Applied Energy 293 (2021) 116856

Contents lists available at ScienceDirect

Applied Energy

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

Occupant behavior modeling methods for resilient building design, operation and policy at urban scale: A review

Bing Dong[a,*], Yapan Liu[a], Hannah Fontenot[a], Mohamed Ouf[b], Mohamed Osman[b], Adrian Chong[c], Shuxu Qin[c], Flora Salim[d], Hao Xue[d], Da Yan[e], Yuan Jin[e], Mengjie Han[f], Xingxing Zhang[f], Elie Azar[g], Salvatore Carlucci[h]

[a] Department of Mechanical & Aerospace Engineering, Syracuse University, 263 Link Hall, Syracuse, NY 13244, United States
[b] Department of Building, Civil and Environmental Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, EV 6.139, Montreal, Quebec H3G 1M8, Canada
[c] Department of Building, School of Design and Environment, National University of Singapore, 4 Architecture Drive, Singapore 117566, Singapore
[d] School of Computing Technologies, RMIT University, Melbourne, Victoria 3000, Australia
[e] Building Energy Research Center, School of Architecture, Tsinghua University, Beijing 100084, China
[f] School of Technology and Business Studies, Dalarna University, 79188 Falun, Sweden
[g] Department of Industrial and Systems Engineering, Khalifa University of Science and Technology, PO Box 127788, Abu Dhabi, United Arab Emirates
[h] Energy, Environment, And Water Research Center, The Cyprus Institute, Nicosia, Cyprus

HIGHLIGHTS

- Review of occupant behavior modeling methods in Building Science and beyond.
- Bridge the knowledge gap between Building Science and other domains.
- Cross comparison of the modeling requirements in different domains.
- Summary of the various metrics for model accuracy evaluation.

ARTICLE INFO

Keywords:
Building science
Urban scale
Occupant behavior modeling
Human mobility modeling
Cross domain
Spatio-temporal data
Performance evaluation

ABSTRACT

Traditional occupant behavior modeling has been studied at the building level, and it has become an important factor in the investigation of building energy consumption. However, studies modeling occupant behaviors at the urban scale are still limited. Recent work has revealed that urban big data can enable occupant behavior modeling at the urban scale – however, utilizing the existing data sources and modeling methods in building science to model urban scale occupant behaviors can be quite challenging. Beyond building science, urban scale human behaviors have been studied in several different domains using more advanced modeling methods, including Stochastic Modeling, Neural Networks, Reinforcement Learning, Network Modeling, etc. This paper aims to bridge the gap between data sources and modeling methodologies in building science by borrowing from other domains. Based on a comprehensive review, we 1) identify the modeling challenges of the current approaches in building science, 2) discuss the modeling requirements and data sources both in building science and other domains, 3) review the current modeling methods in building science and other domains, and 4) summarize available performance evaluation metrics for evaluating the modeling methods. Finally, we present future opportunities in building science with enhanced data sources and modeling methods from other domains.

* Corresponding author.
E-mail address: bidong@syr.edu (B. Dong).

https://doi.org/10.1016/j.apenergy.2021.116856
Received 8 November 2020; Received in revised form 7 February 2021; Accepted 18 March 2021
Available online 21 April 2021

1. Introduction

1.1. Background

By 2050, a staggering 70% of the world's population is projected to live and work in cities, while two-thirds of global primary energy consumption will be attributed to cities, leading to the major production of the global direct energy-related greenhouse gas emissions. Occupant behavior has emerged as one of the leading influences on energy consumption in buildings. Current occupant behavior studies are often isolated and only research individual behavior such as presence or interactions [1–5] in a single space or building [6]. Recent studies have also addressed a breadth of optimization, control and occupancy-related challenges for the operation of individual buildings [7,8]. However, modeling occupant behavior at an urban and city scale for the purpose of building design and operation is limited [9,10].

