Building and Environment 183 (2020) 106964

Contents lists available at ScienceDirect

Building and Environment

journal homepage: http://www.elsevier.com/locate/buildenv

Modelling urban-scale occupant behaviour, mobility, and energy in buildings: A survey

Flora D. Salim [a,*], Bing Dong [b], Mohamed Ouf [c], Qi Wang [d], Ilaria Pigliautile [e], Xuyuan Kang [f], Tianzhen Hong [g], Wenbo Wu [h], Yapan Liu [b], Shakila Khan Rumi [a], Mohammad Saiedur Rahaman [a], Jingjing An [k], Hengfang Deng [d], Wei Shao [a], Jakub Dziedzic [i], Fisayo Caleb Sangogboye [j], Mikkel Baun Kjærgaard [j], Meng Kong [b], Claudia Fabiani [e], Anna Laura Pisello [e], Da Yan [f]

[a] Computer Science and IT, School of Science, RMIT University, Melbourne, VIC, Australia
[b] Department of Mechanical and Aerospace Engineering, Syracuse University, Syracuse, NY, USA
[c] Department of Building, Civil, and Environmental Engineering, Concordia University, Montreal, QC, Canada
[d] Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA
[e] Department of Engineering, University of Perugia, Perugia, Italy
[f] School of Architecture, Tsinghua University, Beijing, China
[g] Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[h] Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, TX, USA
[i] Norwegian University of Science and Technology, Department of Energy and Process Engineering, Norway
[j] Center for Energy Informatics, University of Southern Denmark, Odense, Denmark
[k] Civil Engineering and Architecture, Beijing University, Beijing, China

ARTICLE INFO

Keywords:
Big data
Occupant behaviour
Energy modelling
Mobility
Urban data
Sensors
Machine learning
Energy in buildings
Energy in cities

ABSTRACT

The proliferation of urban sensing, IoT, and big data in cities provides unprecedented opportunities for a deeper understanding of occupant behaviour and energy usage patterns at the urban scale. This enables data-driven building and energy models to capture the urban dynamics, specifically the intrinsic occupant and energy use behavioural profiles that are not usually considered in traditional models. Although there are related reviews, none investigated urban data for use in modelling occupant behaviour and energy use at multiple scales, from buildings to neighbourhood to city. This survey paper aims to fill this gap by providing a critical summary and analysis of the works reported in the literature. We present the different sources of occupant-centric urban data that are useful for data-driven modelling and categorise the range of applications and recent data-driven modelling techniques for urban behaviour and energy modelling, along with the traditional stochastic and simulation-based approaches. Finally, we present a set of recommendations for future directions in data-driven modelling of occupant behaviour and energy in buildings at the urban scale.

1. Introduction

Cities occupy only two percent of the world's land, but they account for more than seventy percent of global CO2 emissions and two-thirds of the world's energy [1].Although simulation technologies are usually the first choice for modelling energy efficiency in the building, transportation and industrial sectors, human behaviours are often overlooked [2]. Understanding how to model human activities at the urban scale is crucial to enable efficient and reliable urban infrastructure, resilient to natural disasters and extreme weather events. In the past decade, occupant behaviour within a building has been largely studied using indoor environmental sensor and energy data [1,3]. Given the advancements of Internet of Things (IoT) and wireless networks, unprecedented growth of *occupant-centric urban data,* from various sources within urban areas, such as traffic, social media, telecommunication data, and different sensors deployed within cities, new opportunities to scale this study up have emerged.

Existing review papers in the area of urban computing and big data

* Corresponding author.
E-mail address: flora.salim@rmit.edu.au (F.D. Salim).

https://doi.org/10.1016/j.buildenv.2020.106964
Received 13 December 2019; Received in revised form 4 May 2020; Accepted 10 May 2020
Available online 11 June 2020

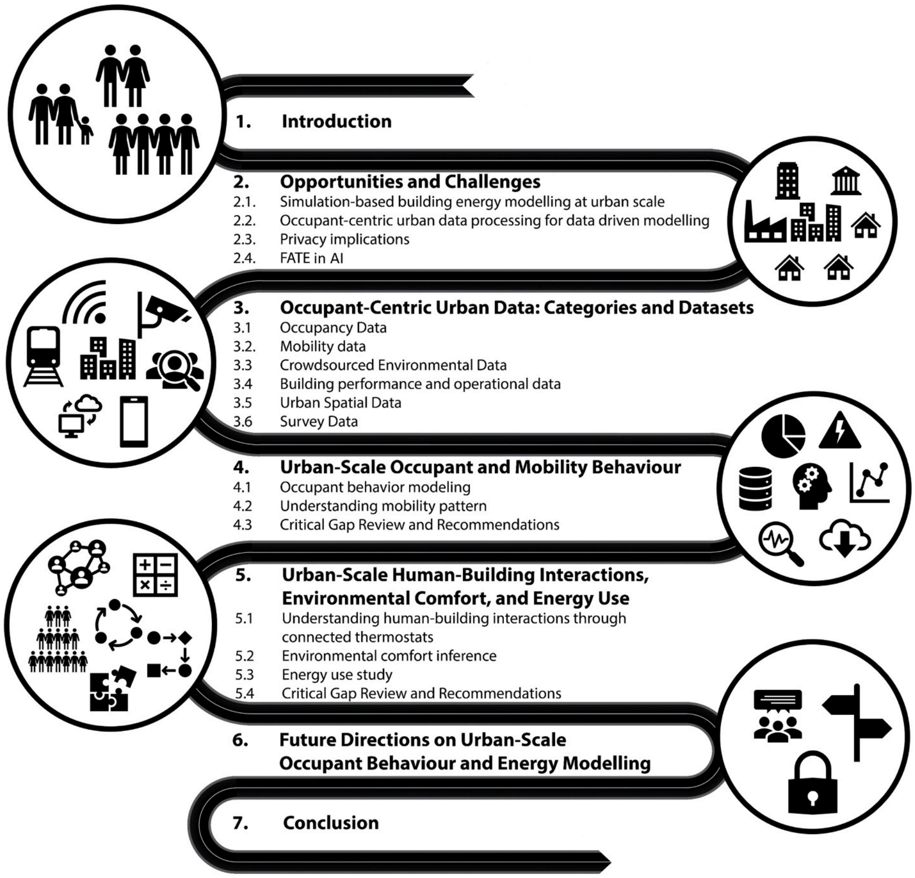

Fig. 1. Manuscript road map.

mainly focus on the different data mining and machine learning techniques to deal with such data (e.g. Ref. [4–6]). Modelling occupant behaviour at the urban-scale and their corresponding energy use have not been the focus. In urban computing, for example, the focus is more on fusing multiple data sources [4] and enabling the multi-task capabilities of the frameworks to harness these types of data simultaneously. Recently, there have been efforts to use this data to analyse occupant behaviours in buildings [7] and the corresponding energy models thereof at an urban scale [8]. This survey paper aims to fill this gap by presenting a detailed overview of the research landscape for occupant behaviour in buildings and cities and how they influence energy use. We review the data sources and methods that have been used to model human activities and energy use in multiple buildings and at urban scale.

We seek to gain insights into the following research questions:

1. What are the important urban-scale human activities and behaviours that influence energy use? e.g. mobility, human outdoor comfort?
2. Which datasets are available and relevant to support and validate those models?
3. What are the key applications of urban-scale occupant modelling and the corresponding ways of modelling that can represent human activities and behaviours at various levels of detail at the urban scale?

Therefore, in this paper, we investigate these questions with opportunities and challenges that come with *occupant centric urban data*.

1.1. Definition of occupant-centric urban data

Occupant-centric urban data can be used to characterise occupant behaviours, encompassing activities done indoors and outdoors, which influence energy use in buildings, and at an urban scale. This data is typically derived and aggregated from multiple heterogeneous sources from building sensors and IoT, mobility data, occupancy and energy data, and surveys.

Kitchin [9] explicitly took the more well-known attributes of big data and associated these with cities. This is largely because cities have become more instrumented and have now experienced the data deluge, useful for more real-time, fine-grained, improved understanding and control of the city infrastructures. According to Kitchin [9], urban big data can have one to many of the following attributes: big in scale, fast-paced, highly dynamic, high in variety, detailed or fine grained, longitudinal or collected over a long period of time, inter-relatable, and scalable.

To clarify the scope of the paper, we focus on urban data that comes from different sensors deployed within cities, and performance and occupant data of multiple buildings in a district, neighbourhood or, possibly, the entire city. The focus is on the usage of urban data for describing occupant behaviour to improve occupant-centric building operations, energy management, building performance simulations, or other related tasks. This, therefore, brings a unique contribution to the body of literature.

Although modelling urban behaviour should ideally also consider transportation, roads, and other infrastructures in the built environment, this paper will only cover relevant papers on mobility with applications related only to dynamic population estimation [10] for occupant behaviour and energy modelling at multiple buildings, neighbourhoods, and city scale.

1.2. Review methodology

In order to explore those questions, we have reviewed more than 400

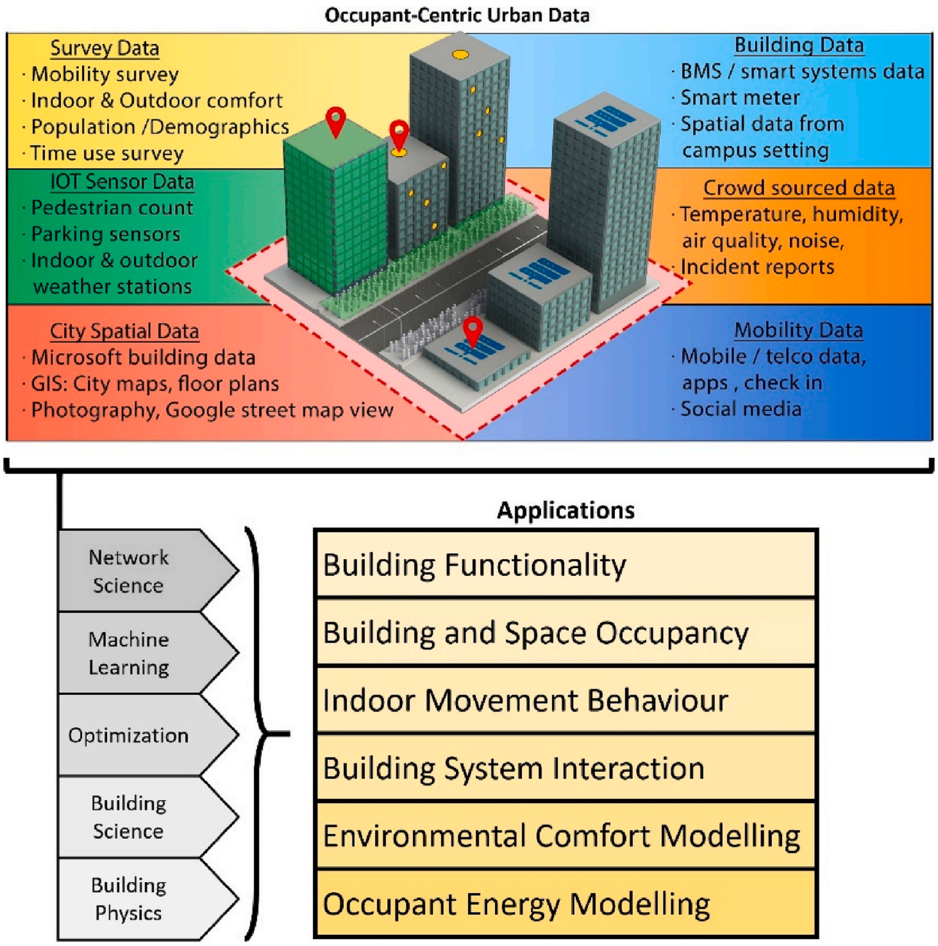

Fig. 2. From occupant-centric urban data to building, occupant behaviour, and energy modelling.

papers from multiple disciplines, although only 270 are finally included in the reference list. This multidisciplinary review includes state-of-the-art approaches in building science, physics, engineering, mathematics, and computer science, in particular, data science and machine learning, to solve the problems covered in this paper.

The review paper aims at pointing out occupant-centric data availability to develop both occupants' behaviour and energy use models referring to the urban scale. In this framework, a multi-disciplinary research is conducted focusing on classifying existing datasets related to both the fields of study, i.e. occupants'/citizens' behaviour and building energy usage (review section 3). In particular, the authors recognise six main categories as follows: occupancy, mobility, environmental, building-related, urban spatial and survey data. The highlighted categories are therefore explored by paying attention on data sources which have to be occupant centric. Specific focuses are pointed out in each category section to present different existing and developing technologies dedicated to occupant-centric data collection in complex urban systems. Therefore, the research survey explores modelling approaches moving from the defined datasets distinguishing in between the two macro-areas. Given the outlined research framework, we chose papers from Google Scholar with publication year after 1990.

This paper is organised as the following. Section 2 provides an overview of the opportunities and challenges that comes with occupant-centric urban data. Section 3 introduces sources of occupant-centric urban data with a range of examples. Section 4 discusses applications and modelling approaches of urban-scale occupant and mobility behaviours. Section 5 provides an overview of applications and modelling techniques of urban-scale human-building interactions, environmental comfort, and energy use. An overarching outlook with discussions, perspective and future directions is given in section 6. Finally, Section 7 concludes the paper. Fig. 1 shows the structure of the whole manuscript, and it can be used as a content navigator.

2. Opportunities and challenges

The widespread adoption of sensing technologies, in smartphones, home devices, and high-speed wireless connectivity, as well as the high availability of The Internet of Things and open data from government and other organisations provide an untapped source of rich knowledge of the city. This brings an unprecedented opportunity to sense and model occupant behaviour and energy usage patterns of buildings at a neighbourhood, regional, or an urban scale. The variety of urban data and the range of applications that have been investigated across different fields are depicted in Fig. 2.

Prior to this ubiquitous computing era, physics-based models have been used extensively to model occupant behaviours and energy usage, based on thermodynamics and other established principles. However, this is mostly done only at a building scale, due to the high-dimensionality and linearity of many physics-based models, making them expensive to compute, even for a single building [11]. To scale this up, recent efforts on physics-based modelling are combined with calibration processes, tested through simulations, as discussed in Section 2.1.

Physics-based models also pose some limitations in capturing the high non-stationarity relationships from the input variables [12]. Data-driven models have therefore become more popular as the complex and dynamic relationships can be captured with the available data, as discussed further in Section 2.2.

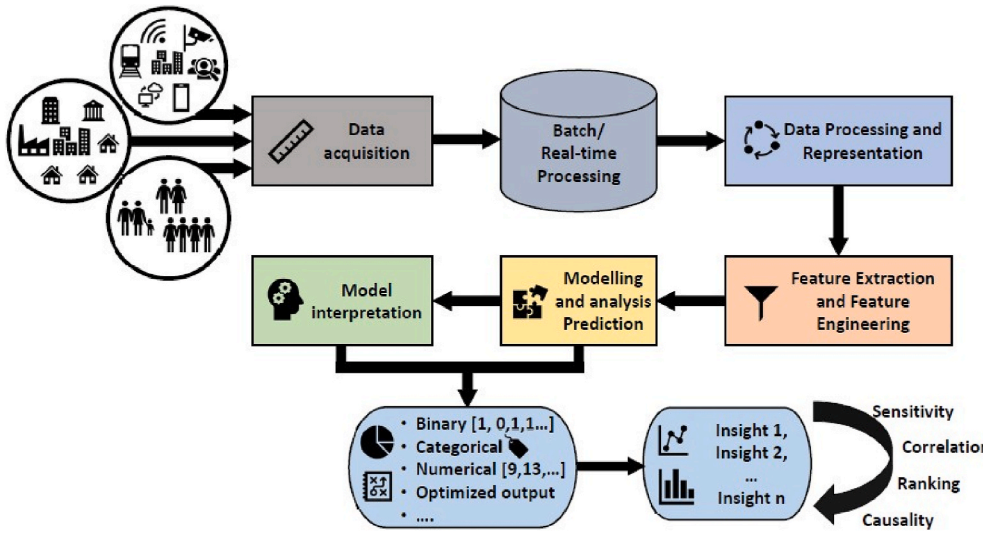

Fig. 3. An overarching view of occupant-centric urban data processing.

This data is typically voluntary generated by users (e.g. through crowdsourcing, or crowdsensing), or automatically generated due to a system or application usage (e.g. logged by mobile devices, or apps, or smart home sensors), as discussed further in Section 3. Therefore, dealing with occupant-centric urban data leads to privacy implications, as discussed in Section 2.3.

