, this paper uses the k-means algorithm. This choice is based on the comparison of clustering algorithms by Chicco [17], which shows that k-means clustering yields clusters populated in a relatively uniform way, contrary to other methods such as hierarchical clustering, that tend to isolate outliers and group the remaining patterns in a single large group. The choice for k-means clustering is in line with the purpose of this paper to find areas with similar demand profiles and within the same order of magnitude. Moreover, k-means clustering is often computationally faster than other algorithms [17]. Thus, k-means clustering is chosen for the type of clusters it yields and its computational speed. The k-means clustering algorithm implemented here is described as follows [32]:

Algorithm 1. k-means clustering

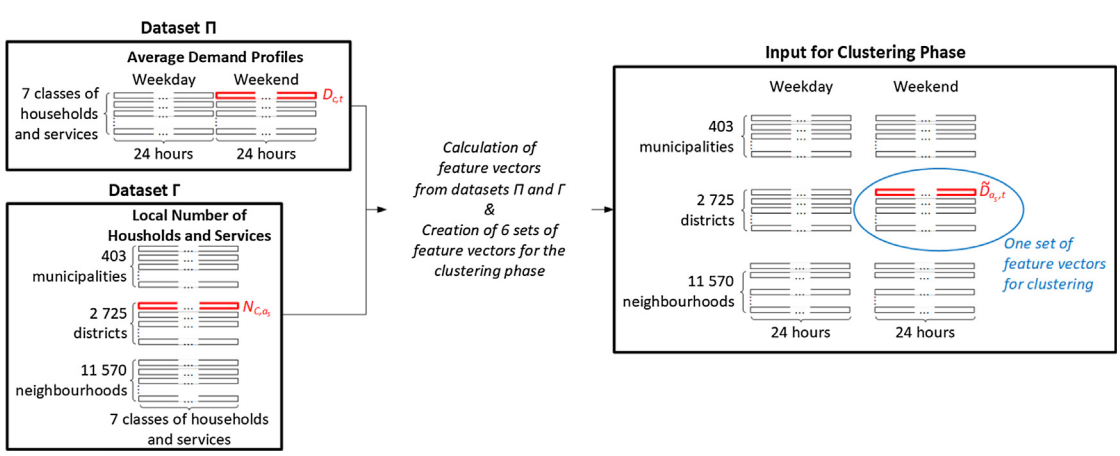

Fig. 2. Flow diagram of the pre-clustering phase. The diagram shows how datasets Π and Γ obtained in the data gathering and processing phase are used to produce an input dataset for the clustering phase. This input dataset contains six sets of so-called *feature vectors*, *i.e.* vectors describing the *features* or *patterns* that are used by a clustering algorithm to form clusters. The feature vectors are called $\widetilde{D}_{a_s,t}$ and are calculated through the multiplication of vectors $D_{c,t}$ from dataset Π by elements n_{c,a_s} from the vector N_{C,a_s} from dataset Γ, summation over all consumer classes $c \in C$ (see Eq. (1)) and normalisation of the resulting area-level profile (see Eq. (2)). The vectors $D_{c,t}$ describe the demand profile of consumer class c on day type t. The integers n_{c,a_s} describe the local number of consumers of class c in the area of interest a_s, and are elements of the vector N_{C,a_s} that contains the numbers of all consumers classes $c \in C$ for the area a_s.

Algorithm 1: k-means clustering

Choose k **cluster centres**. The centres are either k randomly chosen feature vectors or k randomly defined features inside the feature vectors set;
while *feature vectors are reassigned to new cluster centers, or squared error is larger than a pre-set value* **do**
  Assign each feature vector to the closest cluster centre;
  Recompute the cluster centres using the new cluster memberships;
end

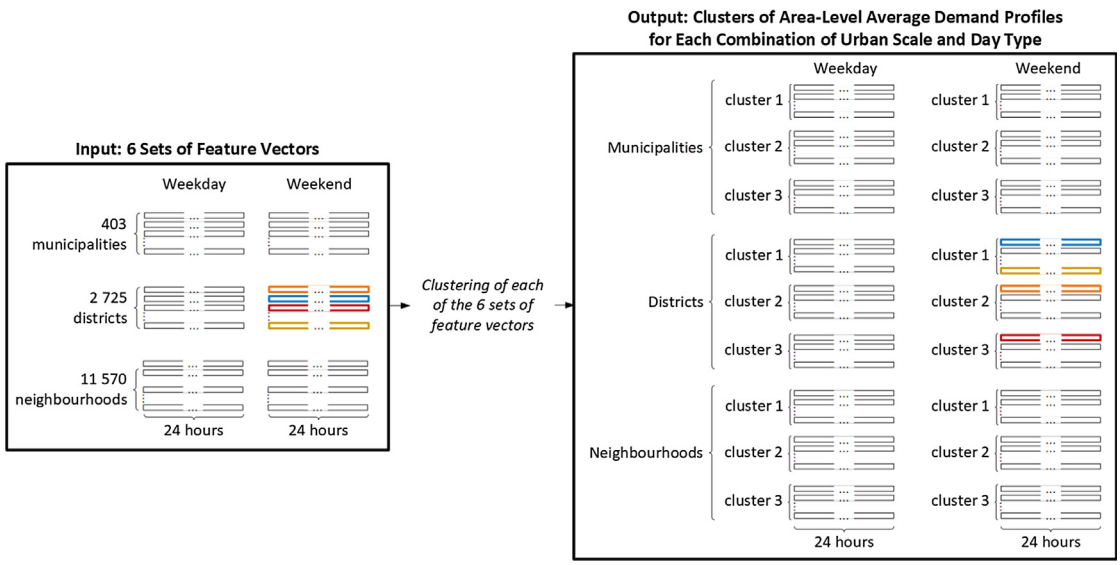

Fig. 3. Flow diagram of the clustering phase. Clustering is carried out six times, once for each combination of day type (weekday and weekend) and urban area (municipalities, districts and neighbourhoods). For each combination, the result is three clusters of similar area-level average demand profiles for one of the two day types. These clusters are analysed in the post-clustering phase, and the analysis results are shown in Figs. 5–11.

The k-means clustering algorithm requires an a priori choice of the number of clusters k. To determine the optimal number of clusters, a cluster validity index (CVI) is calculated for a range of cluster numbers. The CVI used in this paper is the Davies-Bouldin index (*DBI*) [45], chosen because it provides a good balance between cluster compactness and separation [34,46]. The Davies-Bouldin index for k cluster centres is defined as follows:

$$DBI(k) = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j}\left\{\frac{l_i + l_j}{g_{i,j}}\right\}, \quad \text{for } i, j = \{1, \ldots, k\} \tag{3}$$

where l_i is the within cluster distance of cluster i, and $g_{i,j}$ the between cluster distance for clusters i and j. The distances are defined as:

$$l_i = \frac{1}{|i|}\sum_{q_i \in i} \|q_i - m_i\| \tag{4}$$

$$g_{i,j} = \|m_i - m_j\| \tag{5}$$

where q_i is a feature vector[5] in cluster i, m_i is the centre of the cluster i, and $|i|$ is the number of elements in cluster i. The notation $\|q_i - m_i\|$ represents the Euclidean distance between vectors q_i and m_i [32,45].

The Davies-Bouldin indices are calculated for all combinations of urban scale and day type. The result for day type weekday is shown in Fig. 4. The optimal number of clusters is 3 for all urban scales. This is also the optimal number of clusters for the weekend day type for neighbourhoods and districts. The optimal number of clusters for municipalities on weekends is 2, however, for consistency and comparison purposes, the number of clusters is chosen to be 3 for all combinations of urban scale and day type.