Applications in building science:

Currently, there are four main application areas in which urban scale occupant behavior modeling plays key roles:

- District heating and cooling system design and operation

With the development of district heating and cooling systems, the design and operational strategy for heating and cooling should not be limited to a single zone or building but should encompass the urban scale [11]. Heat load profiles have been demonstrated to be related to users' behavior [12,13], and the internal heat gain from occupants is the most influential factor [14,15]. In addition, behavior mode of air conditioner and occupant type composition lead to a 4.5-fold difference in the system efficiency of district cooling systems [16] and simplification of occupant behavior could result in overestimation of the district peak load [17].

- Urban building energy modeling (UBEM) and analysis

UBEMs aspire to become key planning tools for the holistic optimization of buildings, urban design, and energy systems in neighborhoods and districts [10]. However, inappropriate choice of occupant behavioral model could result in oversized district energy systems [18,19] and a wrong estimation of energy saving due to the implementation of energy conservation measures. For example, Barbour [20] investigated urban-scale occupant behavior (U-OB) in the Greater Boston area and showed that the use of observation-derived occupancy profiles can help to reduce energy consumption up to 15% for residential buildings and 21% for commercial buildings. Similarly, several recent studies [21–24] found that energy consumption can be further reduced (up to 60% in some cases) by adopting realistic occupancy profiles. In addition, the analysis of the impact of anthropogenic heat released in urban areas is an important topic that is gathering high interest in the last 5 years. The main research purposes are to understand and predict urban overheating and local urban heat island, to study the interaction between buildings and the built environment, and to assess local microclimate and hotspot risk. Some tools integrate walkability metrics to estimate how friendly an outdoor space is to walking [25].

- Buildings-to-grid integration

Prior studies have developed a framework to understand the coupling of buildings-to-grid integration considering community scale occupant behavior [26–28]. Occupant behavior models have been used to generate appliance usage patterns (human-building interactions) in residential buildings [29] and working schedules (occupant presence) in commercial buildings. Such behavior modeling impacts how buildings respond to demand response signals from the grid [30], building flexibility [31], space-level heating, ventilation, and air conditioning (HVAC) controls [32], the optimization of solar photovoltaic (PV) planning in an energy community [33], and peer-to-peer (P2P) business models for individual PV prosumers in a local electricity market [34].

- Energy policy

There are two main categories of policy mechanisms that affect - and are affected by - the quality of occupant behavior representation [35]. The first category covers regulations, namely building energy codes and standards used to develop building technologies. The development of regulations is often guided by building energy modeling. The second category covers demand-management strategies, mainly occupant interventions that aim to alter existing energy consumption patterns (e.g., energy education, feedback, and incentives).

Applications in other domains:

Modeling human behavior has been studied in other domains such as traffic analysis, epidemiology, disaster management, and marketing through modeling human mobility. Various datasets have been used to analyze human mobility: Geospatial trajectories/ Global Positioning System (GPS) data, Call Detail Records (CDR) data, social media check-ins and apps, location-based service (LBS) data, and transportation data [25]. Individual mobility data can be used for enhanced location-based services and can be applied to solve social and market challenges such as personalized energy services and route optimization, as well as smart and green transportation architectures for citizens and goods.

- Traffic analysis

Leveraging social network services and mobile computing can enable urban planners to investigate city dynamics [36]. Using human mobility and social media data, traffic anomalies can be visualized [37] and detected [36,38,39]. In addition, cyclists' destinations can be predicted by learning human behavior, spatial relationships, and external features [40]. Other studies show that future driving trip paths can be predicted through taxi GPS data [41].

- Epidemiology

Traditionally, disease spread is modeled using models which divide a population into "compartments". For instance, in the susceptible-infectious-recovered (SIR) model, individuals undergo a process of being susceptible to a disease, then infectious, then recovered. In the susceptible-exposed-infectious-recovered (SEIR) model [42], there is a period of being exposed before becoming infectious. Individuals move from one compartment to the next according to some modeled probability. Often these models include an assumption of population homogeneity or invariance. Recently, however, researchers have made use of several large-scale datasets to model human mobility and its impact on disease spread, including Bluetooth and WiFi location data [43]–[45], CDR data [46], and Google aggregated mobility data [47].