Finally, physics-based models and data-driven models can have inherent bias, due to the limited observations and the approximations that are used to generalise this model. In addition, the black-box nature of data-driven models, lead to the need to ensure FATE (Fairness, Accountability, Transparency, and Ethics) in these models. This is further discussed in Section 2.4.

There are additional opportunities and challenges given the rise of cyber-physical systems, digital twins, and new actuating systems. However, this review paper only focuses on sensing and modelling occupant and energy behaviours, therefore, those are out of the scope of this paper.

2.1. Simulation-based building energy modelling at urban scale

The very first application of big data at the urban scale is to model building energy use. Traditionally, urban building energy models (UBEMs) apply physics-based modelling approaches to simulate heat flow in and around buildings and estimate their operational energy as well as indoor and outdoor environmental conditions [8,13]. They can either rely on detailed multi-zone building performance simulation engines (e.g. EnergyPlus, IDA-ICE) or simplified thermal resistance-capacitance (R–C) models [14]. To integrate occupant-centric urban data in UBEMs, Geographic information systems (GIS) has been very useful [15,16] and were previously used in the Energy Atlas Berlin initiative [17], and the work by the European Institute for Energy Research [18]. For many urban areas, the open standard CityGML was also introduced, which provides 3D city models that can be used as inputs for UBEMs. CityGML represents 3D city models at varying levels of detail, which may include additional information for each building in a city, such as the year of construction, building type, and energy use [15].

Irrespective of the level of detail used in creating physics-based UBEMs, uncertainties that arise in the modelling process from various sources, including assumptions regarding occupancy patterns and occupant behaviour, may limit their capabilities [14]. One approach to represent this stochasticity is to use the Monte Carlo simulation to capture the correlations between the variations of the model input and output parameters [19–22]. Another approach is through a calibration process to refine UBEM inputs, so the simulated results match collected data. In the existing literature, Bayesian calibration is a commonly used approach for probabilistic calibration [23–28]. Overall, occupant-centric urban data can provide more information to address the more complex and multi-faceted behaviours and uncertainties in UBEMs, which are typically not captured in modelling outcome and input parameters.

2.2. Occupant-centric urban data processing for data driven modelling

Occupant-centric urban data is usually generated from various sources, thus data-driven modelling aims to turn this vast amount of real-world data into actionable knowledge through practical and applied results [29]. The increased amount of urban data expedites the data-driven approaches to study occupant behaviours or energy usage in city level rather than individual building level only [30,31]. Several steps are involved with any data-driven modelling, including occupant behaviours and energy usage modelling. In general, the associated steps are data acquisition, data cleaning and extraction, data aggregation and representation, data mining and machine learning, modelling and interpretation (Fig. 3).

Occupant-centric urban data can come from heterogeneous sources, as described in Section 3, including diverse individual users and organisations occupying the built environment. Often there is no schema that has been designed to fit the purpose of the data collection. Therefore, acquisition is the first step, and the next one after is data preprocessing and representation. This includes data cleaning, to deal with highly sparse and noisy data. Since the data comes in different shapes and forms, a common approach is to represent the data in memory or storage, prior to the next step, which is feature extraction and engineering. With deep learning, this step is replaced with representation learning, often in an end-to-end manner. The next step is analysis and modelling, although this can also be done in parallel with the previous step, in order to decide on the architecture and models to best capture the behaviours to be estimated or predicted. Finally, these models need to generate actionable insights for the range of applications as discussed further in Section 4. The steps are discussed further in the Appendix.

2.3. Privacy implications

Urban-scale data include several data types, including crowdsourced and mobility data that are produced or relate to individuals. Therefore, if urban-scale data include references to individuals, this result in privacy implications. Many data providers address this problem using either pseudonymization or anonymization [32].

Fung et al. [33,34] present the different categories of attack models comprising of record, attribute and table linkage alongside probabilistic attack to classify the state-of-art privacy models that can guarantee and suffice for privacy protection for each attack model. Using this categorization, a number of frameworks [35–38] utilizing the highlighted privacy models has been proposed for protecting the sensitive information of individuals corresponding to either some form of mobility dataset or dataset obtained from a cyber-physical system.

Recently, a number of empirical studies [39–42] is beginning to assess these privacy models for pseudonymizing and anonymizing urban-scale and cyber-physical data. The result obtained from these studies indicates that no privacy model is suited to protect all protection goals as a combination of different attacks still poses a problem to existing approaches. Similar results are obtained in similar studies presented in [40–42]. These results are indicative of the difficulty and complexity for protecting urban-scale data.

2.4. FATE in AI

The rapid proliferation of pervasive technologies has made our living environments (i.e. cities) “smart”. At the same time, the evolution of smart environments exposes potential risks for individuals and society. For instance, the automation of our everyday activities could delegate the governance of such activities and the whole city to the intelligent engines that can use a set of algorithmically defined policies. Many of these algorithms are designed as black boxes. The end users have very little or no understanding of how they work. However, this lack of understanding does not prevent people from relying on them in most of the cases [43]. Recently, people started questioning the trustworthiness of these models.

The results derived from the available data can have serious issues (e.g., cultural biases and non-conforming logic) when the observation is prioritised over ethical considerations which has an adverse effect in the long run [44,45]. The research in FATE (Fairness, Accountability, Transparency, and Ethics) in AI has been driving investigations into these questions [46]. The early solutions include local surrogate models or interpretable models such as LIME [47] which can be used to explain individual predictions of black-box models.

3. Occupant-centric urban data: categories and datasets

This section summarizes the occupant-centric urban datasets from different sources which are widely used in urban-scale occupant behaviour and energy modelling. Datasets covered in this section are divided into six different categories: Occupancy data, Mobility data, Crowdsourced Environmental data, Building performance and operational data, urban spatial data and survey data.

3.1. Occupancy data

To model human mobility patterns in indoor and outdoor, occupancy data can assist in many ways. Various IOT sensors are widely used to collect the occupancy data. This includes IoT for collecting occupant activities in smart cities and wireless connectivity data within and across building complexes.

3.1.1. IoT and smart cities data

Sensors have been used in previous studies to predict the number of pedestrians in a location including visual, laser and thermal-based sensors [81,82]. Recently, some researchers estimated the dynamic population by measuring the vehicle density using subway and traffic camera in New York City [48].

One important aspect of understanding occupant behaviours in an urban environment is by analysing occupancy of car parks. In the past, most researchers can only study the simulation environment [49]. Recent works on urban-scale parking analytics have largely utilised data from parking sensors, cameras and loop detectors [50–59]. Many cities around the world have published open parking datasets.

Table 1
Overview of different data sources to collect human mobility.

Types of Data	Purpose
Check-in Data	- Predicting friendship between users [70] - Measuring the social diversity [71] - Next location Prediction [72] - Studying the spatial-temporal regularity of user activity [73] - Studying the impact of mobility and social relationship on each other [74] - POI demand modelling [75]
Social media data	- Understanding collective movement pattern [76]
CDR Data	- Population Displacement measurement in natural disaster [77] - Understanding human mobility [78–80]
Building Energy Data	- Electrical data from smart meter [83–88]

3.1.2. WiFi, bluetooth, and other wireless network data

A step towards scaling up to urban-level datasets is the campus-level or complexes of blocks of buildings. Existing work has considered collecting large-scale occupant dataset from such settings using WiFi infrastructures supplemented with other sensors. Ruiz et al. [60] presents a study that utilises WiFi infrastructure to collect spatial data on a whole campus-level and analysed the data for relevant patterns. Sevtsuk et al. [61] presents a similar study for a campus scale. Sangogboye et al. [62] combines WiFi data and datasets from count sensors [63] to provide datasets on occupancy and occupant trajectories. Christensen et al. [64] also considered WiFi enabled devices and other Information and Communications Technologies (ICT) network data to provide data on building occupancy. Das et al. [65] considered how to provide occupant presence datasets based on multi-modal sources based on WiFi data, electricity data, water consumption data. Lastly, Schauer et al. [66] proposed a method for estimating crowd densities and pedestrian flows based on WiFi and bluetooth data in an airport. In large commercial and mixed-use buildings, such as shopping malls and offices, WiFi and Bluetooth data are often used for monitoring visitors [67–69].

3.2. Mobility data

A variety of datasets have been used to study human mobility, each being a potentially valuable resource to inform occupant behaviours in built environment and facilitate energy simulations. While exotic datasets have been analysed occasionally (e.g., $100 banknotes in [89]), datasets from different resources (see Table 1) are used in most of the mobility studies.

3.2.1. Geospatial trajectories/GPS data

The first data set is collected from GPS equipment and have a variety of sampling rates ([90]). Most of the trajectories are logged in a dense representation, e.g. every 1 5 s or every 5–10 m per point, making the recorded movements with almost complete spatial information. Some of the datasets have been made open, such as GeoLife [90], PrivaMov [91], Nokia Mobile dataset [92], etc., making them standard datasets to test new measures and/or algorithms. Smartphone-based Global Positioning System (GPS) log dataset was studied to understand occupant behaviour in urban scale [7]. The dataset was obtained from Cuebiq company which collects data from about 70 million U.S. smartphones through smartphone applications. Human mobility in the city of Chicago has been investigated based on GPS data collected from the social network - Twitter. This dataset contains electricity consumption figures of residential buildings for 801 spatial divisions in Chicago, as well as 8,798, 090 GPS records in the city [93]. Another study was conducted by Kang et al., 2019, which utilised GPS data of social network software from Tencent - one of the largest IT company in China. In this research, active positioning data contains information of occupancy patterns was collected from the social network software [94].

3.2.2. Call detail records (CDR) dataset

This dataset is collected from cell phone communications by telecom companies. When a person makes a phone call or receives/sends a text message, the nearest cell phone tower routes the communication, and its location—roughly the user's locationis recorded [95]. The records include anonymous User ID, longitude, latitude, and timestamp of the phone activity. The accuracy of the location is about 200–300 m.

Research-based on the CDR dataset has uncovered some fundamental knowledge on human mobility including high uniformity [95], ultraslow diffusion [96], high predictability [97], and motif composition [78,98]. 834,690,725 anonymized CDRs of 1 million users in the Boston metropolitan area were collected for a period of two months in 2010 [79] to study urban scale human mobility. Wheatman [99] studied population dynamics based on the CDR data collected from a small European country over nine months in 2015. Despite having enabled important findings, CDR data have their own limitations. First, the precision of the data is determined by the distributions and coverage areas of cell phone towers which is 13km2 per cell phone tower. Such precision has been instrumental when studying human mobility over larger scales (e.g., a state or a country), but it may not be able to capture human mobility at smaller scales (e.g., a district or a city). Second, data privacy and ownership sometimes make verification and validation work challenging [100].

3.2.3. Social media check-ins and apps

When a person uses social media applications, his or her geographical location is recorded. Social media data enables large-scale analyses and scientific discovery, such as universal mobility pattern around the world [101,102] and across cities [103], mobility-based inequality across U.S. cities [104], disaster-induced mobility perturbation [105], etc. Despite these important findings, the representativeness of social media data has been questioned since the beginning; people from certain socio-demographic groups are more likely to use social media and this can especially be true for smaller platforms. In this case, results may only reflect the mobility behaviours from biased sample rather than the entire population. Though there are a few studies that investigated this representativeness issue [106–108], we still lack a comprehensive solution.

3.2.4. Location based service (LBS) data

Location-based service (LBS) data are collected by location-based service providers who usually embed their functions into many smart phone apps. In such a way, these datasets often cover large percentage of populations with decent quantity (e.g., 100 or more records per day) and resolution. LBS data has supported promising new directions of research. However, there are some challenges associated with human mobility data from LBS data. Data acquired from LBS is sparse in nature. Again, data from LBS are generally bias to tech savvy young people [109].

3.2.5. Transportation data

Another source to collect human mobility is from daily transportation, including bicycle, bus, train, or taxi in their daily life. Mobility dynamics in real-time was measured in Ref. [110] using location data from the public transportation system along with cellular network data. The human mobility pattern of a particular area was captured with a spatiotemporal analysis of the shared bicycling system in Ref. [111,112]. In Ref. [113], the authors acquired indoor and outdoor traffic traces from traffic and occupancy sensing system installed by the local government of Hong Kong. This dataset contains traffic information about 617 roads and 118-story building.

Table 2
Overview of different data sources to collect environmental data according to pedestrians' perspective.

Types of Data	Purpose
Sensor networks	- Real time noise/pollution measures for population alert [115] - Fine-grained city air quality map through automobile built-in sensors [119] - Visualize air pollution propagation [118] - IoT platform for public consultancy of air quality [116] - Prototype of IoT-based technology for noise and air quality pollution real-time monitoring [117]
Sensor & Social media	- Monitoring and mitigation of urban noise pollution [126]
Environmental sensor & Survey	- Investigation of dynamic thermal comfort [128] - Extreme learning machine approach to predict thermal comfort in outdoors [129]
Wearables	- Map PM2.5 distribution through miniaturized, personal devices [121] - Map transient outdoor comfort [123] - Understanding dynamic thermal comfort [124] - Environmental mapping according to pedestrian perspective [130] - Enhancement of crowd-sourcing air quality through low costs participating [120]
Wearables & Social media	- Evaluation and representation of sound environment [130] - Soundscapes related to people perception [127] - Sound classification and mapping [131]
Wearables & physiological data	- Physiological response to different microclimates [132]

Table 3
Overview of different data sources to collect building operation and energy use data.

Types of Data	Purpose
BMS data	- Predicting occupant behaviour at the zone level for using lights [133], blinds [134], windows [135] - Predicting thermostat setpoint change requests [136] - Identifying operational variables and their optimal settings to minimize energy use and mantain occupant comfort [137]
Smart thermostat and HVAC use data	- Understanding thermostat usage patterns in different climate zones [138] - Predicting occupancy patterns in residential homes [138] - Estimating thermal time constant values for residential buildings [139]
Smart meter data	- Developing and validating energy disaggregation methods to identify end-use patterns [85,88,140] - Identifying consumers with the most potential for energy reduction through demand response [141,142]

3.3. Crowdsourced Environmental data

Urban environment presents peculiar conditions mainly related to city internal structure, intensity and typology of anthropogenic activities, and lack of greenery and water bodies compared to suburban areas. Cities are further characterized by high heterogeneity which produces dynamic patterns of all the main parameters influencing citizens' life quality (including environmental comfort) and buildings energy consumption. Focusing on pedestrians' well-being, human perception involves different spheres, and thus scientific effort includes data collection related to (i) urban air quality, (ii) noise pollution, (iii) and outdoor thermal comfort. An overview of the most recent studies [114–127] involving these types of data collection at the urban scale is presented distinguishing among different involved monitoring systems as summarized in Table 2.

3.4. Building performance and operational data

Building data that describe different aspects of building performance may include energy use (i.e., smart meter data) or operational data (e.g., smart systems data, as well as occupancy data). The following subsections provide details on the availability of such data at an urban-scale (see Table 3).

3.4.1. Indoor environmental data

The proliferation of indoor environment sensing technologies has enabled the collection of various data points related to the indoor environment. These sensors can collect various information regarding indoor environments such as temperature, humidity, light, CO2 and air velocity in a time-series manner. Combining these data from various environment sensors can provide an unprecedented level of environmental situation awareness inside buildings. There are many indoor environment datasets collected for various purposes [143–146]. Comfort Database [147], and UCI Occupancy Detection Data Set [148]. Some of those are available for public use including ASHRAE Global Thermal.

3.4.2. BMS/smart systems data

Data from building management systems (BMS) can provide a wealth of information regarding occupancy, operational patterns and occupant behaviour. Previous studies used these datasets to develop stochastic occupant behaviour models, albeit at the zone or building level [133–135]. BMS data is typically available for individual buildings and privately owned. However, some researchers obtained BMS data from groups of buildings. For example, Gunay et al. [136] used BMS data from a cluster of three governmental office buildings to collect records of zone temperature, outdoor temperature, and temperature set-points at 30-min intervals. In contrast to BMS, IoT devices such as smart thermostats a provide much larger datasets from thousands of buildings. For example, the 'Donate Your Data' program administered by Ecobee Inc. contains over three million days of data from more than 112,000 thermostats across North American homes [138]. It consists of measurements recorded at 5-min intervals for the indoor and outdoor air temperature, set-points, indoor RH, motion sensors, as well as HVAC equipment run times. Additionally, metadata such as type of home, construction date, and geographical location are available [138]. Other datasets, such as Pecan Street Research Institute's DataPort [87] also contains indoor temperature readings; some of which are using Google Nest's connected thermostats.