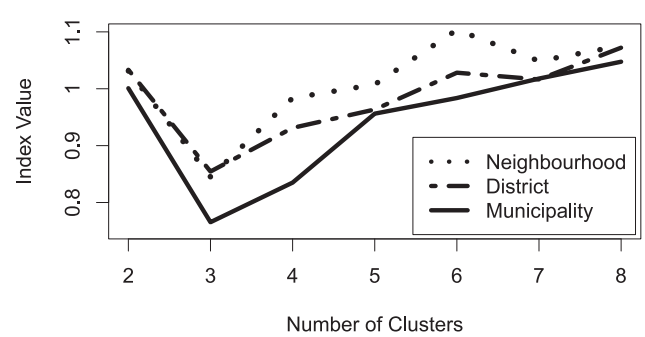

Fig. 4. Davies-Bouldin indices for the three urban scales (neighbourhood, district and municipality) for day type weekday. The optimal number of clusters is determined by the lowest index value, which is 3 for all the urban scales.

3.4. Post-clustering phase

The post-clustering phase covers the analysis of the spatio-temporal demand profile classification obtained through clustering. Each urban

[5] Note that q_i in Eq. (4) and $\widetilde{D}_{a_s,t}$ in Eq. (2) refer to the same feature vectors, but emphasise different aspects. The notation q_i indicates the cluster membership and thus the result of clustering, while the notation $\widetilde{D}_{a_s,t}$ expresses the feature vector construction from the available data, and thus the input for clustering.

scale-day type combination is first characterised in terms of (1) its daily demand profile, and (2) the annual demand of the seven energy user classes analysed. Second, overall and pairwise statistical comparisons of the demand profiles and annual energy user demand compositions are carried out. As the distributions are non-normal, parametric statistical tests are used. The Kruskal-Wallis test is used for the overall comparison, followed by the Mann-Wilcoxon-Whitney test for the pairwise comparison [47]. To avoid increasing the likelihood of Type I errors (*i.e.* finding false positives), all tests are subject to a Bonferroni correction [47]. The family-wise error rate is kept at 5% by correcting for 288 comparisons for demand profiles and for 84 comparisons for energy user annual demand composition.

3.5. Logistic regression

The clustering analysis as described above is based on detailed local households and services registration data. Such data are often not available to other researchers (in this study the data could be accessed by the authors through a research agreement with Statistics Netherlands). To facilitate the application of the results and insights obtained through clustering by other researchers, a logistic regression model has been developed.

In essence, this logistic regression model determines the probabilities that an urban area of interest belongs to each of the three clusters, based solely on relative annual demand data of different energy user classes.[6] Relative annual demand data reflect local energy user composition. This paper shows that this energy user composition differs significantly between each two pairs of clusters, for each combination of urban scale and day type, see further in Section 4.1.2.

As for the clustering analysis, the logistic regression model is built for each of the six combinations of day type (weekday and weekend) and urban scale (municipality, district and neighbourhood). For each combination, 70% of the datapoints are used for logistic regression model calibration, and 30% for its validation. The model correctly classifies over 98% of the areas from the validation dataset. Formally, the logistic regression model determines the probability **P** that an area of interest α_s on urban scale $s \in$ {Municipality, District, Neighbourhood} on day type $t \in$ {Weekday, Weekend} belongs to each of clusters {Residential, Business, Mixed}:

$$\mathbf{P}(\alpha_s \in \text{Residendial}) = 1\cdot(1 + \exp v_{s,t} + \exp w_{s,t})^{-1} \tag{6}$$

$$\mathbf{P}(\alpha_s \in \text{Business}) = \exp v_{s,t}\cdot(1 + \exp v_{s,t} + \exp w_{s,t})^{-1} \tag{7}$$

$$\mathbf{P}(\alpha_s \in \text{Mixed}) = \exp w_{s,t}\cdot(1 + \exp v_{s,t} + \exp w_{s,t})^{-1} \tag{8}$$

where $v_{s,t}$ and $w_{s,t}$ are given by:

$$v_{s,t} = \log\left(\frac{\mathbf{P}(\alpha_s \in \text{Business})}{\mathbf{P}(\alpha_s \in \text{Residential})}\right) = \zeta_{0,s,t} + \sum_c \zeta_{c,s,t}\cdot x_{c,\alpha_s} \tag{9}$$

$$w_{s,t} = \log\left(\frac{\mathbf{P}(\alpha_s \in \text{Mixed})}{\mathbf{P}(\alpha_s \in \text{Residential})}\right) = \eta_{0,s,t} + \sum_c \eta_{c,s,t}\cdot x_{c,\alpha_s} \tag{10}$$

with $c\in$ {Households, Restaurants, Offices, Hotels, Schools, Shops, Warehouses}; $\zeta_{0,s,t}$, $\zeta_{c,s,t}$, $\eta_{0,s,t}$, and $\eta_{c,s,t}$ regression coefficients calculated from Dutch neighbourhoods, districts and municipalities data; and x_{c,α_s} the relative annual energy user demand for energy user class c in the area of interest α_s on scale $s \in$ {Municipality, District, Neighbourhood}, to be provided by the user.

The logistic regression model is provided as a spreadsheet tool in Appendix B (online). The model can be used to determine the type of area (residential, business or mixed) and its average weekday and weekend demand profile for an area of interest based solely on the relative annual demand of the different energy user classes in that area.

[6] Annual demand data are more frequently publicly available than detailed local household and services registration data.

4. Results

This section describes the results obtained through clustering of spatio-temporal demand profiles at three urban scales (neighbourhoods, districts and municipalities) in the Netherlands. Clustering is carried out separately for weekdays and weekends. Overall, the results show that three types of clusters can be distinguished for all three urban scales, to which this paper refers as "residential", "business" and "mixed" clusters. These cluster names are based on the most prevalent energy user classes in each of the clusters: households in the residential cluster, offices in the business cluster, and mixed energy user classes in the mixed cluster. The following paragraphs describe the results in more detail in terms of demand profiles, energy user composition expressed as relative annual demand, and the relative importance of clusters at the different urban scales.

4.1. Cluster demand profiles and composition

Figs. 5–7 show the demand profiles and energy user composition of different clusters on three urban scales for weekdays. Weekend results are shown only for neighbourhoods in Fig. 8, as the trends for districts and municipalities are similar (see further). The following paragraphs first describe the results within one urban scale, then provide the results of the statistic analysis, and finally compare the three urban scales and the two day types.

4.1.1. Single urban scale profile and composition

Within one urban scale, three clusters are distinguished: residential, business and mixed. Both on weekdays (Figs. 5–7) and on weekends (Fig. 8), the **residential cluster** (upper row on all Figures) has a demand profile similar to that of an average household, *i.e.* with a peak in the evening hours. The residential cluster contains areas in which the largest part of the annual electricity demand is consumed by households (*e.g.*, a median of 84% at the neighbourhood scale on weekdays, Fig. 5, right upper panel).

The **business cluster** has a demand profile with a plateau between 9:00 and 16:00 on weekdays, with decreasing demand in the evening hours. On weekends, demand starts decreasing earlier, from 14:00. The business cluster contains areas in which the largest part of the annual electricity demand is consumed by offices (*e.g.*, a median of 39% at the neighbourhood scale on weekdays, Fig. 5, right middle panel).

Finally, the **mixed cluster** has a demand profile with a double-peak between 9:00 and 20:00. The peaks are more pronounced on weekends than on weekdays. The mixed cluster contains areas in which the largest part of the annual electricity demand is consumed by households (*e.g.*, a median of 53% at the neighbourhood scale on weekdays, see Fig. 5, right bottom panel). However, the relative share of the annual demand of households is considerably smaller than in the residential cluster, with a higher share of demand consumed by all other energy user classes, in particular restaurants, offices and warehouses.