- Disaster Management

Understanding of human movements in urban areas plays a key role in improving our disaster response, evacuation, and relief plans [48]. By analyzing individuals' movement data collected from Twitter, the influence of natural disasters on human mobility patterns can be better understood [49]. In addition, population displacement during or after natural disasters can be estimated through analyzing CDR data [50]. Furthermore, recent research presents a social media-based approach to assess disaster impacts on highways [51].

- Visitor Analytics, Smart Retail, and Recommender Systems

Recommender systems investigate and model people's cyber-physical actions (moving, searching, browsing, etc.) in order to predict user demographics [52] and customers' shopping intent [53,54], make contextual recommendations [55] to shoppers, or predict their revisit intention [56]. Similarly, graph-based methods can also be used to predict point of interest (POI) and user activities [57], as well as behavioral trends [58–60]. These methods make use of large-scale smartphone datasets including smartphone continuous sensing data, WiFi, Bluetooth, and CDR data.

1.2. Key Challenges, and research questions

Despite the different approaches to modeling occupant behavior on the urban scale, there are still several key challenges in simulating occupant behavior:

1) Insufficient spatial diversity among all different buildings with different functions and occupant behaviors. All current measured occupant behavior datasets are not large enough (both spatially and temporally) to represent all building types at a district or urban level;

2) Insufficient temporal and longitudinal diversity among the same type of building with the same occupancy type. Occupant behavior is stochastic in nature. Even for the same type of occupant, the behavior will be different over time. It is a challenge to represent such diversity in simulating occupant behavior at an urban scale.

Hence, the goal of this paper is to bridge the data sources and methodology gap between building science and beyond. In order to achieve this goal, the following research questions are answered throughout this paper:

- What are the modeling requirements of occupant behavior at a community level?
- What data sources have been used in other domains that could potentially enhance the modeling capabilities for current building science applications?
- What are the current modeling methods of occupant behavior at a community level?
- What modeling methods have been used in other domains that could potentially enhance the modeling capabilities for building science applications?
- What are the potential future research directions for building design, operation, and policies at a community level, with enhanced data sources and modeling methods from other domains?

1.3. Review methods and structure of the article

This work aims to understand current modeling approaches as well as big data requirements in building science, and investigate recent modeling methods in other domains which include transportation, epidemiology, and disaster management among others. We used Google Scholar to perform a literature search with keywords including "modeling building occupant behavior," "urban scale occupant behavior," "human mobility patterns," "human dynamics," and "spatio-temporal data." After first reading through the abstracts and evaluating the papers according to their relevance to the purpose of this study, we then filtered out the articles which do not provide detailed modeling approaches and data sources. Based on our best knowledge, 206 publications were chosen and further considered in this review work. We have categorized those publications by topic and modeling method. Based on the research questions listed above, we structure the paper as the following: Section 2 analyzes the modeling requirement for the aforementioned four applications in building science and data used in other domains. Section 3 discusses the current occupant behavior modeling methods at community and urban scale. Section 4 presents the research opportunities with occupant behavior modeling methods from other domains and Section 5 concludes this paper.

2. Modeling requirement analysis and data sources

This section answers the following two research questions:

- What are the modeling requirements of occupant behavior for the aforementioned four applications at a community level and how do current approaches meet those requirements?
- What data sources have been used in other domains that could potentially enhance the modeling inputs for current building science applications?

Specifically, we will review spatial and temporal characteristics of occupant behavior modeling requirements for the aforementioned four applications in building science. In addition, data sources and characteristics from other domains, which have been used to model occupant behavior, will be reviewed to cross compare with the requirements from building science.