3.4.3. Building energy data

Smart meters record electricity usage of households. It is estimated that the number of readings will surge from 24 million a year to 220 million per day for a large utility company with advanced metering infrastructure (AMI) [149]. A review of available smart meter datasets is conducted by Babaei [150]. Two datasets covered more than 100 buildings. The first dataset [83] includes 1-min whole building electricity consumption data of 1 day for 400 houses. The other dataset [84] contains whole buildings level consumption data of 400 residential buildings during one to two years. Other datasets covered fewer buildings [83,85,86]. Pecan Street Research Institute [151] provides the world's largest source of disaggregated customer energy data, which contains electricity data collected from 722 houses in the U.S [87]. It also includes energy audit results and annual surveys which provided physical characteristics of the homes and sociodemographic descriptions of the occupants [88].

The databases of building characteristics and their annual energy consumption can present an overview of building energy use trends at the national scale. In the U.S. the Commercial Building Energy Consumption Survey (CBECS) collects building information such as the address, building name, size, and use category, which is provided along with energy use and its breakdown by energy source. In its most recent available datasets from 2012, 6720 records are provided representing commercial buildings from every state [152]. Similar information is provided for residential buildings in the Residential Energy Consumption Survey (RECS), whose latest datasets published in 2015 includes data from 5686 households in that year [153].

3.5. Urban spatial data

More cities are making their data public, e.g., building footprint, assessor's records, building permits, available at their open data portals. These data after curation and integration with other data sources enable the creation of 3D city models that describe physical urban objects at various levels. Cities like New York City and Berlin make their 3D city models available in CityGML which is an international Open Geospatial Consortium standard for the representation and exchange of 3D city models. CityGML defines the 3D geometry, topology, semantics, and appearance of urban objects, including buildings and their components, bodies of water, city furniture (street lighting, traffic lights), transportation infrastructure (streets, roads, bridges, tunnels), and vegetation. CityGML has the flexibility of representing urban objects at various levels of details which is a critical feature enabling 3D city models to be enriched with more becoming available data. In this paper, we discussed urban spatial data based on their sources, including a) geospatial information, maps and floor plans and b) Microsoft building footprint datasets.

3.5.1. Geospatial information, maps, and floor plans

The route information in the world map can be acquired from Open Street Map (OSM) [154] and Google Maps API [155]. OSM is a user-generated open-world street map which is free to use and editable [156]. The road network on OSM maps can be fused with other data sources to aid various mobility applications [157,158]. Google Maps API provides the static and dynamic world map with route and place information. The road network on google map provides real-time traffic information. Travel distance between two places can be measured on this map. Different geospatial information, like Point of Interest (POI), can be obtained from Google Maps API. Location of POI in longitude and latitude format and the type of POI can be obtained from Google Places API [159]. However, some geospatial information like airport tarmac, pedestrianized paths and shortcuts may not be available in these maps. In Ref. [160], the authors proposed a GPS trajectory-based deep-learning framework named COLTRANE to generate maps for different environments.

3.5.2. Microsoft building footprint datasets

Creating a UBEM typically starts with having building footprints of the area of study. Then gradually adding data such as building use and building height to each footprint. These footprints can be gathered from local authorities or in some cases are available online to the public. Microsoft made very significant efforts by providing building footprint datasets free of cost. As of today, these datasets contain over 125 million [161] building footprints of all 50 U S. states and near 12 million [162] building footprints in all Canadian provinces and territories. The datasets are in GeoJSON format which includes the information of building coordinates and building footprint polygon geometries.

3.6. Survey data

Various types of survey data by many users and organizations have been used to provide groundtruth for statsitical analysis of occupant behaviour and energy in buildings, as described in the following subsections.

Table 4
Overview of time-use survey in occupant behaviour research at urban scale.

Survey	Data size	Methods and Purpose
French time-use survey	15,441 individuals of 7949 house- holds. from Feb.1998 to Feb. 1999, at 10-min interval	Methods: Stochastic models based on statistics and Markov process. Purpose: Predicted at-home probability, the conditional probability to start an activity, probability distribution of activity duration [166]
American time-use survey	Data from 13,000 individual 24-h time diaries from 2006, time step unfixed	Methods: Bootstrap sampling. Purpose: Typical household occupancy profiles and overall residential energy use profiles [165]
UK time-use survey	Data of 11,700 UK citizens from 6500 households in 2000, at 10- min interval	Methods: Markov Chain Monte Carlo (MCMC), K-mode clustering. Purpose: Household occupancy model as an input for energy modelling [167]. Occupancy model to generate stochastic appliance-use or occupancy profiles at daily basis [168]. Occupancy profiles as an input for residential building energy end-use models [169].
Spanish time-use survey	Data of 19,295 people from 9541 homes in 2009–2010, at 10-min interval	Methods: Markov Chain Monte Carlo (MCMC). Purpose: The profile of electricity demand in the residential sector [170]
Belgian time-use survey And Household Budget Survey	Data of 6400 respondents from 3474 households, 24-h profiles at 10-min interval	Methods: Hierarchical Clustering. Purpose: The occupancy profiles of households and its relation to socio-economic variables [171]
Japan time-use survey	Data from 80,000 households in Japan, at 10-min interval	Methods: Markov Chain. Purpose: Stochastic model for occupancy profiles in residential buildings [172]
Japanese Survey on Time Use and Leisure Activities (STULA)	Occupant daily activities at 15- min interval	Methods: Probabilistic analysis. Purpose: Aggregated energy saving potential on daily basis [173]
Danish time-use survey	Data of 9640 individuals from 4679 households, at 10-min interval	Methods: Probabilistic analysis. Purpose: Representative electricity profile based on synthetic occupancy profiles for households [174].

3.6.1. *Mobility survey*

The United States Federal Highway Administration (FHWA) released the National Household Travel Survey (NHTS) data[1], which is the authoritative source on the travel behaviour of the American public. It is the only source of national data that allows one to analyse trends in personal and household travel. It includes daily non-commercial travel by all modes, including characteristics of the people traveling, their household, and their vehicles.

Another comprehensive data source that combines different personal travel patterns of residents is the Household Travel Survey (HTS) data [163] collected by Transport for New South Wales, Australia. The Victorian state government in Australia also conducted a similar survey called 'Victorian Integrated Survey of Travel and Activity (VISTA) survey' over four financial years from 2012 to 2016. This survey dataset covers a total of 18,152 households and 46, 562 people [164]. The participants selected randomly were asked to complete a travel diary on a single specified day as part of their tasks. The collected information includes all personal travel information - from walking the dog, through to interstate travel.

3.6.2. *Time use survey*

As one of the major approaches of occupant behaviour profile survey, time of use survey has been conducted by many countries as a part of national census. Time use survey aims to report statistics on the amount of time people spend on various activities on daily basis. Researchers have taken advantage of time use survey to support occupant behaviour research and building scale energy modelling. For details, see Table 4.

4. Urban-scale occupant and mobility behaviour

In this section, we provide an overview of specific application-wise modelling, methodologies and approaches.

4.1. *Occupant behaviour modelling*

Space utilisation and occupant behaviour analytics from different building systems data is very crucial as it can provide intelligent insights. To better understand the impact of occupant behaviour on building energy consumption at an urban scale, building occupancy at an unprecedented scale was studied based on massive and passively collected mobile phone data [80]. By integrating the simulated occupancy data with the urban building energy model, the estimated energy consumption showed a significant difference compared to the reference data from DOE. Based on the same data set, Jiang [79] examined semantic-enriched land uses surrounding destinations in individual's motifs to infer trip purposes.

Urban scale occupancy patterns of two buildings were derived from smartphone-based GPS data to understand urban mobility [7]. Daily and weekly occupancy profiles were extracted based on positional data from social network software [94]. The difference of energy consumption was analysed according to occupancy schedules derived from positional data and energy codes, showing that occupancy data has a significant influence on energy consumption. After studying the positional record data, a spatiotemporal dependency was demonstrated between human mobility and energy use in urban area [93].

Occupancy data has been used to study short term occupancy in commercial and residential buildings [175–178]. Space utilisation prediction has been investigated leveraging historical and current contextual data (i.e. occupancy, spatial, temporal [179]). Occupancy data also has been used for building energy management [180]. A recent research uses the WiFi technology to detect and predict building occupancy [181, 182]. A fusion-based approach that combines cross-source data related to CO2 concentration, temperature, and Wi-Fi signal indicator is used for building occupancy analysis [183].

Another research uses occupancy and occupant activity data to investigate different patterns in hospital patient rooms [184] while researchers in Ref. [185] utilised building layout and occupant activity data to devise associations between building layout, physical activity and sitting time in an office setting. Ren et al. [67,186]. analysed how people use WiFi to access the Web in indoor retail spaces while navigating through a mall. Using only WiFi logs, Ren et al. also found strong correlations between behaviours and user demography [187]. Further, it is also shown that the online and movement patterns can be used to predict the visiting intent of building occupants [188].

Researchers also utilise survey data to investigate building occupancy profiles at homes and support building or community level energy modelling. Some research [169,171] investigated the time-use survey data and generated occupancy profiles for households. Barbour [80]

[1] https://nhts.ornl.gov/assets/2017UsersGuide.pdf.

developed a method for estimating building occupancy at urban-scale by extending the TimeGeo framework [189]. It classified all the sample urban buildings into three categories: residential, commercial, and industrial. Jiang [79] identified stays and pass by from each user's trajectory by setting the roaming distance as 300 m and setting temporal constraint as 10 min. A grid-based clustering method which is over the k-means algorithm and the density-based OPTICS (ordering points to identify the clustering structure) clustering algorithm [190] was used to cluster stay-points to get stay-regions. Jiang's work showed the process of translating large urban mobile phone traces into trip chains, activity sequences, and travel paths which will be helpful for urban planning. When extracting occupancy patterns from positional data collected by social network software, max normalization and k-means clustering method was used [94]. Discrepancy scores (DS) was computed to quantify the degree of divergences between DOE reference occupancy pattern and empirical hourly occupation rates [7].

Occupancy in two residential homes was monitored using the coarse-grained data produced by commodity smart meters, which record a home's electricity usage anywhere between one to 15 min [191]. Occupants' smartphone data were used as the ground truth occupancy, then a threshold-based Non-Intrusive Occupancy Monitoring (NIOM) Algorithm was developed to analyse the electricity data from smart meters for obtaining occupancy information in two homes.

Researchers also utilised data from time-use survey to acquire building occupancy information. This enables detailed occupant behaviour modelling. One of the approaches is clustering-based methods, to extract representative and determinant occupancy or energy use profiles among massive samples [165,171]. Clustering methods are most likely top-down approaches and commonly associated with statistic or probabilistic analysis [173,174].

Another approach is Markov-process based methods, which are usually used for dynamic and stochastic modelling of occupancy and appliance energy use. Markov-process based methods are designed to explore change of state by time shift, and is very efficient for stochastic occupant behaviour modelling The most popular algorithm is Markov Chain Monte Carlo (MCMC), and is widely applied in many researches for bottom-up modelling [166–169,170,172]

4.2. Understanding mobility patterns

The study of mobility patterns is to understand and analyse the movements of occupants in spatio-temporal resolution. The unprecedented scale and resolution of mobility data have generated impactful application for intra-urban mobility analysis [192]. The functionality of a spatial region can be investigated with an assist of human mobility inferred from shared bicycle system or taxi trajectories [193–195]. Recently, researchers in the field of energy simulation in the built environment started to integrate energy consumption and mobility analysis. Mohammadi and Taylor [93] used a full year of individual footprints collected from Twitter and developed a multivariate autoregressive model in reduced principal component analysis space. Their work, incorporating the spatiotemporal energy use fluctuations of urban population activities, created more reliable predictions of demand in a major U.S. city. Dong et al. [7] developed occupant behaviour model based on mobile position data for 456 buildings in San Antonio, Texas, U.S. and compares it with commonly used occupant schedules from U.S. Department of Energy Commercial Building Reference Model. In Ref. [113], the authors studied mobility patterns from the mobile network and indoor and outdoor traffic data.

Existing mobility studies primarily use mathematical frameworks and physical models to capture human mobility patterns. Brockmann et al. [89] proposed to extrapolate the universality of human travel patterns as a random-walk process with the displacements (i.e., the consecutive steps of movements) following a power-law distribution while the stay duration at each location was observed with exponential decay. Confirming the validity of the models, Gonzalez et al. [95] showed human mobility has a high degree of temporal and spatial regularity which can be captured by a unified spatial probability model. Rhee et al. [196] developed a truncated Levy walk mobility (TLW) model based on human mobility traces collected from five cities worldwide. Song et al. [96] developed the Exploration and Preferential Return (EPR) model to describe the ultra-diffusion of human mobility.

Other models can capture the macro-level mobility patterns based on individual patterns, forming critical foundations in linking human mobility to energy usage in urban areas. Isaacman et al. [197] developed a framework to forecast individuals' visitation destinations and found that individuals tend to choose popular destinations from overall populations over their most frequently visited locations. Jiang et al. [189] formulated the TimeGeo model, which integrates location information (home/work/other) for each visit. It also incorporates the regional land use data to estimate collective daily mobility patterns and achieved high accuracy. Pappalardo et al. [198] introduced the d-EPR model that infuses the gravity model into the original EPR model, extending the model to monitor the large-scale spatial mobility flux. Hoteit et al. [199] used location data from smartphones and tested multiple trajectory interpolation methods that have been previously proposed on subsampled user paths. The radiation model [200] can accurately estimate the interurban commuting fluxes compared to the gravity model for it considers the population in between both locations and resembles the process of radiation in physics. Recently, Mazzoli et al. [201] compared both models using Twitter data. It was found that the gravity-based model outperformed radiation-based models in terms of both accuracies on the flow and angles (with R2 greater than 0.95 for both).

In [202], the authors presented a graph-based optimization algorithm to design patrol route using large-scale mobility information from Foursquare. Also, graph-based approaches enabled modelling the parking space locations and the connections between them to manage the car parking violation in Ref. [54]. More recently, in Refs. [113], the authors proposed a semi-absorbing urban mobility model, which is designed to capture the structural impacts on urban mobility pattern in 3-D space.

4.3. Critical gap review and recommendations

Significant challenges exist despite advances in modelling occupant movement and behaviours in buildings. Lack of adequate dataset to represent diverse occupants, at various social-demographics, various building types, various locations, is a big gap. Models that capture the diversity of occupants are still rare as most of them represent aggregated occupant behaviours. Fit-for-purpose occupant modelling [203] is the direction to go, which can help address the tradeoff between model complexity and value. Presenting and interpreting the stochastic nature of results from the use of stochastic occupant models can be a challenge also. An occupant modelling guide, with available dataset, models and simulation tools, is needed to promote industry adoption of occupant modelling to support building design and operations.

In urban scale occupant behaviour modelling based on crowdsourced data, there are both theoretical and practical gaps which call for linking human mobility and building occupancy. Theoretically, despite that intra-urban social mobility could be well captured and modelled to predict urban scale energy demand in the cities, there is still a lack of understanding of how both the occupant behaviour and the temporal shifts of building energy performance are associated with the recurrent human mobility. Emerging studies start to utilise mobility data as a proxy of occupancy in buildings, yet the detailed occupant behaviour and building facility heterogeneity have not been quantified and incorporated into the models. The transformative next step is to integrate dynamic modelling, information on occupant behaviour, and more complete data sets.

In practice, when the aforementioned human mobility models are applied with the real-world data, potential issues rise. For example, the scarcity of fine-grained human mobility data and uncertainty from representativeness of the populations are observed and discussed [93]. Also, large-scale models demand high scalability. Hence, it is critical to devise a framework to identify the proper urban scale model based on the quantity and the quality of the data for optimal spatiotemporal modelling. In addition, human mobility data usually comes from crowdsourcing platforms which are often with biases in terms of both quantity and quality across different populations. Therefore, it is critical to developing new mitigation strategies to quantify and reduce the risk caused by such inherent sociodemographic bias.

5. Urban-scale human-building interactions, environmental comfort, and energy use

Urban big data have been applied in several areas including building retrofitting studies, modelling of human-building interactions, environmental comfort inferences, and energy usage at an urban scale. This section provides an overview of previous studies in those areas.