4.1.2. Statistical comparison

Statistical comparison of the **demand profiles** at each urban scale shows that the differences are significant between nearly all pairs of clusters, both on weekdays and on weekends. At the **neighbourhood scale**, on weekdays, p-values of all pair-wise profile comparison tests are smaller than the family-wise cut-off p-value of 5.21×10^{-4}. At the **district scale** on weekdays, all pairs of clusters are statistically different (p-value less than 5.21×10^{-4}), except the residential and business clusters at 6:00, the business and mixed clusters at 11:00 and 14:00, and the residential and mixed clusters at 18:00. At the **municipality scale** on weekdays, clusters are pairwise statistically different (p-value less than 5.21×10^{-4}), with a few exceptions, namely mixed and

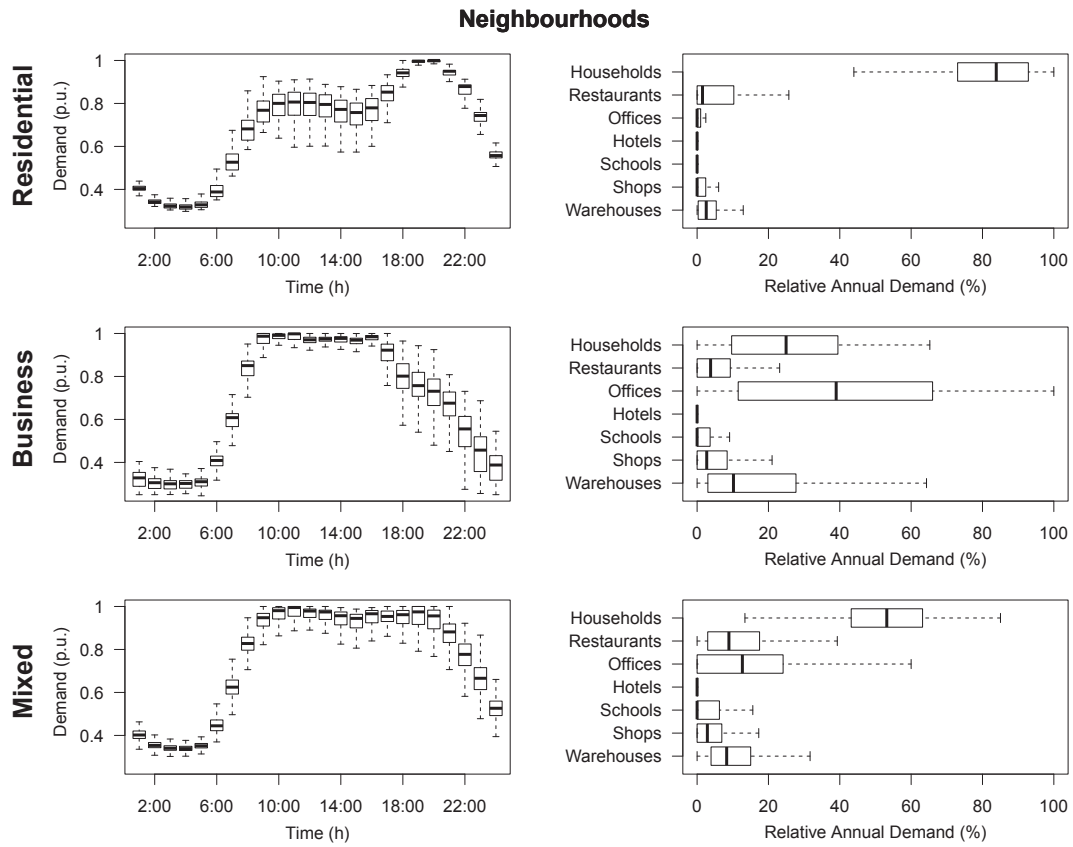

Fig. 5. Characterisation of clusters at the neighbourhood scale. Left panels show the per unit (p.u.) demand profiles, right panels the energy user composition in terms of relative annual demand. The three rows represent the three clusters. Clusters are formed based on similarity of weekday spatio-temporal demand profiles. Both profiles and energy user compositions are shown as boxplots, with the middle line representing the median, the boxplots edges the 25% and 75% percentiles, and the whiskers the minima and maxima. Due to privacy rules of Statistics Netherlands [39], outliers are not shown.

residential clusters at 2:00 and 18:00, and business and mixed clusters at 8:00 and 13:00. On weekends, pairwise statistical difference in profiles is similarly found for nearly all hours and all cluster pairs.

Pairwise comparison of **energy user composition** shows that most clusters are pairwise statistically different on all urban scales and for both day types. At the **neighbourhood scale** on weekdays, statistically significant distinction can be made between all cluster pairs for all energy user classes in terms of their annual demand (family-wise cut-off p-value of 1.78×10^{-3}). On weekends, annual energy demand by different user types is significantly different, except for that of hotels in the residential and business clusters. At the **district scale**, both on weekdays and on weekends, all clusters are pairwise statistically different (family-wise cut-off p-value of 1.78×10^{-3}), except for the annual demand of hotels in residential and business clusters, and the annual demand of shops in business and mixed clusters. At the **municipality scale** on weekdays and weekends, most clusters are statistically different (family-wise cut-off p-value of 1.78×10^{-3}), with a few exceptions: on weekdays, the annual demand of restaurants and hotels in residential and business clusters, that of schools and shops in business and mixed clusters, and that of hotels in the residential and mixed clusters is not statistically different; and on weekends, no statistical distinction can be made in annual demand of restaurants in residential and business clusters, that of hotels in residential and mixed clusters, and that of schools, shops and warehouses in mixed and business clusters.

In summary, despite the conservative Bonferroni correction, the statistical analysis shows that, with a few exceptions, significant differences exist in demand profiles and in energy user composition between all cluster pairs, on all urban scales, both for weekends and weekdays and for most hours and energy user classes.

4.1.3. Comparison across urban scales and day types

Both the profiles and the energy user composition of same-type clusters across urban scales are similar. However, the variation in both the profiles and the energy user composition decreases with increasing urban scale. The three clusters – residential, business, and mixed – as described above exist for all three urban scales. There is less variation in both the profiles and the energy user composition at the municipality scale (Fig. 7) than at the district scale (Fig. 6), and at the neighbourhood scale (Fig. 5). This is the case both on weekdays (Figs. 5–7) and on weekends. On weekends, only the results for neighbourhoods are shown (Fig. 8), the results for higher urban scales are similar in terms of demand profile and energy user composition, but with smaller variations (and are therefore not shown).

4.2. Relative cluster importance

Fig. 9 shows the relative importance of the three clusters (residential, business and mixed) for the three urban scales, both on weekdays and on weekends. Two metrics are used: relative share of areas, and relative share of annual demand. On weekdays (upper row), the residential cluster contains the least number of areas (25% of the neighbourhoods to 8.6% of the municipalities), covering an even smaller part of the annual demand (10% at the neighbourhood scale to 3% at the municipality scale). For all three urban scales, the mixed type cluster contains the most areas (46–49%), while the business type cluster covers the largest part of the annual demand (46–68%). On weekends (lower row), more areas are classified into the residential-type cluster, and less into the business-type. Approximately the same number of areas remains classified as mixed type, however these areas cover more demand than during weekdays.