2.1. Urban scale building design and operation

2.1.1. District heating and cooling system design and operation

Modeling district heating and cooling (DHC) systems requires extracting data from each connected building, including its characteristics, energy use profiles, thermostat setpoints, and occupancy status [61–64]. For these district systems, previous research was conducted using field measurements [16,17,65]. The critical balance between energy supply and demand in such systems required extracting on-site measurements from connected smart meters and sensors with temporal granularity ranging from minutes [66] up to one hour [67,68]. However, the occupancy state in each building and the interactions of its occupants with building systems are key inputs for cooling or heating load simulation, which are not typically considered. Smart thermostat datasets with indoor and outdoor temperature measurements as well as setpoint profiles can provide this information and identify the impact of external parameters such as season and energy price on occupant behavior [13,69–71]. The duration of data collection can range from several months [72,73] up to several years [74], with a temporal resolution from 10 to 15 min, up to one hour [61,66,75]. Upscaling the research objective to the urban scale requires improving the method of data collection accordingly. Therefore, questionnaire surveys [76,77] can be used to supplement the related data at a large spatial scale.

2.1.2. Urban scale building energy performance modeling and analysis

Salim et al. [25] reviewed the use of occupant-centric urban data to model occupant behavior and address the uncertainties in UBEMs. Modeling occupancy in UBEM requires data with fine temporal resolution (seconds or minutes) and spatial resolution defined by the latitude and longitude of each location with relatively high accuracy. Thankfully, the proliferation of big data technologies enables this level of data granularity. With derived urban scale building occupancy profiles from big data, recent studies have integrated urban scale occupant behaviors with UBEM [20,22,23]. For example, Barbour et al. [20] used CDR data of 40 days, with a temporal resolution of 10 min and spatial resolution of 300 m, to model commercial buildings' occupancy using the novel TimeGeo model [78]. The TimeGeo framework was established on a time-inhomogeneous Markov chain model for modeling temporal choices, and a rank-based exploration and preferential return (r-EPR) model for generating spatial choices. Alternatively, aggregated data such as the total number of occupants and their hourly distribution over the day were enough to create the typical occupancy schedules [79]. Location-based services (LBS) are promising data sources that presently include only public access buildings; however, they are expected to have residential buildings and private offices data in upcoming years [80]. The Google popular time (GPT) data is used to develop dynamic occupancy schedules for different types and scales of public buildings [80–82]. Moreover, Mohammadi and Taylor [83] used one year of location records extracted from Twitter to predict energy consumption of residential buildings. Lu et al. [84] showed a slight difference between the schedules developed based on different LBS datasets, reinforcing the significance of validation phases. Moreover, smart connected sensors provide an opportunity to gather standardized and scalable on-site data with high temporal resolution regarding occupancy and occupant's behavior, as shown in [85–88]. Wang et al. [89] used census data, transportation accessibility data, and population data to upscale a building scale dynamic occupancy model to the city level.

2.1.3. Buildings-to-grid (B2G) integration

Currently, the B2G framework typically couples the physical models of buildings and power grid. The building model is represented by a reduced order thermal capacitance and resistance model [90], and the power grid model is represented by power flow equations that are solved by optimal power flow [91]. Typically, building load is generated either through ad-hoc assumption [26,92] or simulation models [93] without considering the occupant behavior. However, in reality, occupant behavior becomes a leading factor for building energy usages. The occupant behavior models at each building are important to generate more accurate building loads at any given time so that the power grid model can optimize the power flow accurately. Such occupant behavior models should include both appliance usages in residential buildings and occupant presence in commercial buildings. Unfortunately, large scale field measurement of occupant behavior is not available with the same scale of power grid distribution network. The current approach from few studies to generate such behavior models includes sampling from limited field measurement for appliance usages and presence [28,94]. The measured data include thermostat behavior, water heater usages, kitchen appliance usages, and electric vehicle (EV) charge behaviors. Due to the coupled control framework, those data need to be at a fine granularity of every 15 or 30 min, and at each building level [95]. The same scenario occurs for commercial buildings. A stochastic occupant behavior simulation tool is often used to generate occupancy presence data [96,97].