5.1. Understanding human-building interactions through connected thermostats

The availability of connected thermostats provides large-scale datasets that enable studying and predicting occupant-building interactions at the urban or regional scales. Huchuck et al., 2018 [204] investigated how users' comfort decisions and thermostat usage patterns are affected by exterior stimuli such as climate regions, seasonal patterns, and utility rates. In a different study, Huchuk et al., 2019 [138] used the same dataset to predict future motion states in homes and explored the prediction quality versus seasonal effects, time of day, and occupancy behavioural types. Another study [142] used analysed HVAC cycling data and weather data and derived metrics to indicate 1) which households are most likely to have extraordinarily high or low thermostat setpoints, 2) which households are most likely to have thermostat setpoints that change at fixed times each day, and 3) which households could benefit from a change in the hysteresis setting of a thermostat. Gunay et al., 2018 [136] used BMS data from groups of government office buildings to predict the frequency of temperature setpoint change requests. Models were developed to identify indoor temperatures that minimize the frequency of thermal complaints or thermostat overrides. With regards to modelling techniques, Huchuck et al., 2018 [204] used statistical methods such as linear regression to investigate the relationship between outdoor temperature and system interactions (i.e. thermostat usage). Other clustering techniques were also used to identify user types based on how they operate their thermostats. In a different study, Huchuk et al., 2018 [138] used machine learning models ranging from logistic regression, random forest, Markov model, hidden Markov model (HMM), and recurrent neural networks (RNN) to evaluate their accuracy in predicting occupancy patterns. Using a set of candidate features consisting of previous motion states, time of day, and weekday vs. weekend status, they found that a simple random forest machine learning model outperforms all other models. Gunay et al., 2018 [136] used multivariate logistic regression models to predict the likelihood of observing a setpoint decrease and increase in the next 30 min in a thermal zone based on the indoor and outdoor temperature.

5.2. Environmental comfort inference

Human perception deals with multi-physical stimuli and thus a comprehensive assessment leans on thermal, air quality, acoustic, and visual investigation. Moving from experimental data collection and surveys campaigns, correlations among built environment and citizens perception have been highlighted and implemented for environmental comfort modelling in space and time [205]. Environmental data collection in cities is mainly adopted to map real-time urban status in terms of air quality, noise pollution, and citizens' thermal stress. In order to get prediction of parameters distribution even under alternative scenarios, i.e. urban space renovation, or future weather forecast in the main framework of climate change, researchers rely upon models. Microclimate models reproduce the thermal performance of the built environment.

Two main physics-based approaches can be distinguished:(i) energy balance-based scheme, such as Town Energy Budget (TEB) [206] and Urban Canopy Models (UCM), both of which solve for the energy budget of the urban canopy layer [207] thanks to urban surface parametrization [208], and (ii) CFD-based models. The introduction of human physiology in microclimate models allow to compute human comfort by means of some of the most widely diffused thermal-comfort indicators such as PMV (Predicted Mean Vote) and PET (Physiological Equivalent Temperature). This possibility paved the way to a synergistic approach combining microclimate modelling, in-field surveying, and outdoor comfort investigations in a compelling method for assessing urban development effect on temperature distribution, wind flows, and consequently, pedestrians' thermal comfort [209,210]. Concerning acoustic comfort in cities, two main numerical approaches can be used: purely physical and soundscape models. The former, focuses on the effect of urban geometry on sound pressure levels and distribution [211]. In this context, full-wave-finite-difference time-domain method (FDTD) represents one of the most widely applied computational methods [212]. It can be used to explore sound quality by coupling information from emission and propagation models to geo-references data [213].

Soundscape models, on the contrary, integrate human perception and sound pressure modelling in search for interrelationships among spatio-temporal patterns in soundscape, acoustic and urban morphological indicators. In this approach, physical acoustic and subjective data are measured, morphological indicators from specific geographic information system (GIS) are analysed, and the most important relationships among them are explored [214]. Through this approach, researchers point to establish links between human senses, human perception and optimized urban design, through global and local spatial regression models [215], and ultimately artificial neural networks [216].

Energy saving potentials and occupants well-being, health, mood, and productivity are strongly influenced by daylight availability into buildings and thus a careful design of city shape could improve visual comfort in indoors and simultaneously reduce electrical energy consumption [217,218]. On the other hand, modelling of artificial lighting generally leans on different software such as DIALux [219] and Relux [220] which are used for indoors but even street lighting design. Multi-objective evolutionary algorithms can be implemented in order to account for both illuminance uniformity (safe and comfortable night vision) and energy efficiency or installation costs as Hammad and Akbarnexhad did in [221]. In [222], the authors revealed that multi-nodal thermal regulation model is potential to address different features for outdoor environment.

5.3. Energy use study

The study of energy usage is to infer and understand various factors that influence various energy usage patterns. The data-driven investigations of energy usage and associated factors can aid better building design as well as energy-efficient urban planning.

5.3.1. Urban energy modelling for buildings and districts

Existing UBEMs in the literature are generally dominated by two approaches: top-down and bottom-up. Top-down (i.e., data-driven) models are based on macro-economic modelling principles and techniques are intended to include important economic variables and statistical information. Such models tend to be used to investigate the interrelationships between the energy sector and the economy at large. They are also used to analyse energy use of building stock and identify energy conservation measures for retrofits.

A top-down method was concluded which treated the residential buildings energy consumption as an energy sink rather than individual end-users [223]. The common variables for the top-down models were enumerated, including macroeconomic indicators, gross domestic product (GDP), employment rates, price indices, climatic conditions, housing construction/demolition rates, and estimates of appliance ownership in the residential sector. This approach is used to provide long-term future predictions in the absence of energy supply, pricing shocks, and technological breakthroughs [224]. [225] used topic modelling, specifically latent Dirichlet allocation, to discover the thematic structure and spatial-temporal patterns of building renovation and adaptive reuse in cities. Because of its lack of energy consumption breakdowns by end-users, top-down models are not suitable to determine key parameters for energy use reductions, such as identifying the effectiveness of implementing an energy-saving measure [226].

In contrast, the bottom-up approach (i.e., Model-based building performance simulation) aims to group buildings with similar characteristics, such as building main use, structural properties, and construction geometries, into one category known as an archetype. It traditionally includes thermal and energy modelling, lighting and daylighting modelling, acoustic modelling, and indoor air quality modelling. These models are often treated as a method to identify the most cost-effective options to achieve given carbon reduction [227,228] or comfort [229] targets among available technologies and processes. Multiple models have been developed based on collected building energy consumption data [230] and occupancy profile [80,231,232] via different mathematical and statistical approaches, including Integrated Nested Laplace Approximation (INLA) [229], Automatic Building Energy Model (AutoBEM) [227], and deep learning [233–235]. Modelling of each functionality can use different approaches at various levels of details to capture the physics in buildings. For example, building thermal/energy performance can be modelled physically (white-box models) considering the detailed and dynamic heat and mass transfer in all inter-connected zones of a building (e.g., tools like EnergyPlus, ESP-r, and DeST). It can also be done with reduced-order (gray-box) RC models like ISO 13790. Kavgic et al. [226] created a detailed comparison between the two approaches and concluded that although the bottom-up approach was widely used in building energy models to study the impact of different combinations of input data and energy-saving measures, there were several limitations associated with those models. The most common challenge of this approach is the lack of publicly available input data due to privacy issues.

5.3.2. Urban building energy modelling platforms

Urban building energy modelling (UBEM) platform can provide key support for energy management of city-scale buildings during planning, design and operation stages. Popular platforms include SUNtool [236], CitySim [237], UMI [238], CityBES [239], TEASER [240], and HUES [241]. These platforms often apply a bottom-up approach for energy modelling at the urban scale. UBEM tools like CityBES are designed for urban buildings modelling, analysis and visualization. It uses an international open data standard, CityGML, to represent and exchange 3D city models. CityBES employs EnergyPlus to simulate building energy use and savings from energy-efficient retrofits. It can be used for urban building energy benchmarking, energy retrofit analysis, renewable energy analysis, building performance visualization, as well as urban climate data analysis and visualization.

UBEM platforms also rely on inputs of occupant schedules. There are three major modelling approaches of occupant behaviours in UBEM: static schedules, stochastic generation of schedules and stochastic/probabilistic models. Static schedules rely on pre-defined temporal schedules of occupants, lighting, plug-loads and HVAC system operation, and are used in CityBES and UMI [238,239]. Stochastic schedule generation, which is used by HUES and TEASER, combines Richardson's model [242] to generate schedules for occupant presence and other behaviours [240,243]. These schedules are only related to time, building type and occupant number. The stochastic model, however, considers the occupant presence model from Page et al. [244]. In the stochastic model, the probability of occupant's presence can be calculated based on the Markov Chains and their previous status. The present states of lighting, window, blind and appliances rely on the probability of corresponding actions. This model is applied in SUNtool and CitySim [236,237,245,246]. Also, different UBEM platforms use different building geometry model definitions, rely on different thermal energy simulation engines, and expect different model outputs and applications. Table 5 offers an overview of occupant behaviour modelling approaches in UBEMs.

Table 5
Overview of occupant behaviour modelling approaches in urban-scale building energy simulation platforms.

Platform	City model generation	Occupant behaviour rules	Thermal energy simulation engine	Outputs
SUNtool	Based on XML input files defined by users through GUI	Stochastic models	Gray-box model	Operational building energy use
CitySim	Based on XML input files defined by users through GUI	Both deterministic schedules and stochastic rules	Custom R–C model	Operational building energy use
UMI	Defined by users in Rhinoceros 3D environment	Deterministic schedules	EnergyPlus	Operational building energy use profiles, daylighting, outdoor comfort, and walkability analysis
CityBES	Automatic genera- tion from CityGML /GeoJson	Deterministic schedules	EnergyPlus	Operational building energy use, retrofit analysis
HUES	Reuse of existing models, or defined by user-self	Stochastic generation of occupancy and appliance use schedules	EnergyPlus	Operational building energy use, optimization results of buildings and energy system
TEASER	Automatic generation from CityGML	Both deterministic schedules and stochastic generation of schedules	ROMs in Modelica	Operational building energy use

5.4. Critical gap review and recommendations

Modelling human-building interactions at the urban-scale have been largely enabled due to large-scale datasets originating from various building systems. However, access to these datasets is typically restricted, especially if they are obtained through building automation systems. In many studies, the analysis of human-building interactions only focused on a handful of buildings in which researchers were able to gain permissions for retrieving data from the BMS. Although smart and connected thermostats can theoretically overcome this issue since data is stored centrally or in cloud platforms, many smart thermostat companies (e.g., Google Nest) do not provide access to their data, while those that provide them only cover residential buildings, thus no information is available for other building types. Another limitation when using such data to investigate human-building interactions at the urban-scale is that despite their relatively broad coverage of thousands of

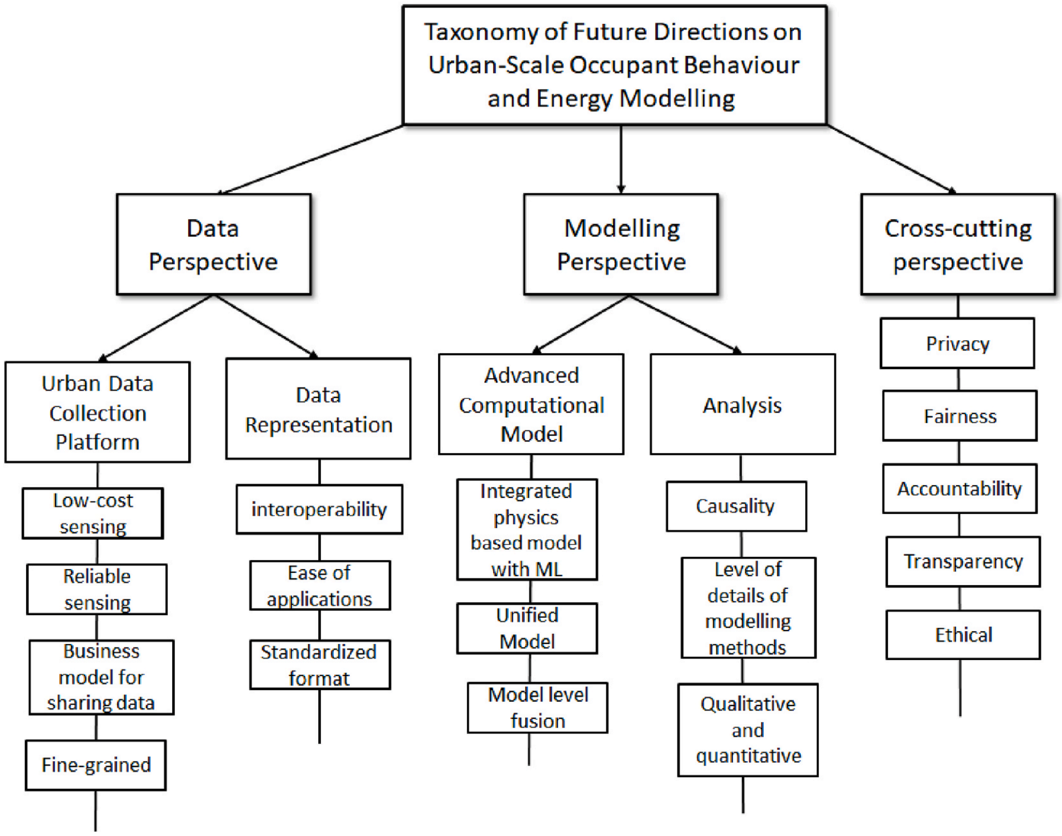

Fig. 4. Taxonomy of future research directions on urban-scale occupant behaviour and energy modelling.

buildings, they are not randomly sampled and represents a specific segment of consumers (e.g., early technology adopters). Therefore, one of the main challenges is to generalise results to draw conclusions on overall trends of human-building interactions at the urban-scale. One solution to address this challenge is to validate findings with different datasets that are randomly sampled to be statistically representative, such as time of use surveys although they do not typically have the same level of granularity. Concerning environmental comfort inference on citizens' well-being and energy consumption, interesting results have been achieved in perception modelling and mapping by taking into account each different comfort spheres singularly. These models are mainly physical-based model, e.g. accounting for urban energy balance or sound propagation laws, while human physiological and psychological filters to the physical data are still not critically addressed. A further effort of the scientific community should thus move forward a comprehensive assessment of human comfort perception simultaneously accounting for different comfort spheres and domains, i.e. physical, psychological, and physiological. In this perspective, new research is needed for novel models taking advantage of different data collection methods and thus, the type and amount of data needed.

6. Future directions on urban-scale occupant behaviour and energy modelling

This section provides discussions, perspectives and future directions based on the review. Specifically, privacy issues and data quality are discussed, and perspectives on new ways of multidisciplinary modelling are presented.

As cities' major energy and environmental goals cannot be met by a single sector, there is a need to integrate the modelling and analysis of multi-domain models including buildings, transportation, urban climate, and the electrical grid. Representing and capturing the interdependencies of these sectors (e.g., buildings release heat to the urban environment which leads to temperature raise; the urban temperature rise further influences energy demand in buildings, i.e., increasing cooling demand and reducing heating demand) are crucial to achieve optimization of planning, design and operation of urban energy and environment. Future developments of a computational framework are needed to facilitate the coupled simulation of multi-physics models at multi-scale, leveraging the emerging urban big data, exascale computing, and advances in artificial intelligence.

Major challenges in urban scale data include: (1) low-cost and reliable sensing technologies that can be deployed at the urban scale to continuously collect the data, (2) standardized data models and schema to represent the collected data for interoperability and ease of applications, (3) a business model to motivate and enable creation and sharing of the data, and (4) applications to demonstrate the application and values of the collected data.

Standardized data models and schema are particularly crucial for developing UBEMs to simulate energy use in neighbourhoods, districts and urban areas. Although some of the existing UBEM tools automate model generation using GIS-based building datasets. These datasets, if available, are typically provided in very different formats which requires significant manual consolidation efforts. Because these datasets also originate from different sources (e.g. tax assessment records, CityGML, or land use data), data quality is typically not consistent and requires manual checking and cleaning.

To deal with the challenges of collecting high-quality data from multiple urban areas, in particular high-quality labels or annotated data, approaches such as semi-supervised learning [247], or transfer learning [248] could be used to adapt data-driven models from one city to another. However, for these techniques, or similar black-box AI techniques, to be widely accepted in practice, not just in research domain, future works require models that enable feedback and collaboration across disciplines, incorporating expert views of the models. Thus, models need to be explainable. More efforts that enable neural network and deep learning models to be explainable is required, for models to be trustworthy and transparent, and relevant actions to be prescribed or recommended. This is why integrating physics-based models with machine-learning (or deep-learning) is a promising research direction in this field, requiring physics-guided design of ML or DL architectures [12] for applications in this field.