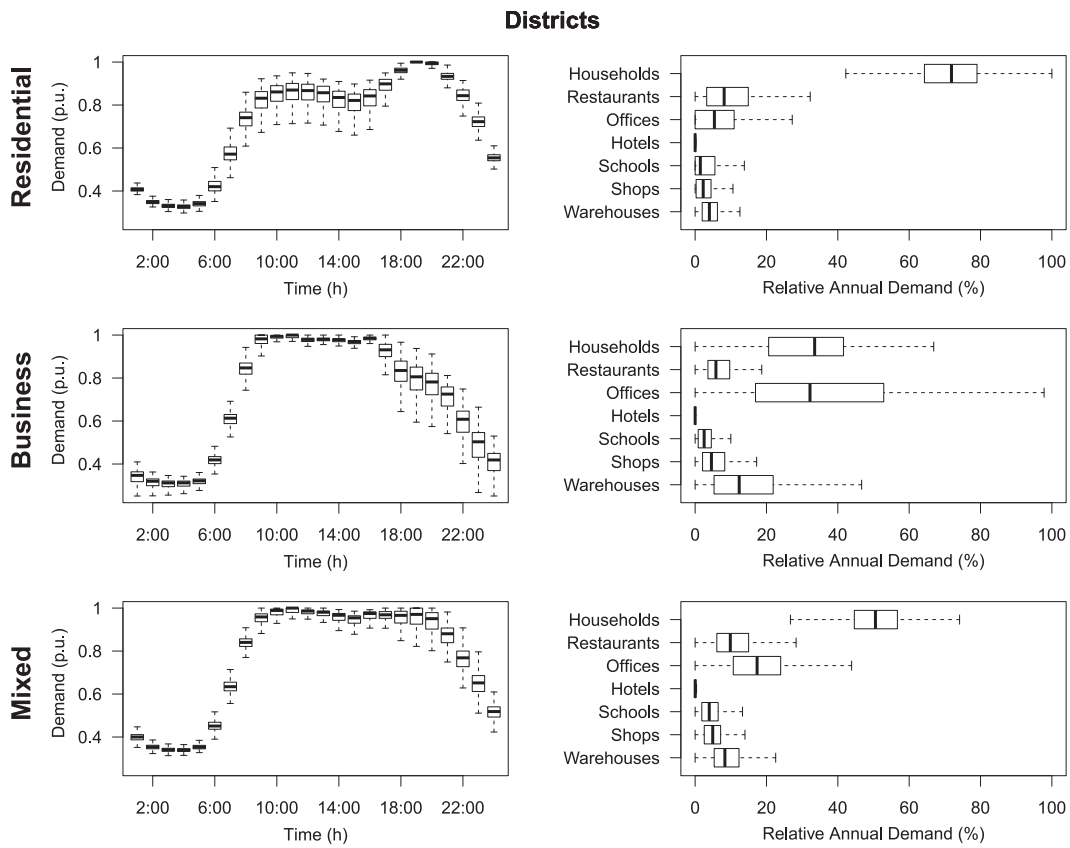

Fig. 6. Characterisation of clusters at the district scale. Left panels show the per unit (p.u.) demand profiles, right panels the energy user composition in terms of relative annual demand. The three rows represent the three clusters. Clusters are formed based on similarity of weekday spatio-temporal demand profiles. Both profiles and energy user compositions are shown as boxplots, with the middle line representing the median, the boxplots edges the 25% and 75% percentiles, and the whiskers the minima and maxima. Due to privacy rules of Statistics Netherlands [39], outliers are not shown.

4.3. Interaction between urban scales

Fig. 10 shows the distribution of lower-scale clusters across higher-scale clusters on weekdays. The distribution is similar on weekends, accounting for the higher share of residential-type clusters, and the lower share of business-type clusters. The interpretation of Fig. 10 is described for the left panel (neighbourhoods per district). The panel shows how the 11 570 *neighbourhoods* are distributed across *districts*.

The residential *neighbourhoods*, 2876 in total, or 25% of all neighbourhoods (see also left panel in Fig. 9) are classified across all three district types: 1181 (10%) are classified in residential districts, 329 (3%) in business districts, and 1366 (12%) in mixed districts.

The residential *districts* consist of 1875 neighbourhoods (16% of all neighbourhoods), the mixed districts consist of 6038 neighbourhoods (52% of all neighbourhoods). Thus, although the absolute number of residential neighbourhoods in both residential and mixed districts is approximately the same, 62% of residential districts consists of residential neighbourhoods, while only 23% of mixed districts consists of residential neighbourhoods. The distribution of business and mixed neighbourhoods across respectively business and mixed districts is similar, 60% of business districts consists of business neighbourhoods, and 59% of mixed districts consists of mixed neighbourhoods.

These results show a correlation between clusters at lower-level and at higher-level urban scales, although clustering is carried out independently at each scale. Note that on the middle and right panels of Fig. 10 there are more business municipalities than mixed municipalities, while on the left panel in Fig. 9 there are more mixed-type municipalities than business-type. This is seemingly contradictory, but can be explained as follows: Fig. 10 shows the relative number of *neighbourhoods* (middle panel) and *districts* (right panel) classified across municipalities, while the right bar on the left panel in Fig. 9 shows the relative number of *municipalities* themselves. Thus, although fewer municipalities are classified as business-type than as mixed-type, the business municipalities contain a higher number of districts and neighbourhoods than the mixed municipalities.

4.4. Interaction between day types

Fig. 11 shows how clusters formed based on weekend profiles relate to clusters formed based on weekday profiles at each urban scale. In general, Fig. 11 shows that the areas that are classified as residential on weekends, are also classified as such on weekdays. However, on weekends, more areas are classified as residential. Primarily areas that are classified as mixed on weekdays are reclassified as residential on weekends. A similar pattern can be distinguished for business areas. Areas that are classified as business on weekends, are also classified as such on weekdays. However, on weekdays, more areas are classified as business, these areas are classified predominantly as mixed on weekends. In both reclassification cases, mixed to residential and business to mixed, this reclassification occurs due to smaller demand of non-household energy users on weekends than on weekdays.

5. Discussion

The results presented in this paper provide insights in the demand profiles and energy user composition on three urban scales. These insights can improve existing and future energy system models, used to assess and support the transition to renewable generation and electrification of transportation and heating. Most existing energy models assume that the local demand in the area of study is solely residential.

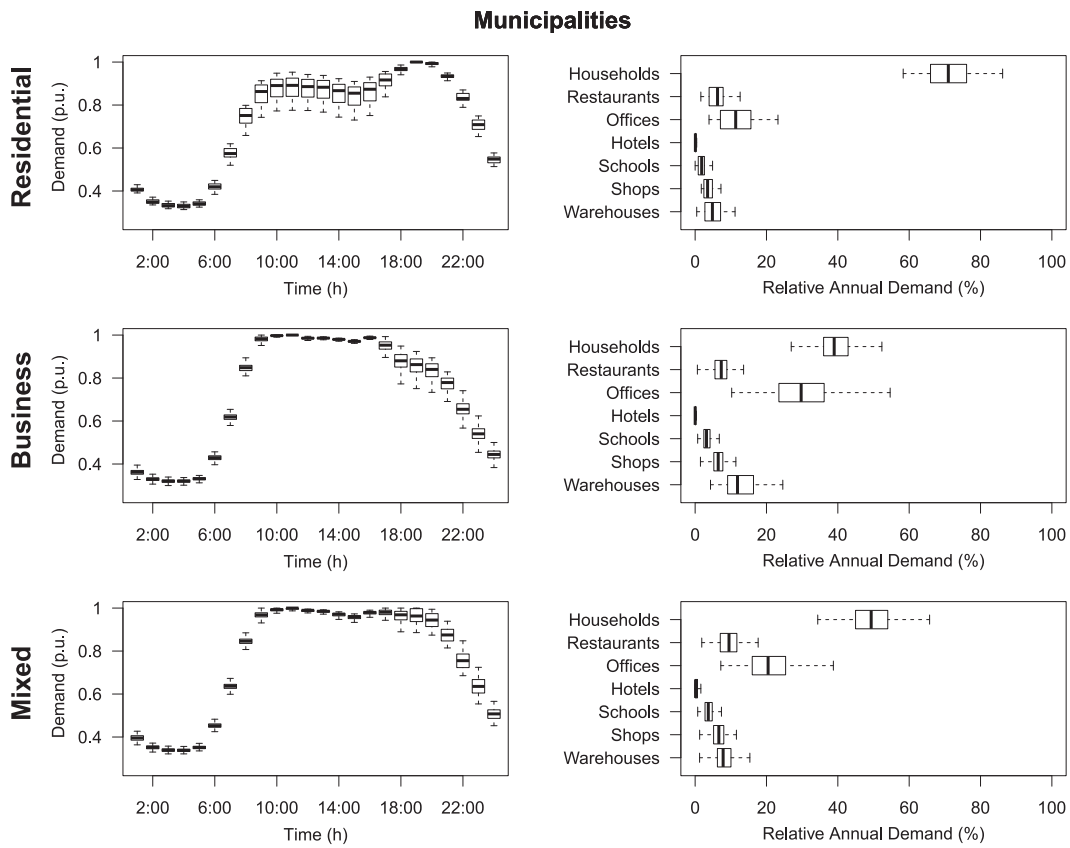

Fig. 7. Characterisation of clusters at the municipality scale. Left panels show the per unit (p.u.) demand profiles, right panels the energy user composition in terms of relative annual demand. The three rows represent the three clusters. Clusters are formed based on similarity of weekday spatio-temporal demand profiles. Both profiles and energy user compositions are shown as boxplots, with the middle line representing the median, the boxplots edges the 25% and 75% percentiles, and the whiskers the minima and maxima. Due to privacy rules of Statistics Netherlands [39], outliers are not shown.