2.1.4. Energy policy

Energy policy covers two main types of mechanisms, namely regulations and demand-side management strategies. Regarding regulations, building energy modeling is often used to guide the design of building codes and building technology standards. Consequently, simplistic occupant behavior representation in models could lead to "unsuitable technology adoption and investment in building energy efficiency" [35]. Despite advances in sensing and information and communications technologies (ICT) to gather occupant behavior data from diverse sources, current data collection efforts remain disaggregated. There is a growing need to expand and merge current datasets to have representative occupant behavior data for different building types and geographical areas, avoiding generic occupant behavior assumptions (e.g., ASHRAE's diversity profiles) [35]. As for demand-side management strategies, modeling and simulations are increasingly used to experiment with and guide the design of such methods. One such example is the work of Azar and Al Ansari [98], who proposed an Agent-Based Modeling (ABM) framework to simulate and optimize feedback methods at the urban level, accounting for the interaction of occupants within and between buildings through social networks. A common gap in simulation-driven occupant behavior intervention studies is the lack of data to validate the predictions of the developed models. Future data collection efforts should include (1) data on the multidisciplinary drivers of behaviors (e.g., physiological, social, economic, and psychological) and (2) longitudinal occupant behavior data pre- and post-intervention (e.g., 6 months before and 12 months after) to help validate existing models and guide the design of future intervention mechanisms.

2.2. Beyond building science

2.2.1. Transportation

Recently, mobility data generated in or near real-time are extensively integrated into travel models and different intelligent transportation systems (ITS) applications [99–101]. Kaffash et al. [99] reviewed the algorithms used in the ITS field and showed that artificial neural networks and deep learning algorithms were the most frequently used methods. Considering this review's context, the spatial–temporal dynamic models developed to estimate human activities in transportation systems have a promising potential to be repositioned in the building design research field. In ITS research, the origin–destination matrices in travel demand models are generated using data sources such as GPS tracking, smart ticketing, mobile network connections, and multimedia data (speech, text, image, etc.) [102,103]. CDR data has also been used to model the universal home-work commute pattern in different countries [104–106]. Researchers have developed an exploration and preferential return (EPR) model to classify individuals as returners and explorers and predict the returners' frequently visited locations such as work or home [105,107]. Fekih et al. [104] developed an expansion factor that upscales CDR data to represent the whole population. Moreover, social media platforms have emerged as a promising data source due to both their widespread popularity in many countries and their high temporal resolution. Researchers have demonstrated the practicality of extracting mobility data from Twitter and Google maps to predict users' future activities and locations [49,100]. Considering the strengths and weaknesses of the available data sources, CDR and GPS data overcome the limitations faced by other data sources such as scale, resolution, adaptation, and representation [104]. However, GPS and CDR companies intentionally reduce the temporal and spatial accuracy of data to anonymize them and protect users' privacy [108].

2.2.2. Epidemiology

The COVID-19 pandemic has necessitated the expansion of the concept of human behaviors to include such things as wearing a face-mask, improving personal hygiene, and practicing social distancing. Such impacts of the pandemic on human behavior have received little attention in the literature thus far, and most epidemic models consider only the impact of human mobility on disease spread [109]. Traditionally, disease spread is modeled either without the inclusion of human movement (i.e., humans are modeled as a spatially fixed population [110] within a compartmentalized model such as susceptible-exposed-infection-recovered, or SEIR) or with a simulated model of human movement that is inferred from large-scale human mobility data but not necessarily extracted from said data. In one such approach, human trajectories were modeled as a random walk in which the step lengths are assumed to follow a Lévy distribution [111]. Other studies use census data to model mobility using such a distribution [112], and mobility patterns have also been extracted from bank note trajectories [113]. When it comes to large-scale datasets, there are varying data requirements depending on the research need. Dong, Pentland, and Heller used a graph-coupled hidden Markov model [43] and a Bayesian, discrete-time multi-agent model of infection [44] to model the spread of an infectious disease within a social network, training the model using the Social Evolution dataset [114] which consists of Bluetooth- and WiFi-enabled occupant tracking data (collected every six minutes) and daily flu symptom surveys completed by undergraduate dormitory residents. Madan et al. analyzed the same dataset to track disease spread by characterizing changes in face-to-face interactions and individual trajectories [45]. By using CDR data, Frias-Martinez et al. generated an agent mobility model with a granularity of 1 h to estimate the position of each agent at each moment in time, and a social network model to identify the set of individuals with whom a person has close contact [46].