Finally, occupant behaviour is influenced by many factors including physical, biological, psychological, and social (the interaction among occupants). Unfortunately, there is not a single model considering all the aforementioned factors. Fundamentally, future computational models of occupant behaviour need to integrate with domain experts and modelling methods from building science, social science and psychology to simulate the root cause of occupant behaviour within the built environment. Future research also may include the study of causal relationship among different factors. Furthermore, modelling occupant behaviour from a single building to the urban scale is another challenge. A few research questions could be explored: 1) level of details (LoD) of the modelling methods: a fine-grained occupant behaviour for a single space or a building may not be applicable for urban scale. It worth to explore the model complexity v.s. accuracy of modelling occupant behaviour; 2) energy v.s. urban infrastructure: at a single building level, almost all studies of occupant behaviour are related to the energy consumption of the building. However, at an urban scale, modelling occupant behaviour is not only for the urban scale energy consumption but also urban infrastructures such as transportation or urban mobility services. Although the averaging effect of occupant behaviour at the urban scale may limit its influence on total energy use, their effect on spatial and temporal peak demand patterns can be significant.

Based on the above discussion, we categorise future research directions on urban-scale occupant behaviour and energy modelling into different categories and illustrate them in Fig. 4.

7. Conclusion

In this paper, we present a holistic multidisciplinary overview of the opportunities and challenges from occupant-centric urban data for modelling occupant and mobility behaviour and energy usage patterns at the urban scale. We started by reviewing concepts and definitions of occupant-centric urban data. We then review multiple datasets that have been used to capture human activities at various levels of details, from spatial data, occupancy data, building system and control data, crowdsourced sensor data, to mobility and survey data. Several categories of applications are reviewed, ranging from the analysis of building functionality to occupant and energy usage behaviour at building and urban scale, along with the range of approaches and methodologies that have been proposed, targeting the aforementioned applications. . Lastly, we discussed how different disciplines could collaborate on this challenging and complex problem of modelling occupant and energy behaviour at the urban scale, for the sustainable design, planning, and development of our future cities, and present several key problems and research directions for modelling with occupant-centric urban data.

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.

Acknowledgment

This paper is an output of IEA EBC Annex 79 SubTask 2. Flora Salim acknowledges the support of the Alexander von Humboldt Foundation and Bayer Foundation for her Humboldt-Bayer Fellowship, and also Australian Research Council Discovery DP190101485. Bing Dong would like to thank the support from the U.S. National Science Foundation CAREER Award (Award No. 1949372). Tianzhen Hong's work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the United States Department of Energy (Contract No. DE-AC02-05CH11231). Fisayo Caleb Sangogboye and Mikkel Baun Kjærgaard would like to acknowledge funding by Danish Energy Agency EUDP (Grant, n. 64018–0558). Da Yan acknowledges the support of National Natural Science Foundation of China, grant number 51778321.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.buildenv.2020.106964.

Appendix A. Steps for Urban big data processing for data driven modelling

Appendix A.1. Data Acquisition

Data acquisition from open data released by governments, voluntary-generated (e.g. crowdsourced), or system-generated (e.g. IoT, sensor data loggers) sources is the very first step in data-driven modelling. Since data acquisition is generally a pipeline of data crawlers and loggers collecting data from open data and APIs. Since occupant-centric urban data can capture activities from a wide variety of individual users and organisations occupying the built environment, often there is no schema that has been designed to fit the purpose of the data collection. Various platforms have been designed to acquire data from different sources. For example, one of the European FP7 projects, SOCIETIES (Self Orchestrating CommunIty ambiEnT IntelligEnce Spaces) design and develop platform for pervasive communities to collect urban mobility data [249]. The nature and types of the urban data, which are potential for implementing data-driven approaches for occupant behaviours and energy usage modelling will be discussed further in Section 3.

Appendix A.2. Data Preprocessing and Representation

Urban data collected from various platforms come in different structure and format. Hence, data preprocessing or cleaning is required in data-driven modelling. For example, occupant-centric building and energy data are usually noisy, which requires special attention. Data preprocessing involves data curation and cleaning of noisy data. This may involve imputation of missing data as well. Sometimes data fusion techniques are applied to combine the data from heterogeneous sources [250].

As discussed, urban data has different structure and format. Since there is often no established schema before the need for urban big data collection arises, therefore, various methods are applied to represent data in a common format.

One popular approach to represent aggregated data from multiple sources is using matrices or tensors. Matrix-based approaches such as collaborative filtering and latent matrix factorization are popular in modelling the space occupancy behaviours.

Another popular representation approach is graphs or networks. In a real-world scenario, many applications regarding space occupancy behaviours can be represented with graph-based approaches. A graph is used to characterise the interaction between objects which consists of a set of nodes and edges [251]. The nodes generally represent some data points, and the edges connect the nodes based on the relationship between the nodes.

Appendix A.3. Feature Extraction and Engineering

Once, data is processed for implementation, and feature engineering is required to develop a predictive model. Feature engineering involves extracting useful features from different raw data in the data mining task. It determines the important factors in machine learning model for a specific task [252]. The extracted features represent either spatial or temporal or spatio-temporal characteristics. Sometimes, feature normalization, feature selection may be necessary before applying the extracted features from urban data in predictive models.

With deep neural networks, this step is often skipped completely, with raw data that have been represented as matrices or graphs, as discussed in the previous section, channelled directly into deep learning models.

Appendix A.4. Analysis and Modelling

Analysis and modelling is a decider step in data-driven modelling. Predictive models are used to decide the occupant behaviours or energy usage in building or city level. A few machine learning methods have been applied in analysing and modelling big data. It includes clustering, classification, regression and deep learning-based models.

Clustering is used in identifying the patterns of energy usage in building [253,254]. Some popular classification algorithms are Random Forest [255], Support Vector Machine [256], Logistic Regression [257], Decision Tree [258], which have been used in predictive modelling of building energy and occupant behaviour.

Regression is another predictive modelling approach, which is used for predicting numerical outcomes. Each model is trained with a set of features to predict the numerical output value. Some popular regression algorithms are Linear Regression [259], Negative Binomial [260] etc. Energy usage density is generally a numerical value. Therefore, regression-based approaches have been widely used to predict future energy consumption [261].

Table A.6
Overview of different deep learning neural networks

Types of DNN	Main Usage
Auto-Encoder (AE) [262]	Feature extraction and de-noise
Convolutional Neural Network (CNN) [263]	Image-processing, object recognition and tracking
Recurrent Neural Network (RNN) [264]	Time-series prediction
Graph Neural Network (GNN) [265]	Traffic, optimization problem and biology
Generative adversarial network (GAN) [266]	Data augment, video games and advertisement
Reinforcement Learning [267]	Navigation, game play and control system

Recently, deep learning has attracted extensive attention in predicting building energy consumption [233,268]. However, the different deep learning architectures listed in Table A.6 can be used for various purposes.

Appendix A.5. Model Interpretation

Data-driven models need to be interpreted to help and improve decision making processes in specific application areas. Several papers [269,270] refers to this step as to reach actionable knowledge as the goal of data ingestion. The interpretability of the models, however, varies between different machine learning models. Deep learning models are least interpretable. Statistical models like Naive Bayes or decision-tree are more interpretable.

Several techniques are used to extract insights from outputs of data-driven models, aside from visualization. Correlation analysis across features can be used to evaluate the most dominant variables influencing the performance of the models. The scoring and ranking of different contributing variables can also be derived from ablation study and statistical test, such as the paired *t*-test. Sensitivity analysis can be used to check the robustness of the models against the input data or sample size. Causality analysis can be used to check whether a variable affects another independent variable. This, however, needs to be done with rigorous control experiments and causality testing, and clear hypothesis and metrics to be established.

Appendix B. Table of Terminologies

Table B.7
Table of Terminologies

Abbreviation	Definitions
AI	Artificial Intelligence
API	Application Programming Interface
BMS	Building management system
CBECS	Commercial Building Energy Consumption Survey
CDR	Call Detail Records
CFD	Computational Fluid Dynamics
CityBES	City Building Energy Saver
DL	Deep Learning
DNN	Deep learning neural network. List of different DNN in Table A.6
EPR	Exploration and Preferential Return

(*continued on next page*)

Table B.7 (*continued*)

Abbreviation	Definitions
FATE	Fairness, Accountability, Transparency, and Ethics in AI
GIS	Geographic information system
HMM	Hidden Markov model
HUES	Holistic Urban Energy Simulation Platform for Effective Model Integration
IoT	Internet of Things
LBS	Location-Based Service data
ML	Machine Learning
NHTS	National Household Travel Survey
NPTS	Nationwide Personal Transportation Survey
POI	Point of Interest
RC	First-order Resistance-Capacitance model
ROMs	Reduced Order Models
SUNtool	New modelling paradigm for simulating and optimising urban sustainability
TEASER	Open tool for urban energy modelling of building stocks
TimeGeo	Modelling framework for urban mobility
UBEM	Urban Scale Building Energy Modelling
UCM	Urban Canopy Model
UMI	Urban Modelling Interface

References

[1] T. Hong, D. Yan, S. D'Oca, C.-f. Chen, Ten questions concerning occupant behavior in buildings: the big picture, Build. Environ. 114 (2017) 518–530.
[2] S. D'Oca, T. Hong, J. Langevin, The human dimensions of energy use in buildings: a review, Renew. Sustain. Energy Rev. 81 (2018) 731–742.
[3] T. Hong, S. Taylor-Lange, S. D'Oca, D. Yan, S. Corgnati, Advances in research and applications of energy-related occupant behavior in buildings, Energy Build. 116 (2016) 694–702.
[4] Y. Zheng, L. Capra, O. Wolfson, H. Yang, Urban computing: concepts, methodologies, and applications, ACM Trans. Intell. Syst. Technol. 5 (3) (2014), 38:1-38:55.
[5] Y. Zheng, Trajectory data mining: an overview, ACM Trans. Intell. Syst. Technol. 6 (3) (2015) 1–29, 41.
[6] F. Salim, U. Haque, Urban computing in the wild: a survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and internet of things, Int. J. Hum. Comput. Stud. 81 (2015) 31–48.
[7] B. Dong, W. Wu, Q. Wang, S. Abdelmutaal, V. Prakash, Derive Urban Scale Occupant Behavior Profiles from Mobile Position Data : A Pilot Study, Proceeding of IBPSA 2019 Conference, 2019.
[8] T. Hong, Y. Chen, X. Luo, N. Luo, S. H. Lee, Ten Questions on Urban Building Energy Modeling, Building and Environment. 168. https://www.sciencedirect.com/science/article/abs/pii/S0360132319307206.
[9] R. Kitchin, The real-time city? big data and smart urbanism, Geojournal 79 (1) (2014) 1–14, 47.
[10] F. Calabrese, L. Ferrari, V.D. Blondel, Urban sensing using mobile phone network data: a survey of research, ACM Comput. Surv. 47 (2) (2014) 1–25, 20.
[11] D.P. Zhou, Q. Hu, Q, C.J. Tomlin, Quantitative comparison of data-driven and physics-based models for commercial building HVAC systems, in: American Control Conference, ACC, 2017, pp. 2900–2906, 2017.
[12] J. Willard, X. Jia, S. Xu, M. Steinbach, V. Kumar, Integrating physics-based modeling with machine learning: a survey, arXiv, 2003, https://arxiv.org/pdf/2003.04919.pdf, 2020.
[13] C.F. Reinhart, C.C. Davila, Urban building energy modeling – a review of a nascent field, Build. Environ. 97 (2016) 196–202, https://doi.org/10.1016/j.buildenv.2015.12.001. http://www.sciencedirect.com/science/article/pii/S0360132315003248.
[14] G. Happle, J.A. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, Energy Build. 174 (2018) 276–292, https://doi.org/10.1016/j.enbuild.2018.06.030. http://www.sciencedirect.com/science/article/pii/S0378778817333583.
[15] U. Eicker, M. Zirak, N. Bartke, L.R. Rodr_guez, V. Coors, New 3d model based urban energy simulation for climate protection concepts, Energy Build. 163 (2018) 79–91, https://doi.org/10.1016/j.enbuild.2017.12.019. http://www.sciencedirect.com/science/article/pii/S0378778816319016.
[16] Y. Chen, T. Hong, X. Luo, B. Hooper, Development of city buildings dataset for urban building energy modeling, Energy Build. 183 (2019) 252–265.
[17] J.-M. Bahu, A. Koch, E. Kremers, S. Murshed, Towards a 3d spatial urban energy modelling approach, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences (1) (2013) 33–41.
[18] R. Kaden, T.H. Kolbe, City-wide total energy demand estimation of buildings using semantic 3d city models and statistical data. Proc. Of the 8th International 3D GeoInfo Conference, 2013, p. 48.
[19] K.J. Lomas, H. Eppel, Sensitivity analysis techniques for building thermal simulation programs, Energy Build. 19 (1) (1992) 21–44, https://doi.org/10.1016/0378-7788(92)90033-D. http://www.sciencedirect.com/science/article/pii/037877889290033D.
[20] S. de Wit, G. Augenbroe, Analysis of uncertainty in building design evaluations and its implications, Energy Build. 34 (9) (2002) 951–958, https://doi.org/10.1016/S0378-7788(02)00070-1, a View of Energy and Bilding Performance Simulation at the start of the third millennium, http://www.sciencedirect.com/science/article/pii/S0378778802000701.
[21] T.A. Reddy, I. Maor, C. Panjapornpon, Calibrating detailed building energy simulation programs with measured data|part ii: application to three case study office buildings (rp-1051), HVAC R Res. 13 (2) (2007) 243–265, https://doi.org/10.1080/10789669.2007.10390953.
[22] B. Eisenhower, Z. O'Neill, V.A. Fonoberov, I. Mezifc, Uncertainty and sensitivity decomposition of building energy models, Journal of Building Performance Simulation 5 (3) (2012) 171–184, https://doi.org/10.1080/19401493.2010.549964.
[23] Y. Heo, G. Augenbroe, R. Choudhary, Quantitative risk management for energy retrofit projects, Journal of Building Performance Simulation 6 (4) (2013) 257–268, https://doi.org/10.1080/19401493.2012.706388.
[24] Y. Heo, G. Augenbroe, D. Graziano, R.T. Muehleisen, L. Guzowski, Scalable methodology for large scale building energy improvement: relevance of calibration in model-based retrofit analysis, Build. Environ. 87 (2015) 342–350, https://doi.org/10.1016/j.buildenv.2014.12.016. http://www.sciencedirect.com/science/article/pii/S0360132314004272.
[25] Q. Li, L. Gu, G. Augenbroe, C.J. Wu, J. Brown, Calibration of dynamic building energy models with multiple responses using Bayesian inference and linear regression models, Energy Procedia vol. 78 (2015) 979–984, https://doi.org/10.1016/j.egypro.2015.11.037, 6th International Building Physics Conference, IBPC 2015, http://www.sciencedirect.com/science/article/pii/S1876610215017695.
[26] Q. Li, G. Augenbroe, J. Brown, Assessment of linear emulators in lightweight bayesian calibration of dynamic building energy models for parameter estimation and performance prediction, Energy Build. 124 (2016) 194–202, https://doi.org/10.1016/j.enbuild.2016.04.025. http://www.sciencedirect.com/science/article/pii/S0378778816302808.
[27] Y. Kang, M. Krarti, Bayesian-emulator based parameter identification for 50 calibrating energy models for existing buildings, Building Simulation 9 (4) (2016) 411–428, https://doi.org/10.1007/s12273-016-0291-6.
[28] J. Berger, H.R. Orlande, N. Mendes, S. Guernouti, Bayesian inference for estimating thermal properties of a historic building wall, Build. Environ. 106 (2016) 327–339, https://doi.org/10.1016/j.buildenv.2016.06.037. http://www.sciencedirect.com/science/article/pii/S036013231630244X.
[29] R. Kitchin, Big data, new epistemologies and paradigm shifts, Big data& society 1 (1) (2014), 2053951714528481.
[30] G. Tardioli, R. Kerrigan, M. Oates, O. James, D. Finn, Data driven approaches for prediction of building energy consumption at urban level, Energy Procedia 78 (2015) 3378–3383.
[31] C.E. Kontokosta, C. Tull, A data-driven predictive model of city-scale energy use in buildings, Appl. Energy 197 (2017) 303–317.
[32] C. Dwork, Differential privacy: a survey of results, in: M. Agrawal, D. Du, Z. Duan, A. Li (Eds.), Theory and Applications of Models of Computation, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 1–19.
[33] B.C. Fung, K. Wang, A.W.-C. Fu, S.Y. Philip, Introduction to Privacy Preserving Data Publishing: Concepts and Techniques, Chapman and Hall/CRC, 2010.
[34] K. Wang, R. Chen, B. Fung, P. Yu, Privacy-preserving Data Publishing: A Survey on Recent Developments, ACM Computing Surveys.
[35] R. Jia, F.C. Sangogboye, T. Hong, C. Spanos, M.B. Kaergaard, Pad: protecting anonymity in publishing building related datasets, in: Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments, ACM, 2017, p. 4.51.
[36] H. Lu, C.S. Jensen, M.L. Yiu, Pad: privacy-area aware, dummy-based location privacy in mobile services, in: Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access, ACM, 2008, pp. 16–23.