The results in this paper show that three types of areas can be distinguished: residential, business, and mixed. Statistic analysis demonstrates that these area types are significantly different, both in terms of their daily demand profiles and their energy users composition. Moreover, this paper shows that the residential-type demand assumed by many existing energy models is representative only of a minority of areas, and accounts only for a small share of the total urban demand. The following paragraphs validate the approach used, and discuss the importance of the obtained results for urban energy systems modelling.

5.1. Approach validation

This paper uses a bottom-up approach to model urban scale spatio-temporal demand profiles. In urban energy systems literature, this approach is preferred over the alternative top-down approach as it yields more detailed demand profiles [8,19,20]. A drawback of this bottom-up approach is the need for a large number of data sources. In particular, as detailed spatial-temporal energy data are not publicly available, data sources from other fields need to be combined [14,25–27]. The best validation for this bottom-up approach is arguably the comparison of the resulting demand profiles with measured profiles of statistically representative urban areas, along with metadata of these areas, such as the local energy user composition. However, such data are currently not publicly available. This is the very issue highlighted by this paper. The validation therefore has to rely on the comparison of the results with available data, and prior studies. The validation consists of two parts: (1) comparison of the total annual modelled demand to published annual demand data, and (2) comparison of the results to those published by other authors.

5.1.1. Annual demand validation

The obtained results are compared with the total annual electricity demand of Dutch households and service sector energy users. Such annual country-level demand data are publicly available, but do not suffice to generate detailed spatio-temporal demand profiles. The Netherlands Environmental Assessment Agency attributes 22.7 TWh of Dutch electricity consumption to households, and 43.8 TWh to the combination of the service sector, waste and wastewater treatment, and agriculture and fisheries. The service sector alone consumes 77% of this value [37], *i.e.* 33.6 TWh. Statistics Netherlands reports service sector consumption of 30.6 TWh.[7] Thus, the combined consumption of households and the service sector lies between 53 and 56 TWh per year. This paper models 41 TWh of electricity demand, of which 19 TWh consumed by households, and 22 TWh by the service sector. This means that the modelled demand covers 73–77% of the total demand, 84% of household demand, and 65–71% of the service sector demand. This validation indicates that the demand modelled in this paper is in line with the measured Dutch household and service sector demand data. The missing remainder includes unaccounted for subsectors, in particular in the service sector (*e.g.*, healthcare, leisure), inaccuracies in demand profiles, and in data conversions.

5.1.2. Literature-based validation

Several studies [14,22–25] are used to validate four subsequent parts of this paper: (1) data combination, (2) bottom-up demand modelling based on linear regression, (3) clustering of areas instead of individual energy users, and (4) obtained results.

[7] The discrepancies in published data likely arise from the lack of unified definitions, an issue also raised by other researchers [48–50].

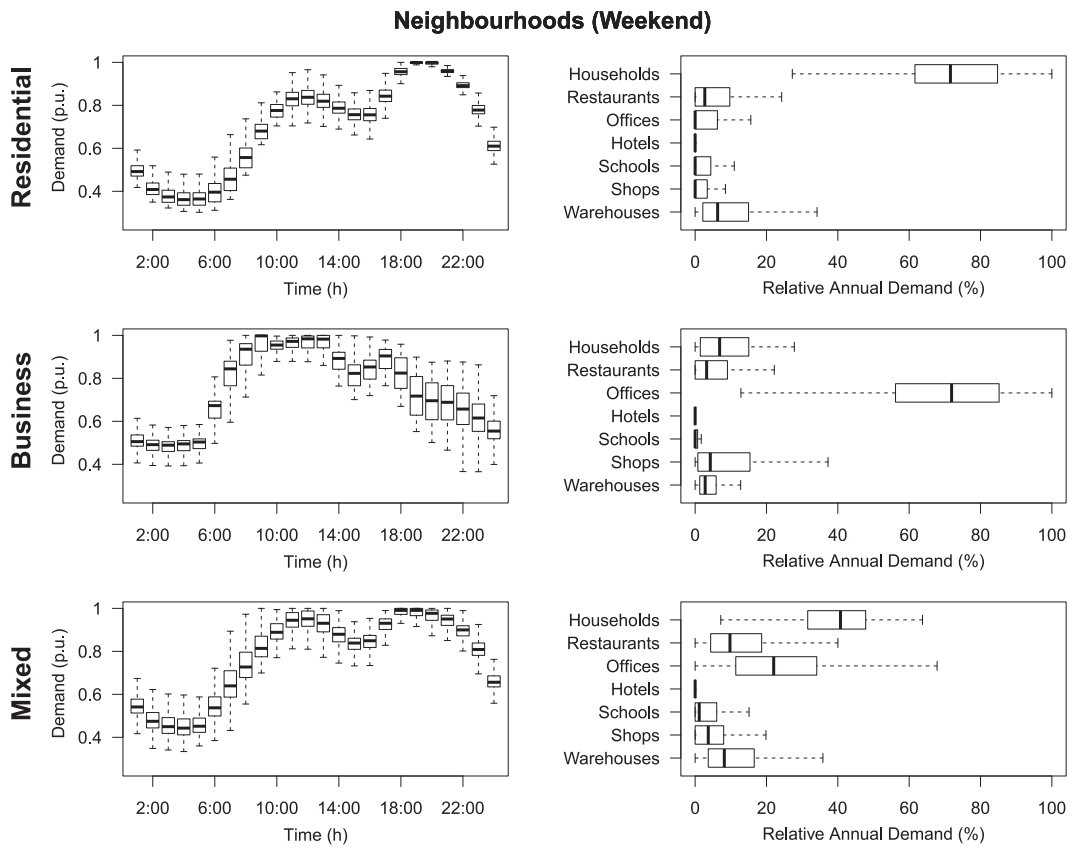

Fig. 8. Characterisation of clusters at the neighbourhood scale for weekend days. Left panels show the per unit (p.u.) demand profiles, right panels the energy user composition in terms of relative annual demand. The three rows represent the three clusters. Clusters are formed based on weekend spatio-temporal demand profiles. Both profiles and energy user composition are shown as boxplots, with the middle line representing the median, the boxplots edges the 25% and 75% percentiles, and the whiskers the minima and maxima. Due to privacy rules of Statistics Netherlands [39], outliers are not shown.

5.1.2.1. Data combination. The data used in this paper are a combination of individual building demand profiles for the temporal dimension, and administrative registration data for the spatial dimension. A similar data combination has been used by Brownsword et al. [24].