The ongoing global pandemic has spurred a large body of research involving human mobility modeling. The Google aggregated mobility research dataset[1], used in [47] and [115] for COVID-19 forecasting aggregates inter-region flows of users both spatially (to a region of 5 sq. km) and temporally (on a weekly basis), and it can be further aggregated to obtain lower-resolution flows (inter-county and intra-county) depending on the required analysis. Further, the dataset includes “mobility trends” which may be used to infer social distancing-based metrics. Chang et al. [116] integrated two mobility networks - hourly movements of people in census block groups to POIs, derived from CDR data - to inform a SEIR model for modeling the spread of COVID-19 and identify more effective reopening strategies. Similarly, Ramchandani et al. [117] uses the SafeGraph dataset[2], which combines population activities data (including POI and social distancing data), mobility data (including travel data), and census features (including sociodemographic features), to develop an interpretable deep learning model, DeepCOVIDNet, for forecasting increase in infected cases for the next seven days in the United States. Panagopolous et al. [118] utilize the Facebook Data For Good datasets[3] to develop a Graph Neural Network for COVID-19 forecasting. Hu et al. [119] utilize large-scale mobility data from several sources, including Apple[4], C2SMART[5], Cuebiq[6], Google, and SafeGraph to develop a generalized additive mixed model for similar COVID-19 forecasting purposes.

2.2.3. Disaster management

Geo-tagged tweets collected from Twitter have a very fine resolution, both spatially and temporally. This level of precision is necessary to map and measure the effects of natural disasters on human mobility on a relatively small scale (i.e., within a city) [49,51,120]. CDR data, by contrast, has a fine temporal resolution but coarser spatial resolution, which is still useful for assessing national-level population movements in the wake of a natural disaster with a much wider area of impact (in the case of [50], an earthquake). Researchers in Japan constructed a large human mobility database which has GPS records from nearly 1.6 million mobile phone users with a time span from 1 August 2010 to 31 July 2011 [121,122]. The temporal and spatial resolution of this kind of GPS data can reach the level of seconds and meters. Built on Markov Decision Process, Song et al. [121] developed a general probabilistic model to simulate and predict population evacuations during severe disasters. In another study [122], a Hidden Markov Model based disaster behavior model was proposed and used to predict human emergency behavior and mobility under large-scale disasters. Another study [123] focused on capturing pedestrians’ movement behavior during disaster evacuation. With 41 GPS traces used in this study, the sampling rate of those GPS data was every 30 seconds and the spatial resolution was at meter level.

2.2.4. Visitor analytics, smart retail, and recommender system

In the fields of marketing analysis and social recommendation, large-scale, fine-resolution smartphone-based datasets are used to model user movements and actions. Kaur et al. [53,54] analyzed a dataset collected over WiFi in a large shopping mall which includes logs of WiFi access-point association and web queries. The analysis involved two steps: first, semantic categorization of physical locations to find semantic similarity between user queries and physical points; second, classification of user trajectories into two categories, intentful and intentless. In [52], a location-query-browse (LQB) graph was introduced for making contextual recommendations to customers. The LQB was trained using the same dataset (WiFi access-point association logs and web browsing logs) as mentioned previously. Kim and Lee [56] also made use of WiFi data to predict user behavior - specifically, they analyzed data at 5-meter granularity to detect the semantic location of customers and explore their revisiting behavior. In [59] and [60], a Simultaneous Extraction of Context and Community (SECC) model was proposed and demonstrated on several datasets, including the Reality Mining [124] and SigComm [125] datasets, both of which use Bluetooth proximity signals, and the StudentLife dataset [58], which contains continuous sensing and WiFi data from students’ smartphones. Datasets used in these and similar analyses require fine spatial granularity (several meters) in order to make useful predictions regarding user movements and behaviors.