[37] P. Zhong, R. Lu, Pad: privacy-preserving data dissemination in mobile social networks, in: IEEE International Conference on Communication Systems, IEEE, 2014, pp. 243–247, 2014.
[38] F.C. Sangogboye, R. Jia, T. Hong, C. Spanos, M.B. Kjaergaard, A framework for privacy-preserving data publishing with enhanced utility for cyber-physical systems, ACM Trans. Sens. Netw. 14 (3–4) (2018) 30.
[39] M. Wernke, P. Skvortsov, F. Durr, K. Rothermel, A classification of location privacy attacks and approaches, Personal Ubiquitous Comput. 18 (1) (2014) 163–175.
[40] D. Manousakas, C. Mascolo, A.R. Beresford, D. Chan, N. Sharma, Quantifying privacy loss of human mobility graph topology, Proceedings on Privacy Enhancing Technologies (3) (2018) 5–21.
[41] Y.-A. De Montjoye, C.A. Hidalgo, M. Verleysen, V.D. Blondel, Uniquein the crowd: the privacy bounds of human mobility, Sci. Rep. 3 (2013) 1376.
[42] J. H. Schwee, F. C. Sangogboye, M. B. Kj_rgaard, Evaluating Practical Privacy Attacks for Building Data Anonymized by Standard Methods.
[43] F. Zambonelli, F. Salim, S.W. Loke, W. De Meuter, S. Kanhere, Algorithmic governance in smart cities: the conundrum and the potential of pervasive computing solutions, IEEE Technol. Soc. Mag. 37 (2) (2018) 80–87.
[44] D. Ensign, S. Neville, A. Paul, S. Venkatasubramanian, The complexity of explaining neural networks through (group) invariants, Theor. Comput. Sci. 808 (2020) 74–85.
[45] K. Crawford, M. Whittaker, M.C. Elish, S. Barocas, A. Plasek, K. Ferry man, The AI Now report: the social and economic implications of artificial intelligence technologies in the near-term. AI Now Public Symposium, Hosted by the White, House and New York University's Information Law Institute, 2016. July 7th.
[46] A.D. Selbst, D. Boyd, S.A. Friedler, S. Venkatasubramanian, J. Vertesi, Fairness and abstraction in sociotechnical systems, in: Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, pp. 59–68.
[47] M.T. Ribeiro, S. Singh, C. Guestrin, Why should i trust you?: explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 1135–1144.
[48] P. Wei, X. Jiang, Data-driven energy and population estimation for realtime city-wide energy footprinting, in: Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, ACM, 2019, pp. 267–276.
[49] M. Caliskan, A. Barthels, B. Scheuermann, M. Mauve, Predicting parking lot occupancy in vehicular ad hoc networks, in: IEEE 65th Vehicular Technology Conference, IEEE, 2007, pp. 277–281.
[50] W. Shao, Y. Zhang, B. Guo, K. Qin, J. Chan, F.D. Salim, Parking availability prediction with long short term memory model, in: International Conference on Green, Pervasive, and Cloud Computing, Springer, 2018, pp. 124–137, 53.
[51] E.I. Vlahogianni, K. Kepaptsoglou, V. Tsetsos, M.G. Karlaftis, A realtime parking prediction system for smart cities, Journal of Intelligent Transportation Systems 20 (2) (2016) 192–204.
[52] S. Liu, H. Guan, H. Yan, H. Yin, Unoccupied parking space prediction of chaotic time series, in: ICCTP 2010: Integrated Transportation Systems: Green, Intelligent, Reliable, 2010, pp. 2122–2131.
[53] F. Caicedo, F. Robuste, A. Lopez-Pita, Parking management and modellng of car park patron behavior in underground facilities, Transport. Res. Rec. 1956 (1) (2006) 60–67.
[54] W. Shao, F.D. Salim, T. Gu, N.-T. Dinh, J. Chan, Traveling officer problem: managing car parking violations efficiently using sensor data, IEEE Internet of Things Journal 5 (2) (2017) 802–810.
[55] K.S. Liu, J. Gao, X. Wu, S. Lin, On-street parking guidance with realtime sensing data for smart cities, in: 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), IEEE, 2018, pp. 1–9.
[56] W. Alajali, S. Wen, W. Zhou, On-street car parking prediction in smart city: a multi-source data analysis in sensor-cloud environment, in: International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage, Springer, 2017, pp. 641–652.
[57] O. Bulan, R.P. Loce, W. Wu, Y.R. Wang, E.A. Bernal, Z. Fan, Videobased real-time on-street parking occupancy detection system, J. Electron. Imag. 22 (4) (2013), 041109.
[58] Z. Guo, Residential street parking and car ownership: a study of households with off-street parking in the New York City region, J. Am. Plann. Assoc. 79 (1) (2013) 32–48.
[59] K.K. Qin, W. Shao, Y. Ren, J. Chan, F.D. Salim, Solving multiple travelling officers problem with population-based optimization algorithms, Neural Comput. Appl. (2019) 1–27.
[60] A.J. Ruiz-Ruiz, H. Blunck, T.S. Prentow, A. Stisen, M.B. Kjærgaard, Analysis methods for extracting knowledge from large-scale wifi monitoring to inform building facility planning, in: IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2014, pp. 130–138.
[61] A. Sevtsuk, Mapping the mit campus in real time using wifi, in: Handbook of Research on Urban Informatics: the Practice and Promise of the Real-Time City, IGI Global, 2009, pp. 326–338.
[62] F.C. Sangogboye, M.B. Kjaergaard, Scalable and accurate estimation of room-level people counts from multi-modal fusion of perimeter sensors and wifi trajectories, IEEE International Conference on Mobile Data Management (MDM), IEEE, 2019, pp. 37–42.
[63] F.C. Sangoboye, M.B. Kjaergaard, Plcount: a probabilistic fusion algorithm for accurately estimating occupancy from 3d camera counts, in: Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, ACM, 2016, pp. 147–156.
[64] K. Christensen, R. Melfi, B. Nordman, B. Rosenblum, R. Viera, Using existing network infrastructure to estimate building occupancy and control plugged-in devices in user workspaces, Int. J. Commun. Network. Distr. Syst. 12 (1) (2014) 4–29.
[65] A.K. Das, P.H. Pathak, J. Jee, C.-N. Chuah, P. Mohapatra, Non-intrusive multi-modal estimation of building occupancy, in: Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems, ACM, 2017, p. 14.55.
[66] L. Schauer, M. Werner, P. Marcus, Estimating crowd densities and pedestrian ows using wi-_ and bluetooth, in: Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, ICST, 2014, pp. 171–177.
[67] Y. Ren, M. Tomko, F.D. Salim, K. Ong, M. Sanderson, Analyzing web behavior in indoor retail spaces, Journal of the Association for Information Science and Technology 68 (1) (2017) 62–76.
[68] W. Shao, T. Nguyen, K. Qin, M. Youssef, F.D. Salim, Bledoorguard: a device-free person identification framework using bluetooth signals for door access, IEEE Internet of Things Journal 5 (6) (2018) 5227–5239.
[69] W. Shao, F.D. Salim, T. Nguyen, M. Youssef, Who opened the room? device-free person identification using bluetooth signals in door access, in: IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), IEEE, 2017, pp. 68–75, 2017.
[70] J. Cranshaw, E. Toch, J. Hong, A. Kittur, N. Sadeh, Bridging the gap between physical location and online social networks. Proceedings of the 12th ACM International Conference on Ubiquitous Computing, ACM, 2010, pp. 119–128.
[71] D. Hristova, M.J. Williams, M. Musolesi, P. Panzarasa, C. Mascolo, Measuring urban social diversity using interconnected geo-social networks, in: Proceedings of the 25th International Conference on World Wide Web, ternational World Wide Web Conferences Steering Committee, 2016, pp. 21–30.
[72] Y. Su, X. Li, W. Tang, J. Xiang, Y. He, Next check-in location prediction via footprints and friendship on location-based social networks, in: 2018 19th IEEE International Conference on Mobile Data Management (MDM), IEEE, 2018, pp. 251–256.
[73] D. Yang, D. Zhang, V.W. Zheng, Z. Yu, Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns, IEEE Transactions on Systems, Man, and Cybernetics: Systems 45 (1) (2014) 129–142.
[74] D. Yang, B. Qu, J. Yang, P. Cudre-Mauroux, Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach, in: The World Wide Web Conference, ACM, 2019, pp. 2147–2157.
[75] Y. Liu, C. Liu, X. Lu, M. Teng, H. Zhu, H. Xiong, Point-of-interest demand modeling with human mobility patterns, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2017, pp. 947–955.
[76] J. Cranshaw, R. Schwartz, J. Hong, N. Sadeh, The livehoods project: utilizing social media to understand the dynamics of a city, in: Sixth International AAAI Conference on Weblogs and Social Media, 2012.
[77] R. Wilson, E. zu Erbach-Schoenberg, M. Albert, D. Power, S. Tudge, M. Gonzalez, S. Guthrie, H. Chamberlain, C. Brooks, C. Hughes, et al., Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal earthquake, PLoS currents 8.
[78] S. Jiang, J. Ferreira, M.C. Gonzalez, Activity-based human mobility patterns inferred from mobile phone data: a case study of Singapore, IEEE. Transactions on Big Data 3 (2) (2017) 208–219.
[79] S. Jiang, G.A. Fiore, Y. Yang, J. Ferreira Jr., E. Frazzoli, M.C. Gonzalez, A review of urban computing for mobile phone traces: current methods, challenges and opportunities, in: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, ACM, 2013, pp. 1–9.
[80] E. Barbour, C.C. Davila, S. Gupta, C. Reinhart, J. Kaur, M.C. Gonzalez, Planning for sustainable cities by estimating building occupancy with mobile phones, Nat. Commun. 10 (1) (2019) 1–10.
[81] X. Wang, J. Liono, W. Mcintosh, F.D. Salim, Predicting the city foot traffic with pedestrian sensor data, in: Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, ACM, 2017, pp. 1–10.
[82] M.T. Doan, S. Rajasegarar, M. Salehi, M. Moshtaghi, C. Leckie, Profiling pedestrian distribution and anomaly detection in a dynamic environment, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, ACM, 2015, pp. 1827–1830.
[83] Smart, Optimizing Energy Consumption in Smart Homes, 2012. http://lass.cs.umass.edu/projects/smart/.
[84] N. Isaacs, M. Camilleri, L. Burrough, A. Pollard, K. Saville-Smith, R. Fraser, P. Rossouw, J. Jowett, Energy use in new zealand households: final report on the household energy end-use project (heep), BRANZ Study Report 221 (71) (2010) 15–21.
[85] J.Z. Kolter, M.J. Johnson, Redd: a public data set for energy disaggregation research, in: Workshop on Data Mining Applications in Sustainability (SIGKDD), vol. 25, 2011, pp. 59–62. San Diego, CA.
[86] J. Kelly, W. Knottenbelt, UK-DALE': A dataset recording UK Domestic Appliance-Level Electricity demand and whole-house demand, Uk-dale', arXiv preprint arXiv:1404.0284, 2014.
[87] O. Parson, G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, W. Knottenbelt, A. Rogers, Dataport, nilmtk, A building data set designed for non-intrusive load monitoring, in: IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2015, pp. 210–214. 58.
[88] B. Glasgo, C. Hendrickson, I.M. Azevedo, Using advanced metering infrastructure to characterize residential energy use, Electr. J. 30 (3) (2017) 64–70.