5.1.2.2. Bottom-up demand modelling. The bottom-up demand modelling approach in this paper is based on the premise that the

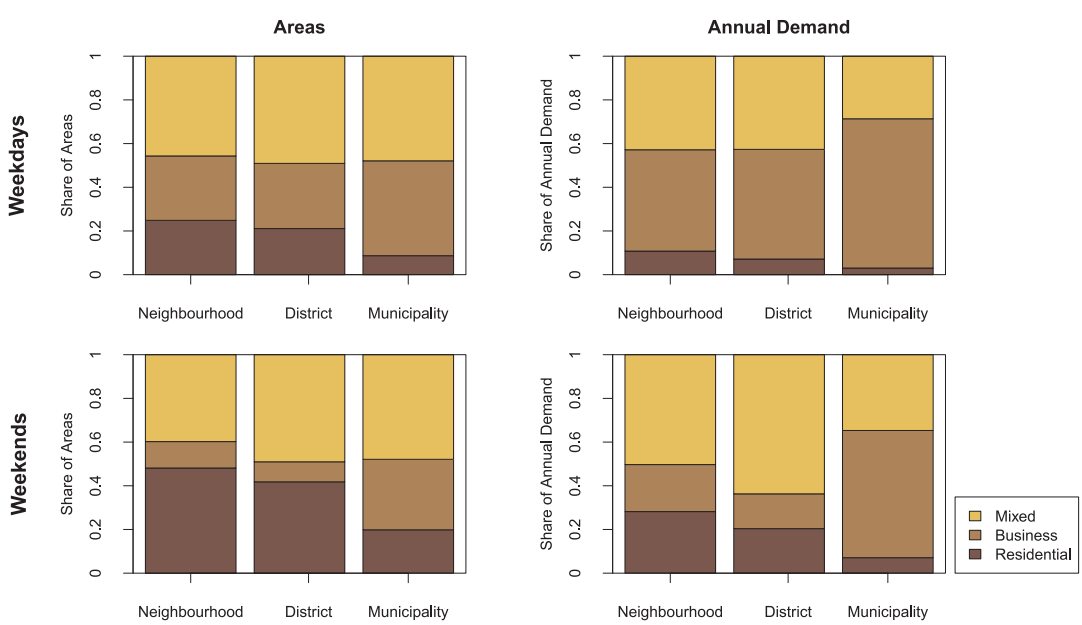

Fig. 9. Relative importance of clusters at three urban scales in terms of the number of areas (left panels) and annual demand (right panels) covered by each cluster. Top row shows clusters based on weekday profiles, bottom row shows those based on weekend profiles. The absolute number of areas is 11 570 at the neighbourhood scale, 2725 at the district scale and 403 at the municipality scale. The total annual demand is 41 TWh/year at all scales.

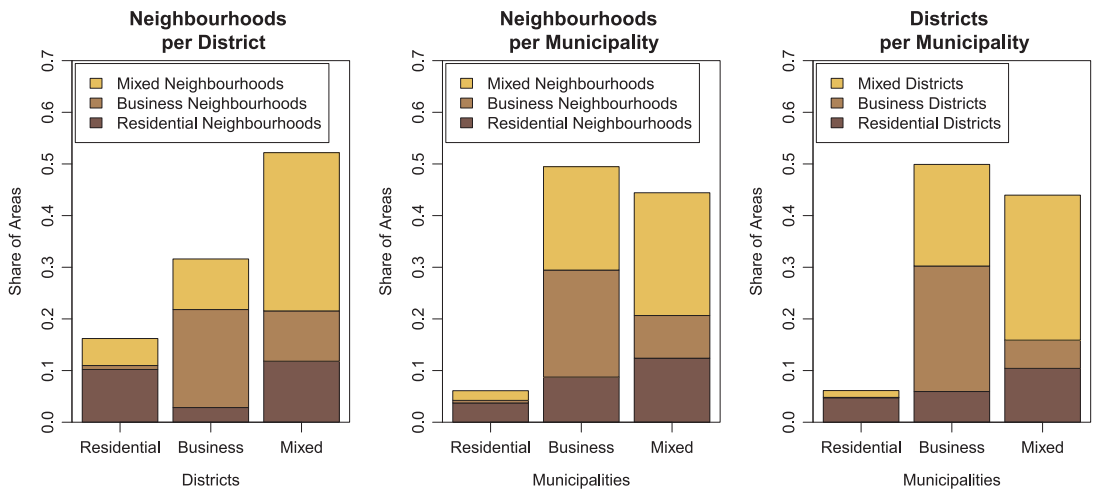

Fig. 10. Distribution of lower-scale clusters across higher-scale clusters on weekdays. In each panel, the sum of all neighbourhoods or districts is 1, the sum of same-type areas (*e.g.*, residential-type) equals the share of that area type in Fig. 9. For instance, the left panel shows the distribution of the 11570 neighbourhoods across districts: 2876 of these neighbourhoods are residential (25%, see also left panel in Fig. 9). Of these residential neighbourhoods, 1181 are classified in residential districts (10%), 329 in business districts (3%), and 1366 in mixed districts (12%). The distribution of neighbourhoods and districts on the other panels should be interpreted similarly.

demand profile of an area is the sum of the demand profiles of the energy users in that area. The same premise is assumed in the work of Andersen et al. [22,23]. These authors determine the local energy user composition based on a combination of transformer-level demand profiles, reference Danish demand profiles, and Nord Pool market data [22,23]. They similarly choose for linear regression as the methodological approach, and use it to estimate the *weight* of the demand of an *energy user class* in a given area [22,23]. Note that this is a similar, but data-wise reversed approach to the one developed in this paper, where the weight of the energy user class is known from annual energy demand data, and linear regression is used to appropriately *scale* reference building *demand profiles*.

5.1.2.3. Area-scale clustering. Clustering is used in this paper to classify areas instead of individual energy users, the latter is more common in literature (*e.g.*, [29,30,33,34]). Yamaguchi et al. [25] also describe clustering of areas, in particular of districts in Osaka, Japan. The authors use floor space of commercial buildings as clustering features. Their analysis results in six clusters: residential, mixed-use commercial, concentrated commercial, urban core, low-rise office and high-rise office. The clustering approach in this paper is based on 24-h demand profiles. The difference in clustering features explains the lower number of clusters in this paper as compared to the results of Yamaguchi et al. [25]. Low-rise and high-rise office buildings are likely to have similar, business-type shapes of electricity demand. Similarly, mixed-use commercial, concentrated commercial and urban core areas are likely to correspond to mixed-type demand profile areas. The work of Yamaguchi et al. [25] shows that clustering is a valid methodology to classify urban areas.

5.1.2.4. Results: three cluster types. The results in this paper show the existence of three clusters on all urban scales, each cluster with a distinct demand profile. Similar results are found by Mikkola and Lund, who have developed a spatio-temporal energy demand model and applied it to 136 neighbourhoods in the city of Helsinki, Finland [14]. Based on their results, the authors distinguish three neighbourhoods

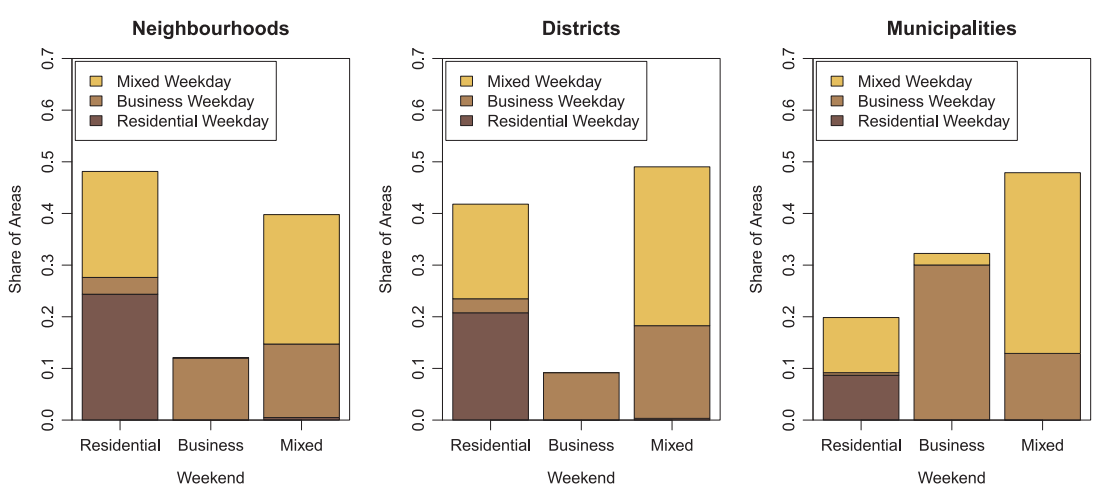

Fig. 11. Distribution of weekend clusters across weekday clusters on different urban scales. In each panel, the sum of all areas is 1. This sum represents 11 570 neighbourhoods (left panel), 2725 districts (middle panel), and 403 municipalities (right panel).

with the same classification as in this paper, and provide examples of each: residential area (Puistola), office buildings area (Kluuvi), and mixed area (Punavuori). The authors also discuss the shape of the demand profiles in each area: the profile peaks during the morning and evening for Puistola, during the day for Kluuvi, and Punavuori is a mix of the two types [14]. Both the classification, and the demand profile shapes are similar to the results found in this paper. Although the results of Mikkola and Lund cover a hundredfold smaller number of areas, they thus validate the findings of this paper.