2.3. Cross comparison

Fig. 1 graphically compares the modeling requirements for applications both in building science and beyond building science. Each application in building science (left side) has various data requirements; spatial requirements span three scales (building, community, city) and temporal requirements span from the scale of minutes to days for most applications. As the figure shows, the four non-building applications (right side) often use data at temporal and spatial resolutions that can fill the gaps for building application data requirements. For example, UBEM requires data spanning community and city scales. Epidemiology and disaster management studies may be able to provide data (or methods for collecting data) at those scales.

It is important to observe and learn the context from which the data are generated, particularly when dealing with heterogeneous high-dimensional data from buildings, cities, and urban areas. One main challenge in spatiotemporal sensor data is to discover meaningful relationships among the numerous sensor channels and other types of data from multiple domains, sampled at different rates, and collected for specific purposes. High quality annotations are often not available. Therefore, the alignment, fusion, or similarity analysis of sensor data needs to be done in the spatial, temporal, and data domain [126] across multiple contextual signals [53,54,59]. Furthermore, it should be noted that data type and volume provided by non-building applications may partially satisfy for building science applications. For example, when considering occupant behavior in the UBEM applications, epidemiology studies can provide human mobility data at a satisfactory temporal and spatial resolutions’ level for UBEM, while UBEM requires appliance usage data which is not usually collected in epidemiology research. This is a challenge for some building applications. The detailed review for data types and sources is conducted by Flora et al. [25]. Due to the wide variety of possible data types required by building science applications as well as those available through non-building applications, such challenges must be met, to some extent, on a case-by-case basis and may not be easily solved using a single data resource from applications beyond building science. Another major challenge is dynamic changes in the real-world, requiring a model to be robust to the fast-changing features of urban dynamics [127]. The high variability of urban big data requires data pre-processing, normalization, and feature engineering to take place prior to any machine learning and modeling tasks [25].

3. Occupant behavior modeling methodologies

This section answers the following two research questions:

- What are the current modeling methods of occupant behavior for the aforementioned four applications at a community scale?
- What modeling methods have been used in other domains that could potentially enhance the modeling capabilities for building science applications?

[1] https://www.google.com/covid19/mobility/
[2] https://www.safegraph.com/dashboard/covid19-commerce-patterns
[3] https://dataforgood.fb.com/tools/disease-prevention-maps
[4] https://covid19.apple.com/mobility
[5] http://c2smart.engineering.nyu.edu/covid-19-dashboard/
[6] https://help.cuebiq.com/hc/en-us/articles/360041285051-Cuebiq-s-COVID-19-Mobility-Insights