[89] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, Nature 439 (7075) (2006) 462.
[90] Y. Zheng, X. Xie, W.-Y. Ma, et al., Geolife: a collaborative social networking service among user, location and trajectory, IEEE Data Eng. Bull. 33 (2) (2010) 32–39.
[91] S.B. Mokhtar, A. Boutet, L. Bouzouina, P. Bonnel, O. Brette, L. Brunie, M. Cunche, S. D'Alu, V. Primault, P. Raveneau, et al., Priva'mov: Analysing Human Mobility through Multi-Sensor Datasets, 2017.
[92] J.K. Laurila, D. Gatica-Perez, I. Aad, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, M. Miettinen, et al., The mobile data challenge: big data for mobile computing research, Workshop on the Nokia Mobile Data Challenge, in: Proceedings of the Conjunction with the 10th International Conference on Pervasive Computing (2012) 1–8.
[93] N. Mohammadi, J.E. Taylor, Urban energy flux: spatiotemporal fluctuations of building energy consumption and human mobility-driven prediction, Appl. Energy 195 (2017) 810–818.
[94] X. Kang, D. Yan, H. Sun, Y. Jin, P. Xu, An Approach for Obtaining and Extracting Occupancy Patterns in Buildings Based on Mobile Positioning Data, Proceeding of IBPSA 2019 Conference, 2019.
[95] M.C. Gonzalez, C.A. Hidalgo, A.-L. Barabasi, Understanding individual human mobility patterns, Nature 453 (7196) (2008) 779.
[96] C. Song, T. Koren, P. Wang, A.-L. Barabasi, Modelling the scaling properties of human mobility, Nat. Phys. 6 (10) (2010) 818–823, https://doi.org/10.1038/nphys 59, doi:10/abs/nphys1760.html#supplementary-information.
[97] C. Song, Z. Qu, N. Blumm, A.-L. Barabasi, Limits of predictability in human mobility, Science 327 (5968) (2010) 1018–1021.
[98] C.M. Schneider, V. Belik, T. Couronn_e, Z. Smoreda, M.C. Gonzalez, Unravelling daily human mobility motifs, J. R. Soc. Interface 10 (84) (2013) 20130246.
[99] B. Wheatman, A. Noriega, A. Pentland, Electricity demand and population dynamics prediction from mobile phone metadata, in: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Springer, 2016, pp. 196–205.
[100] B.A. Huberman, Sociology of science: big data deserve a bigger audience, Nature 482 (7385) (2012) 308.
[101] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, C. Ratti, Geo-located twitter as proxy for global mobility patterns, Cartogr. Geogr. Inf. Sci. 41 (3) (2014) 260–271.
[102] R. Jurdak, K. Zhao, J. Liu, M. AbouJaoude, M. Cameron, D. Newth, Understanding human mobility from twitter, PloS One 10 (7) (2015), e0131469.
[103] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, C. Mascolo, A tale of many cities: universal patterns in human urban mobility, PloS One 7 (5) (2012) 1–10, https://doi.org/10.1371/journal.pone.0037027.
[104] Q. Wang, N.E. Phillips, M.L. Small, R.J. Sampson, Urban mobility and neighborhood isolation in America's 50 largest cities, Proc. Natl. Acad. Sci. 115 (30) (2018) 7735–7740.
[105] Q. Wang, J.E. Taylor, Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster, PLoS One 11 (1) (2016) 1–14.
[106] P. Barberá, G. Rivero, Understanding the political representativeness of twitter users, Soc. Sci. Comput. Rev. 33 (6) (2015) 712–729.
[107] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, J.N. Rosenquist, Understanding the demographics of twitter users, in: Fifth International AAAI Conference on Weblogs and Social Media, 2011.
[108] Y. Liu, C. Kliman-Silver, A. Mislove, The tweets they are a-changin: evolution of twitter users and behavior, in: Eighth International AAAI Conference on Weblogs and Social Media, 2014.
[109] D. Quercia, D. Saez, Mining urban deprivation from foursquare: implicit crowdsourcing of city land use, IEEE Pervasive Computing 13 (2) (2014) 30–36.
[110] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, C. Ratti, Real-time urban monitoring using cell phones: A case study in Rome, IEEE Trans. Intell. Transport. Syst. 12 (1) (2011) 141–151.
[111] J. Froehlich, J. Neumann, N. Oliver, Measuring the pulse of the city through shared bicycle programs, Proc. of UrbanSense08 (2008) 16–20.
[112] J.E. Froehlich, J. Neumann, N. Oliver, Sensing and predicting the pulse of the city through shared bicycling, in: Twenty-First International Joint Conference on Artificial Intelligence, 2009.
[113] Z. Zheng, F. Wang, D. Wang, L. Zhang, Buildings affect mobile patterns: developing a new urban mobility model, in: Proceedings of the 5th Conference on Systems for Built Environments, ACM, 2018, pp. 83–92.
[114] R. Giridharan, M. Kolokotroni, Urban heat island characteristics in London during winter, Sol. Energy 83 (9) (2009) 1668–1682, https://doi.org/10.1016/j.solener.2009.06.007.61.
[115] F. Dominguez, S. Dauwe, N.T. Cuong, D. Cariolaro, A. Touha, B. Dhoedt, D. Botteldooren, K. Steenhaut, Towards an environmental measurement cloud: delivering pollution awareness to the public, Int. J. Distributed Sens. Netw. 10 (3) (2014) 541360, https://doi.org/10.1155/2014/541360.
[116] S. Dhingra, R.B. Madda, A.H. Gandomi, R. Patan, M. Daneshmand, Internet of things mobile-air pollution monitoring system (IoT-Mobair), IEEE Internet of Things Journal 6 (3) (2019) 5577–5584, https://doi.org/10.1109/JIOT.2019.2903821.
[117] A. Sani, P. Beauty Ejiroghene, Smart framework for environmental pollution monitoring and control system using IoT-based technology, Sensors & Transducers 229 (1) (2019) 84–93.
[118] Z. Deng, D. Weng, J. Chen, R. Liu, Z. Wang, J. Bao, Y. Zheng, Y. Wu, AirVis: visual analytics of air pollution propagation, IEEE Trans. Visual. Comput. Graph. (2019), https://doi.org/10.1109/TVCG.2019.2934670.
[119] J. Huang, N. Duan, P. Ji, C. Ma, feng Hu, Y. Ding, Y. Yu, Q. Zhou, W. Sun, A crowdsource-based sensing system for monitoring fine-grained air quality in urban environments, IEEE Internet of Things Journal 6 (2) (2019) 3240–3247, https://doi.org/10.1109/JIOT.2018.2881240.
[120] T. Liu, Y. Zhu, Y. Yang, F. Ye, ALC2: when active learning meets compressive crowdsensing for urban air pollution monitoring, IEEE Internet of Things Journal (2019), https://doi.org/10.1109/jiot.2019.2939552.
[121] S. Xu, B. Zou, Y. Lin, X. Zhao, S. Li, C. Hu, Strategies of method selection for _ne-scale $PM_{2.5}$ mapping in an interurban area using crowdsourced monitoring, Atmospheric Measurement Techniques 12 (5) (2019) 2933–2948, https://doi.org/10.5194/amt-12-2933-2019, 62.
[122] I. Pigliautile, A.L. Pisello, Environmental data clustering analysis through wearable sensing techniques: New bottom-up process aimed to identify intra-urban granular morphologies from pedestrian transects, Build. Environ. 171 (2020) 106641, https://doi.org/10.1016/j.buildenv.2019.106641.
[123] A.S. Nouman, A. Chokhachian, D. Santucci, T. Auer, Prototyping of environmental kit for georeferenced transient outdoor comfort assessment, ISPRS Int. J. Geo-Inf. 8 (2) (2019) 76, https://doi.org/10.3390/ijgi8020076.
[124] A. Chokhachian, K. Ka-Lun Lau, K. Perini, T. Auer, Sensing transient outdoor comfort: a georeferenced method to monitor and map microclimate, Journal of Building Engineering 20 (2018) 94–104, https://doi.org/10.1016/J.JOBE.2018.07.003.
[125] J. Picaut, N. Fortin, E. Bocher, G. Petit, P. Aumond, G. Guillaume, An open-science crowdsourcing approach for producing community noise maps using smartphones, Build. Environ. 148 (2019) 20–33, https://doi.org/10.1016/J.BUILDENV.2018.10.049.
[126] J. P. Bello, C. Silva, O. Nov, R. L. DuBois, A. Arora, J. Salamon, C. Mydlarz, H. Doraiswamy, SONYC: A System for the Monitoring, Analysis and Mitigation of Urban Noise Pollution. arXiv:1805.00889.
[127] L.M. Aiello, R. Schifanella, D. Quercia, F. Aletta, Chatty maps: constructing sound maps of urban areas from social media data, Royal Society Open Science 3 (3) (2016) 150690, https://doi.org/10.1098/rsos.150690.
[128] K.K.-L. Lau, Y. Shi, E.Y.-Y. Ng, Dynamic response of pedestrian thermal comfort under outdoor transient conditions, Int. J. Biometeorol. 63 (7) (2019) 979–989, https://doi.org/10.1007/s00484-019-01712-2.
[129] S. Kariminia, S. Shamshirband, S. Motamedi, R. Hashim, C. Roy, A systematic extreme learning machine approach to analyze visitors' thermal comfort at a public urban space, Renew. Sustain. Energy Rev. 58 (2016) 751–760, https://doi.org/10.1016/j.rser.2015.12.321.
[130] I. Pigliautile, A. Pisello, A new wearable monitoring system for investigating pedestrians' environmental conditions: development of the experimental tool and start-up findings, Sci. Total Environ. 630 (2018), https://doi.org/10.1016/j.scitotenv.2018.02.208.
[131] D. Verma, A. Jana, K. Ramamritham, Classification and mapping of sound sources in local urban streets through AudioSet data and Bayesian optimized Neural Networks, Noise Mapp. 6 (1) (2019) 52–71, https://doi.org/10.1515/noise-2019-0005.
[132] M. Nakayoshi, M. Kanda, R. Shi, R. de Dear, Outdoor thermal physiology along human pathways: a study using a wearable measurement system, Int. J. Biometeorol. 59 (5) (2015) 503–515, https://doi.org/10.1007/s00484-014-0864-y.
[133] C.F. Reinhart, Lightswitch-2002: a model for manual and automated control of electric lighting and blinds, Sol. Energy 77 (1) (2004) 15–28, https://doi.org/10.1016/j.solener.2004.04.003. http://www.sciencedirect.com/science/article/pii/S0038092X04000702.
[134] F. Haldi, D. Robinson, Adaptive actions on shading devices in response to local visual stimuli, Journal of Building Performance Simulation 3 (2) (2010) 135–153, https://doi.org/10.1080/19401490903580759.
[135] H.B. Rijal, P. Tuohy, F. Nicol, M.A. Humphreys, A. Samuel, J. Clarke, Development of an adaptive window-opening algorithm to predict the thermal comfort, energy use and overheating in buildings, Journal of Building Performance Simulation 1 (1) (2008) 17–30, https://doi.org/10.1080/19401490701868448.
[136] H.B. Gunay, W. Shen, G. Newsham, A. Ashouri, Modelling and analysis of unsolicited temperature setpoint change requests in office buildings, Build. Environ. 133 (2018) 203–212, https://doi.org/10.1016/j.buildenv.2018.02.025. http://www.sciencedirect.com/science/article/pii/S0360132318300970.
[137] H.B. Gunay, M. Ouf, G. Newsham, W. O'Brien, Sensitivity analysis and optimization of building operations, Energy Build. 199 (2019) 164–175, https://doi.org/10.1016/j.enbuild.2019.06.048. http://www.sciencedirect.com/science/article/pii/S0378778818334479.
[138] B. Huchuk, S. Sanner, W. O'Brien, Comparison of Machine Learning Models for Occupancy Prediction in Residential Buildings Using Connected Thermostat Data, Build. Environ. vol. 160 (2019) 106177, https://doi.org/10.1016/j.buildenv.2019.106177. http://www.sciencedirect.com/science/article/pii/S0360132319303877.
[139] C. John, C. Vallianos, J. Candanedo, A. Athienitis, Estimating Time Constants for over 10,000 Residential Buildings in North America: towards a Statistical Characterization of Thermal Dynamics, 2018, pp. 1383–1388, https://doi.org/10.14305/ibpc.2018.ps17.
[140] C. Holcomb, Pecan street inc.: a test-bed for nilm, in: International Workshop on Non-intrusive Load Monitoring, Pittsburgh, PA, USA, 2012.
[141] R. Rajagopal Albert, Finding the right consumers for thermal demand response: an experimental evaluation, IEEE Transactions on Smart Grid 9 (2) (2018) 564–572, https://doi.org/10.1109/TSG.2016.2555985.

[142] S.J. Smullin, Thermostat metrics derived from hvac cycling data for targeted utility efficiency programs, Energy Build. 117 (2016) 176–184, https://doi.org/10.1016/j.enbuild.2016.02.018. http://www.sciencedirect.com/science/article/pii/1570 S0378778816300652.
[143] T. Ramos, S. Dedesko, J.A. Siegel, J.A. Gilbert, B. Stephens, Spatial and temporal variations in indoor environmental conditions, human occupancy, and operational characteristics in a new hospital building, PloS One 10 (3) (2015), e0118207.
[144] M. Karami, G.V. McMorrow, L. Wang, Continuous monitoring of indoor environmental quality using an arduino-based data acquisition system, Journal of Building Engineering 19 (2018) 412–419.
[145] A.S. Ali, Z. Zanzinger, D. Debose, B. Stephens, Open source building science sensors (osbss): a low-cost arduino-based platform for long-term indoor environmental data collection, Build. Environ. 100 (2016) 114–126, https://doi.org/10.1016/j.buildenv.2016.02.010.
[146] O. Irulegi, A. Serra, R. Hernandez, Data on records of indoor temperature and relative humidity in a university building, Data in brief 13 (2017) 248–252.
[147] ASHRAE, Ashrae global thermal comfort database ii (Last accessed, http://www.comfortdatabase.com/. (Accessed October 2019).
[148] L. Candanedo, Occupancy detection data set (Last accessed, https://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+. (Accessed October 2019).
[149] L. Wigle, How big data will make us more energy efficient. https://www.weforum.org/agenda/2014/05/big-data-will-make-us-energy-efficient/66, 2014.
[150] T. Babaei, H. Abdi, C.P. Lim, S. Nahavandi, A study and a directory of energy consumption data sets of buildings, Energy Build. 94 (2015) 91–99.
[151] P.S. Dataport, Pecan street dataport. https://dataport.pecanstreet.org, 2016.
[152] H. Deng, D. Fannon, M.J. Eckelman, Predictive modeling for us commercial building energy use: a comparison of existing statistical and machine learning algorithms using cbecs microdata, Energy Build. 163 (2018) 34–43.
[153] S. Karatasou, M. Santamouris, Socio-economic status and residential energy consumption: a latent variable approach., Energy and Buildings.
[154] Open street map. https://www.openstreetmap.org. (Accessed 11 October 2019).
[155] Google maps platforms. https://cloud.google.com/maps-platform/. (Accessed 11 October 2019).
[156] M. Haklay, P. Weber, Openstreetmap: user-generated street maps, IEEE Pervasive Computing 7 (4) (2008) 12–18.
[157] M.S. Rahaman, Y. Mei, M. Hamilton, F.D. Salim, CAPRA: a contourbased accessible path routing algorithm, Inf. Sci. 385 (2017) 157–173.
[158] M.S. Rahaman, M. Hamilton, F.D. Salim, Coact: a framework for context-aware trip planning using active transport, in: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE, 2018, pp. 645–650, 2018.
[159] Google Places, 2019, 10-11. 67, https://developers.google.com/places/.
[160] A. Prabowo, P. Koniusz, W. Shao, F.D. Salim, Coltrane: convolutional trajectory network for deep map inference, in: Proceedings of the 6th ACM International Conference on Systems for Energy Efficient Buildings,Cities, and Transportation, ACM, New York, NY, USA, 2019, pp. 21–30.
[161] Microsoft, US building footprints. https://github.com/microsoft/USBuildingFootprints, 2018.
[162] Microsoft, Canadian building footprints. https://github.com/microsoft/CanadianBuildingFootprints, 2018.
[163] Household travel survey (hts) data (Last accessed: October, 2019), https://opendata.transport.nsw.gov.au/dataset/household-travel-survey-200708-%E2%80%2593-201718.
[164] Victorian integrated survey of travel & activity (vista) data (Last accessed: October, 2019), https://transport.vic.gov.au/about/data-and-research/vista/vista-data-and-publications.
[165] Y.-S. Chiou, K.M. Carley, C.I. Davidson, M.P. Johnson, A high spatial resolution residential energy model based on american time use survey data and the bootstrap sampling method, Energy Build. 43 (12) (2011) 3528–3538.
[166] U. Wilke, F. Haldi, J.-L. Scartezzini, D. Robinson, A bottom-up stochastic model to predict building occupants' time-dependent activities, Build. Environ. 60 (2013) 254–264.
[167] I. Richardson, M. Thomson, D. Infield, C. Clifford, Domestic electricity use: a high-resolution energy demand model, Energy Build. 42 (10) (2010) 1878–1887.
[168] T.S. Blight, D.A. Coley, Sensitivity analysis of the effect of occupant behaviour on the energy consumption of passive house dwellings, Energy Build. 66 (2013) 183–192.
[169] G. Buttitta, W. Turner, D. Finn, Clustering of household occupancy profiles for archetype building models, Energy Procedia 111 (2017) 161–170.
[170] M. Lopez-Rodriguez, I. Santiago, D. Trillo-Montero, J. Torriti, A. Moreno-Munoz, Analysis and modeling of active occupancy of the residential sector in Spain: an indicator of residential electricity consumption, Energy Pol. 62 (2013) 742–751.
[171] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, F. Descamps, Discrete occupancy profiles from time-use data for user behaviour modelling in homes, in: 13th Conference of International Building Performance Simulation Association, 2013, pp. 2421–2427.
[172] Y. Yamaguchi, Y. Shimoda, Evaluation of a behavior model of occupants in home based on Japanese national time use survey. Proceedings of Building Simulation, 2015.
[173] S. Nakano, A. Washizu, Which time slots can people save power? An analysis using a Japanese survey on time use, Sustainability 11 (16) (2019) 4444.
[174] K. Foteinaki, R. Li, C. Rode, R.K. Andersen, Modelling household electricity load profiles based on Danish time-use survey data, Energy Build. 202 (2019) 109355.
[175] Z. Li, B. Dong, Short term predictions of occupancy in commercial buildings performance analysis for stochastic models and machine learning 1670 approaches, Energy Build. 158 (2018) 268–281.
[176] Z. Chen, Y.C. Soh, Comparing occupancy models and data mining approaches for regular occupancy prediction in commercial buildings, Journal of Building Performance Simulation 10 (5–6) (2017) 545–553.
[177] Z. Li, B. Dong, A new modeling approach for short-term prediction of occupancy in residential buildings, Build. Environ. 121 (2017) 277–290.
[178] M.S. Rahaman, H. Pare, J. Liono, F.D. Salim, Y. Ren, J. Chan, S. Kudo, T. Rawling, A. Sinickas, Occuspace: towards a robust occupancy prediction system for activity based workplace, in: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE, 2019, pp. 415–418, 2019.
[179] A.A. Adamopoulou, A.M. Tryferidis, D.K. Tzovaras, A context-aware method for building occupancy prediction, Energy Build. 110 (2016) 229–244.
[180] V.L. Erickson, M.A. Carreira-Perpinan, A.E. Cerpa, Occupancy modeling and prediction for building energy management, ACM Trans. Sens. Netw. (TOSN) 10 (3) (2014) 42.
[181] W. Wang, J. Chen, T. Hong, N. Zhu, Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology, Build. Environ. 138 (2018) 160–170.
[182] M.M. Ouf, M.H. Issa, A. Azzouz, A.-M. Sadick, Effectiveness of using WiFi technologies to detect and predict building occupancy, Sustainable Buildings 2 (2017) 1–10.
[183] W. Wang, T. Hong, N. Xu, X. Xu, J. Chen, X. Shan, Cross-source sensing data fusion for building occupancy prediction with adaptive lasso feature filtering, Build. Environ. 162 (2019) 106280.
[184] S. Dedesko, B. Stephens, J.A. Gilbert, J.A. Siegel, Methods to assess human occupancy and occupant activity in hospital patient rooms, Build. Environ. 90 (2015) 136–145.
[185] L. Smith, M. Ucci, A. Marmot, R. Spinney, M. Laskowski, A. Sawyer, M. Konstantatou, M. Hamer, G. Ambler, J. Wardle, et al., Active buildings: modelling physical activity and movement in office buildings. An observational study protocol, BMJ open 3 (11) (2013), e004103.
[186] Y. Ren, M. Tomko, F.D. Salim, J. Chan, C.L. Clarke, M. Sanderson, A location-query-browse graph for contextual recommendation, IEEE Trans. Knowl. Data Eng. 30 (2) (2017) 204–218.
[187] Y. Ren, M. Tomko, F.D. Salim, J. Chan, M. Sanderson, Understanding the predictability of user demographics from cyber-physical-social behaviours in indoor retail spaces, EPJ Data Science 7 (1) (2018) 1.
[188] M. Kaur, F.D. Salim, Y. Ren, J. Chan, M. Tomko, M. Sanderson, Shopping intent recognition and location prediction from cyber-physical activities via wi-_ logs, in: Proceedings of the 5th Conference on Systems for Built Environments, ACM, 2018, pp. 130–139.
[189] S. Jiang, Y. Yang, S. Gupta, D. Veneziano, S. Athavale, M.C. Gonzalez, The timegeo modeling framework for urban mobility without travel surveys, Proc. Natl. Acad. Sci. 113 (37) (2016) E5370–E5378.
[190] V.W. Zheng, Y. Zheng, X. Xie, Q. Yang, Collaborative location and activity recommendations with gps history data, in: Proceedings of the 19th International Conference on World Wide Web, ACM, 2010, pp. 1029–1038.
[191] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, P. Shenoy, Non-intrusive occupancy monitoring using smart meters, in: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, ACM, 2013, pp. 1–8.
[192] M.S. Rahaman, M. Hamilton, F.D. Salim, Using Big Spatial Data for Planning User Mobility, Springer International Publishing, Cham, 2018, pp. 1–6.
[193] M.S. Rahaman, M. Hamilton, F.D. Salim, Predicting imbalanced taxi and passenger queue contexts in airport, PACIS, 2017, p. 172.
[194] M.S. Rahaman, M. Hamilton, F.D. Salim, Queue context prediction using taxi driver knowledge, in: Proceedings of the Knowledge Capture Conference, ACM, 2017, p. 35.
[195] M.S. Rahaman, Y. Ren, M. Hamilton, F.D. Salim, Wait time prediction for airport taxis using weighted nearest neighbor regression, IEEE Access 6 (2018) 74660–74672.
[196] I. Rhee, M. Shin, S. Hong, K. Lee, S.J. Kim, S. Chong, On the levy-walk nature of human mobility, IEEE/ACM Trans. Netw. 19 (3) (2011) 630–643, https://doi.org/10.1109/TNET.2011.2120618.
[197] S. Isaacman, R. Becker, R. Caceres, M. Martonosi, J. Rowland, A. Varshavsky, W. Willinger, Human mobility modeling at metropolitan scales, in: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, ACM, 2012, pp. 239–252.
[198] L. Pappalardo, S. Rinzivillo, F. Simini, Human mobility modelling: Exploration and preferential return meet the gravity model, Procedia Computer Science vol. 83 (2016) 934–939, https://doi.org/10.1016/j.procs.2016.04.188. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016)/the 6th International Conference on Sustainable Energy Information Technology (SEIT-2016)/Affiliated Workshops, http://www.sciencedirect.com/science/article/pii/S1877050916302216.
[199] S. Hoteit, S. Secci, S. Sobolevsky, C. Ratti, G. Pujolle, Estimating human trajectories and hotspots through mobile phone data, Comput. Network. 64 (2014) 296–307.
[200] F. Simini, M.C. Gonzalez, A. Maritan, A.-L. Barabasi, A universal model for mobility and migration patterns, Nature 484 (7392) (2012) 96.
[201] M. Mazzoli, A. Molas, A. Bassolas, M. Lenormand, P. Colet, J.J. Ramasco, Field theory for recurrent mobility, Nat. Commun. 10 (1) (2019) 1–10.
[202] S.K. Rumi, K.K. Qin, F.D. Salim, modeling police patrol route in dynamic environment", in: Proceedings of ICWSM Conference, 2020.