5.2. Implications for urban energy system models

The primary purpose of the classification of areas presented in this paper is to increase understanding of the heterogeneity of urban scale demand profiles. The results emphasise the importance of using local demand profiles in individual areas simulated in urban energy system models.

5.2.1. Importance of demand profile types

Many studies (*e.g.*, [4–6]) assume residential-type demand when assessing the impact of EVs and PVs on local energy systems. The results in this paper show that this assumption is incorrect for the majority of real urban areas. The results show that the majority of areas has a business-type or a mixed-type profile, *i.e.* a profile with demand having a plateau during the day, or a double-peak during the day and the evening. Such profiles can be expected to interact differently than residential-type profile (with the demand peaking during the evening) with solar PVs and with EVs. Solar PV generation peaks during the day. Robinson et al. showed that EV charging peaks during the morning at workplaces, during the day in public charging points and during the evening at home [51]. Thus, the mismatches between PV generation and local demand can be expected to be smaller in business and mixed areas than in residential areas. The effect of EV charging depends on both the type of EV charging points and the area they are located in, and thus require an assessment tailored to the area under study.

5.2.2. Importance of scale and day type

Two main conclusions can be drawn from the comparisons of urban scales and day types. First, the scale at which the impact of a new technology is assessed is important. As shown in Fig. 10, lower-scale demand types overlap with higher-scale demand types in only approximately half of the cases (55% on average). Thus, although data from a higher-scale area (such as a municipality) might be more readily available, they can be expected to correctly predict the demand profile of lower-scale areas only in approximately half of the cases. Appropriate scale data should therefore be used in energy system models.

Second, for individual areas, both weekday and weekend profiles should be taken into account when assessing the local impact of new technologies. In this paper, same-scale areas are classified independently for weekdays and weekends. Results show that most areas, *i.e.* 61% of the neighbourhoods and districts and 74% of municipalities are classified in the same clusters both on weekdays and on weekends. The remainder is reclassified to a cluster with a higher importance of household energy users due to decreased business activity on weekends (Fig. 11). In reality, the difference in demand profiles of weekdays and weekends in a specific areas should be taken into account when assessing the local impact of new technologies.

5.2.3. Improving models despite lacking data

Determining the local profile of a particular area requires local data, which at present are often lacking. This issue has been raised in literature (*e.g.*, [14,18,27,28]), and remains largely unresolved. In absence of detailed hourly demand data, urban energy system models can be improved by using approximations. The logistic regression model, that is calibrated and validated based on the same data as used to cluster areas, is provided to give other researchers and stakeholders the opportunity to gain more insights in their areas of interest based on limited local data. The model is provided as a spreadsheet tool in Appendix B (online). This model can be used to determine the type of demand profile in an area of interest based solely on the relative annual demand of different energy users in that area. This type of cumulative data is more often available than detailed profiles, although, if available, the superiority of local demand profiles remains undisputed and should be used whenever possible.

6. Conclusions

This paper provides a classification and an analysis of local electricity demand profiles in nearly 15000 urban areas. Such systematic spatio-temporal demand profile characterisation has thus far been lacking in literature. The results demonstrate that at neighbourhood, district and municipality scales, three types of area demand profiles can be distinguished, which are termed “residential”, “business”, and “mixed” in this paper, based on the most prevalent local energy users in each of them. Statistical analysis shows that at each scale, these areas are pairwise significantly different from each other, both in terms of their demand profiles and their energy user composition. Moreover, this paper establishes that residential-type demand profiles, used in many energy system models, are found only in a minority of areas, and account for only a small share of the total demand. As a consequence, case studies of local impact of renewables, electric vehicles, etc. that assume solely household demand are representative for only a small share of urban areas and cannot be generalised without errors. Existing and future urban energy system models should therefore be expanded with more realistic and detailed spatio-temporal local demand profiles that account for both household and non-household energy users. Unfortunately, detailed hourly demand profiles, especially for non-household energy users, are currently not publicly available for many urban areas. To overcome this issue and facilitate the use of obtained insights in other models, a spreadsheet tool is provided in Appendix B (online). The tool can be used to approximate local hourly demand based on more readily available energy user composition data in an area of interest. In general, this paper emphasises the importance of efforts to improve urban energy system models through realistic representation of spatial and temporal urban demand heterogeneity.

Acknowledgements

This work was supported by the Netherlands Organisation for Scientific Research (NWO) [Grant No. 408-13-012]. The authors further acknowledge Statistics Netherlands for providing access to detailed urban microdata and for their support during the data analysis.

Appendix A. Data matching

No standardised classification of energy users exists, datasets from different sources often have different subdivisions. In this paper, different data sources are combined: (1) temporal dimension data based on sources from the Netherlands [36,37] for households, and on data from the United States Department of Energy (U.S. DOE) [38] for services, and (2) spatial dimension data based on Statistics Netherlands local administrative registrations [39,40]. Both dimension datasets pertain to households and services, however, the definition of service classes differs between the datasets. The matching between the energy user classes is given in Table A1.

Table A1
Overview of the energy user classes in the temporal and spatial datasets, and the joint spatio-temporal dataset used in this paper.

Temporal data (14 subsectors)	Spatial data (11 subsectors)	Joint data (7 subsectors)
Household	Wonen (*Housing*)	Households
Hospital	Gezondheidszorg (*Healthcare*)	(*not used in this paper*[a])
Hotel	Logies (*Lodging*)	Hotels
Small Hotel		
Large Office	Kantoor (*Office*)	Offices
Medium Office		
Small Office		
Primary School	Onderwijs (*Education*)	Schools
Secondary School		
Restaurant	Bijeenkomst (*Gathering*)	Restaurants
Quick Service Restaurant		
Supermarket	Winkel (*Shop*)	Shops
Stand Alone Retail		
Warehouse	Overig (*Other*)	Warehouses
(no equivalent)	Cel (*Prison*)	*(not used in this paper)*
(no equivalent)	Industrie (*Industry*)	*(not used in this paper)*
(no equivalent)	Sport (*Sports Facility*)	*(not used in this paper)*

[a] Healthcare could not be satisfactorily modelled based on the available datasets due to low R^2-value of the corresponding scaling factor [44].

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.apenergy.2018.08.121.

References

[1] Facchini A, Kennedy C, Stewart I, Mele R. The energy metabolism of megacities. Appl Energy 2017;186:86–95.
[2] Cajot S, Peter M, Bahu J-M, Guignet F, Koch A, Maréchal F. Obstacles in energy planning at the urban scale. Sustain Cities Soc 2017;30:223–36.
[3] Morvaj B, Evins R, Carmeliet J. Decarbonizing the electricity grid: the impact on urban energy systems, distribution grids and district heating potential. Appl Energy 2017;191:125–40.
[4] Paevere P, Higgins A, Ren Z, Horn M, Grozev G, McNamara C. Spatio-temporal modelling of electric vehicle charging demand and impacts on peak household electrical load. Sustain Sci 2014;9:61–76.
[5] Neaimeh M, Wardle R, Jenkins AM, Yi J, Hill G, Lyons PF, et al. A probabilistic approach to combining smart meter and electric vehicle charging data to investigate distribution network impacts. Appl Energy 2015;157:688–98.
[6] Ramirez Camargo L, Zink R, Dorner W, Stoeglehner G. Spatio-temporal modeling of roof-top photovoltaic panels for improved technical potential assessment and electricity peak load offsetting at the municipal scale. Comput Environ Urban Syst 2015;52:58–69.
[7] Capozzoli A, Savino M, Brandi S. Mining typical load profiles in buildings to support energy management in the smart city context. In: 9th International conference on sustainability in energy build; 2017. p. 865–74.
[8] Heiple S, Sailor DJ. Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles. Energy Build 2008;40:1426–36.
[9] Mirakyan A, Guio RD. Integrated energy planning in cities and territories: a review of methods and tools. Renew Sustain Energy Rev 2013;22:289–97.
[10] Allegrini J, Orehounig K, Mavromatidis G, Ruesch F, Dorer V, Evins R. A review of modelling approaches and tools for the simulation of district-scale energy systems. Renew Sustain Energy Rev 2015;52:1391–404.
[11] Lund PD, Lindgren J, Mikkola J, Salpakari J. Review of energy system flexibility measures to enable high levels of variable renewable electricity. Renew Sustain Energy Rev 2015;45:785–807.
[12] Davila CC, Reinhart CF, Bemis JL. Modeling Boston: a workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets. Energy 2016;117:237–50.
[13] Howard B, Parshall L, Thompson J, Hammer S, Dickinson J, Modi V. Spatial distribution of urban building energy consumption by end use. Energy Build 2012;45:141–51.
[14] Mikkola J, Lund PD. Models for generating place and time dependent urban energy demand profiles. Appl Energy 2014;130:256–64.
[15] Keirstead J, Jennings M, Sivakumar A. A review of urban energy system models: approaches, challenges and opportunities. Renew Sustain Energy Rev 2012;16(6):3847–66.
[16] Stoimenov L, Davidovic N, Stanimirovic A, Bogdanovic M, Nikolic D. Enterprise integration solution for power supply company based on GeoNis interoperability framework. Data Knowl Eng 2016;105:23–38.
[17] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy 2012;42:68–80.
[18] Pfenninger S, DeCarolis J, Hirth L, Quoilin S, Staffel I. The importance of open data and software: is energy research lagging behind? Energy Policy 2017;101:211–5.
[19] Mastrucci A, Baume O, Stazi F, Leopold U. Estimating energy savings for the residential building stock of an entire city: a GIS-based statistical downscaling approach applied to Rotterdam. Energy Build 2014;75:358–67.
[20] Fonseca JA, Schlueter A. Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. Appl Energy 2015;142:247–65.
[21] Voulis N, Warnier M, Brazier FMT. Impact of service sector loads on renewable resource integration. Appl Energy 2017;205:1311–26.
[22] Andersen FM, Larsen HV, Boomsma TK. Long-term forecasting of hourly electricity load: identification of consumption profiles and segmentation of customers. Energy Convers Manage 2013;68:244–52.
[23] Andersen FM, Larsen HV, Juul N, Gaardestrup RB. Differentiated long term projections of the hourly electricity consumption in local areas. The case of Denmark West. Appl Energy 2014;135:523–38.
[24] Brownsword RA, Fleming PD, Powell JC, Pearsall N. Sustainable cities - modelling urban energy supply and demand. Appl Energy 2005;82(2):167–80.
[25] Yamaguchi Y, Shimoda Y, Mizuno M. Proposal of a modeling approach considering urban form for evaluation of city level energy management. Energy Build 2007;39(5):580–92.
[26] Chow M, Zhu J, Tram H. Application of fuzzy multi-objective decision making in spatial load forecasting. IEEE Trans Power Syst 1998;13(3):1185–90.
[27] Alhamwi A, Medjroubi W, Vogt T, Agert C. GIS-based urban energy systems models and tools: Introducing a model for the optimisation of flexibilisation technologies in urban areas. Appl Energy 2017;191:1–9.
[28] Best RE, Flager F, Lepech MD. Modeling and optimization of building mix and energy supply technology for urban districts. Appl Energy 2015;159:161–77.
[29] Pitt BD, Kitschen DS. Application of data mining techniques to load profiling. In: Power industry computer applications, 1999. PICA'99. Proceedings of the 21st 1999 IEEE international conference; 1999. p. 131–6.
[30] Gerbec D, Gasperic S, Smon I, Gubina F. Allocation of the load profiles to consumers using probabilistic neural networks. IEEE Trans Power Syst 2005;20(2):548–55.
[31] Andersen FM, Larsen HV, Gaardestrup RB. Long term forecasting of hourly electricity consumption in local areas in Denmark. Appl Energy 2013;110:147–62.
[32] Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv 1999;31(3):264–323.
[33] Figueiredo V, Rodrigues F, Vale Z, Gouveia JB. An electric energy consumer characterization framework based on data mining techniques 2005;20(2):596–602.
[34] Räsänen T, Voukantsis D, Niska H, Karatzas K, Kolehmainen M. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Appl Energy 2010;87(11):3538–45.
[35] Lamedica R, Santolamazza L, Fracassi G, Martinelli G, Prudenzi A. A novel methodology based on clustering techniques for automatic processing of MV feeder daily load patterns. In: Power engineering society summer meeting; 2000. p. 96–101.
[36] NEDU. Verbruiksprofielen. < http://nedu.nl/portfolio/verbruiksprofielen > [last accessed online 05/06/2018].
[37] Energy Research Centre of the Netherlands. Energie-Nederland, Netbeheer Nederland, Energietrends 2016. Tech rep. The Hague, the Netherlands; 2016.
[38] Deru M, Field K, Studer D, Benne K, Griffith B, Torcellini P, et al. U.S. Department of

Energy commercial reference building models of the national building stock. Tech rep. Golden (Colorado, USA): National Renewable Energy Laboratory; 2011.
[39] Statistics Netherlands. Microdata: conducting your own research. < https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research > [last accessed online 16/05/2018].
[40] Statistics Netherlands. Kerncijfers wijken en buurten. < http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLnl&PA=83220NED&LA=nl > [last accessed online 16/05/2018].
[41] Organisation for Economic Co-operation and Development. OECD E-Government Studies: Netherlands. Paris: OECD Publishing; 2007.
[42] Statistics Netherlands. Begrippen. < https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen?tab=b\#id=buurt > [last accessed online 29/04/2018].
[43] Benabbas F, Khadir MT. Fuzzy C-Means clustering and Kohonen maps for the identification of regional electricity load day types. Int J Hybrid Intell Syst 2011;8:81–92.
[44] Voulis N, Warnier M, Brazier FMT. Statistical data-driven regression method for urban electricity demand modelling. In: IEEE environment and electrical engineering conference; 2018 [in press].
[45] Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979:224–7.
[46] Chicco G, Napoli R, Piglione F. Comparisons among clustering techniques for electricity customer classification. IEEE Trans Power Syst 2006.
[47] McDonald J. Handbook of biological statistics. 3rd ed. Baltimore (Maryland, U.S.): Sparky House Publishing; 2014.
[48] Jakob M, Kallio S, Bossmann T. Generating electricity demand-side load profiles of the tertiary sector for selected European countries. In: 8th International conference improving energy efficiency in commercial buildings; 2014. p. 15.
[49] Mairet N, Decellas F. Determinants of energy demand in the French service sector: a decomposition analysis. Energy Policy 2009;37(7):2734–44.
[50] Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. Energy Build 2008;40(3):394–8.
[51] Robinson AP, Blythe PT, Bell MC, Hübner Y, Hill GA. Analysis of electric vehicle driver recharging demand profiles and subsequent impacts on the carbon content of electric vehicle trips. Energy Policy 2013;61:337–48.