[203] I. Gaetani, P. Hoes, J. Hensen, Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy, Energy Build. 121 (2016) 188–204.
[204] B. Huchuk, W. O'Brien, S. Sanner, A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data, Build. Environ. 139 (2018) 199–210, https://doi.org/10.1016/j.buildenv.2018.05.003. http://www.sciencedirect.com/science/article/pii/S0360132318302634.
[205] V.L. Castaldo, I. Pigliautile, F. Rosso, F. Cotana, F. De Giorgio, A.L. Pisello, How subjective and non-physical parameters affect occupants, environmental comfort perception 178 (2018) 107–129, https://doi.org/10.1016/j.enbuild.2018.08.020.
[206] V. Masson, A physically-based scheme for the urban energy budget in atmospheric models, Boundary-Layer Meteorol. 94 (3) (2000) 357–397, https://doi.org/10.1023/A:1002463829265.
[207] D. Li, E. Bou-Zeid, Quality and sensitivity of high-resolution numerical simulation of urban heat islands, Environ. Res. Lett. 9 (2014) 73, https://doi.org/10.1088/1748-9326/9/5/055001.
[208] G.F. Garuma, Review of urban surface parameterizations for numerical climate models, Urban Climate 24 (2018) 830–851, https://doi.org/10.1016/j.uclim.2017.10.006. September 2017.
[209] N. Antoniou, H. Montazeri, M. Neophytou, B. Blocken, CFD simulation of urban microclimate: validation using high-resolution field measurements, Sci. Total Environ. 695 (2019) 133743, https://doi.org/10.1016/J.SCITOTENV.2019.133743.
[210] C. Piselli, V. Castaldo, I. Pigliautile, A. Pisello, F. Cotana, Outdoor comfort conditions in urban areas: on citizens' perspective about microclimate mitigation of urban transit areas, Sustainable Cities and Society 39 (2018) 16–36, https://doi.org/10.1016/j.scs.2018.02.004.
[211] V. Horoshenkov, C. Hothersall, E. Mercy, Scale modelling of sound propagation in a city street canyon, J. Sound Vib. 223 (5) (1999) 795–819, https://doi.org/10.1006/jsvi.1999.2157.
[212] G.M. Echevarria Sanchez, T. Van Renterghem, P. Thomas, D. Botteldooren, The effect of street canyon design on traffic noise exposure along roads, Build. Environ. 97 (2016) 96–110, https://doi.org/10.1016/j.buildenv.2015.11.033.
[213] C. Lavandier, P. Aumond, S. Gomez, C. Domingues, Urban Soundscape Maps Modelled with Geo-Referenced data, doi:10.1515/noise-2016-0020.
[214] J.Y. Hong, J.Y. Jeon, Exploring spatial relationships among soundscape variables in urban areas: a spatial statistical modelling approach, Landsc. Urban Plann. 157 (2017) 352–364, https://doi.org/10.1016/j.landurbplan.2016.08.006.
[215] J.Y. Hong, J.Y. Jeon, Relationship between spatiotemporal variability of soundscape and urban morphology in a multifunctional urban area: a case study in Seoul, Korea, Build. Environ. 126 (2017) 382–395, https://doi.org/10.1016/j.buildenv.2017.10.021.
[216] V. Puyana Romero, L. Ma_ei, G. Brambilla, G. Ciaburro, Modelling the soundscape quality of urban waterfronts by artificial neural networks, Appl. Acoust. 111 (2016) 121–128, https://doi.org/10.1016/j.apacoust.2016.04.019.
[217] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (8) (2009) 1819–1835.
[218] J. Li, Q. Ban, X.J. Chen, J. Yao, Glazing sizing in large atrium buildings: a perspective of balancing daylight quantity and visual comfort, Energies 12 (4) (2019) 701.
[219] N. Nasrollahi, E. Shokri, Daylight illuminance in urban environments for visual comfort and energy performance, Renew. Sustain. Energy Rev. 66 (2016) 861–874.
[220] R.A. Mangkuto, Validation of DIALux 4.12 and DIALux evo 4.1 against the analytical test cases of CIE 171: 2006, Leukos 12 (3) (2016) 139–150.
[221] C.E. Ochoa, M.B. Aries, J.L. Hensen, State of the art in lighting simulation for building science: a literature review, Journal of Building Performance Simulation 5 (4) (2012) 209–233.
[222] A.W. Hammad, A. Akbarnezhad, Sustainable lighting layout in urban areas: maximizing implicit coverage and minimizing installation cost, Frontiers in built environment 4 (2018) 42.
[223] Y. Xie, T. Huang, J. Li, J. Liu, J. Niu, C.M. Mak, Z. Lin, Evaluation of a multi-nodal thermal regulation model for assessment of outdoor thermal comfort: sensitivity to wind speed and solar radiation, Build. Environ. 132 (2018) 45–56.
[224] B. Howard, L. Parshall, J. Thompson, S. Hammer, J. Dickinson, V. Modi, Spatial distribution of urban building energy consumption by end use, Energy Build. 45 (2012) 141–151.
[225] Y Lai, CE Kontokosta, Topic modeling to discover the thematic structure and spatial-temporal patterns of building renovation and adaptive reuse in cities, Comput. Environ. Urban Syst. 78 (2019) 101383.
[226] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summer_eld, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, Build. Environ. 45 (7) (2010) 1683–1697.
[227] J. New, M. Adams, P. Im, H. L. Yang, J. Hambrick, W. Copeland, L. Bruce, J. A. Ingraham, Automatic building energy model creation (AutoBEM) for urban-scale energy modeling and assessment of value propositions for electric utilities, Proceedings of the INternational Conference on Energy Engineering and Smart Grids (ESG).
[228] N. Rivers, M. Jaccard, Combining top-down and bottom-up approaches to energy-economy modeling using discrete choice methods, Energy J. (2005) 83–106.
[229] M. Braulio-Gonzalo, P. Juan, M.D. Bovea, M.J. Rua, Modelling energy efficiency performance of residential building stocks based on Bayesian 75 statistical inference, Environ. Model. Software 83 (2016) 198–211.
[230] A. Galante, M. Torri, et al., A methodology for the energy performance classification of residential building stock on an urban scale, Energy Build. 48 (2012) 211–219.
[231] Z. Pang, P. Xu, Z. O'Neill, J. Gu, S. Qiu, X. Lu, X. Li, Application of mobile positioning occupancy data for building energy simulation: an engineering case study, Build. Environ. 141 (2018) 1–15.
[232] G. Jiefan, X. Peng, P. Zhihong, C. Yongbao, J. Ying, C. Zhe, Extracting typical occupancy data of different buildings from mobile positioning data, Energy Build. 180 (2018) 135–145.
[233] H. Song, A.K. Qin, F.D. Salim, Multi-resolution selective ensemble extreme learning machine for electricity consumption prediction, in: International Conference on Neural Information Processing, Springer, 2017, pp. 600–609.
[234] H. Song, A.K. Qin, F.D. Salim, Evolutionary multi-objective ensemble learning for multivariate electricity consumption prediction, in: International Joint Conference on Neural Networks (IJCNN), IEEE, 2018, pp. 1–8, 2018.
[235] H. Song, A. Qin, F.D. Salim, Evolutionary model construction for electricity consumption prediction, Neural Comput. Appl. (2019) 1–18.
[236] D. Robinson, N. Campbell, W. Gaiser, K. Kabel, A. Le-Mouel, N. Morel, J. Page, S. Stankovic, A. Stone, Suntool-a new modelling paradigm for simulating and optimising urban sustainability, Sol. Energy 81 (9) (2007) 1196–1211.
[237] D. Robinson, F. Haldi, P. Leroux, D. Perez, A. Rasheed, U. Wilke, Citysim: comprehensive micro-simulation of resource flows for sustainable urban planning, in: Proceedings of the Eleventh International IBPSA Conference, CONF, 2009, pp. 1083–1090.
[238] C. Reinhart, T. Dogan, J.A. Jakubiec, T. Rakha, A. Sang, Umi-an urban simulation environment for building energy use, daylighting and walkability. 13th Conference of International Building Performance Simulation Association, Chambery, France, 2013.
[239] T. Hong, Y. Chen, S. H. Lee, M. A. Piette, Citybes: A Web-Based Platform to Support City-Scale Building Energy Efficiency, Urban Computing vol. 14.
[240] P. Remmen, M. Lauster, M. Mans, M. Fuchs, T. Osterhage, D. Müller, Teaser: an open tool for urban energy modelling of building stocks, Journal of Building Performance Simulation 11 (1) (2018) 84–98.
[241] L.A. Bollinger, R. Evins, Hues: a holistic urban energy simulation platform for effective model integration, in: Proceedings of International Conference CISBAT 2015 Future Buildings and Districts Sustainability from Nano to Urban Scale, CONF, LESO-PB, EPFL, 2015, pp. 841–846.
[242] Y Chen, T Hong, MA Piette, City-scale building retrofit analysis: a case study using CityBES, Build. Simul. (2017).
[243] L.A. Bollinger, R. Evins, Facilitating model reuse and integration in an urban energy simulation platform, Procedia Computer Science 51 (2015) 2127–2136.
[244] J. Page, D. Robinson, N. Morel, J.-L. Scartezzini, A generalised stochastic model for the simulation of occupant presence, Energy Build. 40 (2) (2008) 83–98.
[245] A. Allen, S. Zakery, Y. Mao, D. Robinson, A rapid urban de-carbonization scenario analysis tool, Procedia engineering 198 (2017) 826–835.
[246] F. Haldi, D. Robinson, The impact of occupants' behaviour on building energy demand, Journal of Building Performance Simulation 4 (4) (2011) 323–338.
[247] I.B. Arief-Ang, F.D. Salim, M. Hamilton, Da-hoc: semi-supervised domain adaptation for room occupancy prediction using co2 sensor data, in: Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments, BuildSys '17, ACM, New York, NY, USA, 2017, 1:1-1:10.
[248] W. Hu, Y. Luo, Z. Lu, Y. Wen, Heterogeneous transfer learning for thermal comfort modeling, in: Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation - BuildSys '19, ACM Press, 2019, pp. 61–70, https://doi.org/10.1145/3360322.3360843.
[249] K. Doolin, I. Roussaki, M. Roddy, N. Kalatzis, E. Papadopoulou, N. Taylor, N. Liampotis, D. McKitterick, E. Jennings, P. Kosmides, Societies: where pervasive meets social, in: The Future Internet Assembly, Springer, 2012, pp. 30–41.
[250] Y. Zheng, Methodologies for cross-domain data fusion: an overview, IEEE transactions on big data 1 (1) (2015) 16–34.
[251] P. Latouche, F. Rossi, Graphs in machine learning: an introduction, in: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), Proceedings of the 23-Th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015), 2015, pp. 207–218.
[252] P.M. Domingos, A few useful things to know about machine learning, Commun. ACM 55 (10) (2012) 78–87.
[253] J. Yang, C. Ning, C. Deb, F. Zhang, D. Cheong, S.E. Lee, C. Sekhar, K.W. Tham, k-shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement, Energy Build. 146 (2017) 27–37.
[254] H. Naganathan, W.O. Chong, X. Chen, Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches, Autom. ConStruct. 72 (2016) 187–194.
[255] A. Liaw, M. Wiener, et al., Classification and regression by random forest, R. News 2 (3) (2002) 18–22.
[256] V. Vapnik, The support vector method of function estimation. Nonlinear Modeling, Springer, 1998, pp. 55–85.
[257] D.G. Kleinbaum, K. Dietz, M. Gail, M. Klein, M. Klein, Logistic Regression, Springer, 2002.
[258] S.R. Safavian, D. Landgrebe, A survey of decision tree classifier methodology, IEEE transactions on systems, man, and cybernetics 21 (3) (1991) 660–674.
[259] G.A. Seber, A.J. Lee, Linear Regression Analysis, vol. 329, John Wiley & Sons, 2012.
[260] J.M. Hilbe, Negative Binomial Regression, Cambridge University Press, 2011.
[261] M. Braun, H. Altan, S. Beck, Using regression analysis to predict the future energy consumption of a supermarket in the UK, Appl. Energy 130 (2014) 305–313.

[262] G.E. Hinton, R.S. Zemel, Autoencoders, minimum description length and helmholtz free energy. Advances in Neural Information Processing Systems, 1994, pp. 3–10.

[263] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436.

[264] D.P. Mandic, J. Chambers, Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability, John Wiley & Sons, Inc., 2001.

[265] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Trans. Neural Network. 20 (1) (2008) 61–80.

[266] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets. Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.

[267] L.P. Kaelbling, M.L. Littman, A.W. Moore, Reinforcement learning: a survey, J. Artif. Intell. Res. 4 (1996) 237–285.

[268] C. Li, Z. Ding, D. Zhao, J. Yi, G. Zhang, Building energy consumption prediction: an extreme deep learning approach, Energies 10 (10) (2017) 1525.

[269] L. Cao, C. Zhang, Q. Yang, D. Bell, M. Vlachos, B. Taneri, E. Keogh, S.Y. Philip, N. Zhong, M.Z. Ashra, et al., Domain-driven, actionable knowledge discovery, IEEE Intell. Syst. 22 (4) (2007) 78–88.

[270] L. Cao, Actionable knowledge discovery and delivery. Meta Synthetic Computing and Engineering of Complex Systems, Springer, 2015, pp. 287–312.