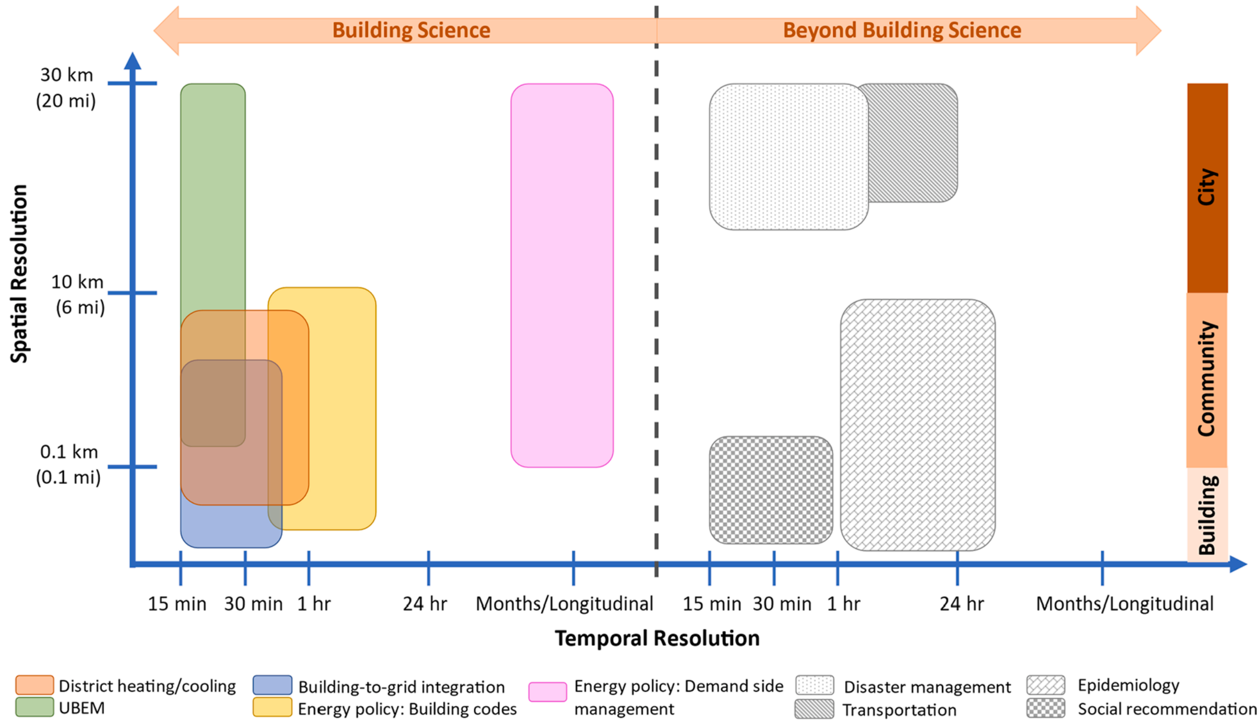

Fig. 1. Modeling requirements for applications both in building science and beyond.

3.1. *Current modeling methods in building science*

3.1.1. *Summary of current approaches*

Current urban scale occupant behavior modeling studies are briefly reviewed and summarized in Salim et al. [25], and Happle et al. [10] without details on specific modeling methods. In general, there are two main approaches: 1) Deterministic approach. A fixed occupancy schedule is often pre-defined for different building types. Such schedules come from either ad-hoc rules or standards defined by ASHRAE [128], NREL [129], and DOE [130]. Most of these schedules lack diversity, assuming that all occupants have the exact same schedule. Some have occupancy diversity through measured building occupancy density, census data, use of discrete sets of appliances in residential buildings, and random sampling from probabilistic distributions of model parameters [63]. 2) Stochastic approach. Most prior research has implemented first order non-homogeneous Markov Chain Monte Carlo technique to generate synthetic data from time-use survey data, without diversity [131]. Heterogeneous discrete-time Markov chain was used to model a three-state stochastic active-occupancy model with four types of occupants in residential buildings with diversity [132]. In addition, stochastic sampling was used to obtain individual occupant behavior and appliance usage [133].

3.1.2. *Challenges*

Based on previous reviews, stochastic models have demonstrated better performance than deterministic models at the urban scale [17]. In addition, stochastic individual-based models seem to have the capability of generating urban scale individual occupant behavior. However, at the urban scale, all prior studies only generated one type of occupant behavior per building type, mostly residential while some others are office buildings. Not a single study covers more than two building types. In addition, almost all studies neglect to discuss the method by which the urban scale occupant behavior was generated apart from a simple description of stochastic sampling based on limited measurements [133]. Furthermore, mixed-use scenarios involving different types of occupancy in a single building are not modeled unless using a deterministic approach without stochasticity. Finally, at the urban-scale, humans move from one building to another on a daily basis through certain types of transportation. Their activities are based on their presence at different buildings, and these activities relate to the energy and environmental control in those buildings. Hence, urban-scale occupant behavior is a much more complex problem that goes beyond building science and relies on knowledge from other relevant urban and heath care areas.

3.2. *Modeling methods in other domains*

This section describes in detail some of the modeling methods used in other domains, such as transportation, epidemiology, and disaster management, which can be leveraged for modeling occupant behavior at the urban scale. Specifically, this section focuses on describing advanced stochastic models, neural networks, and reinforcement learning among other modeling methods, which represent recent approaches to modeling human behavior in different domains. Table 1 provides a summary of these methods, their loss functions, and examples of their applications in different domains as well as some open source documentations of these implementations.

3.2.1. *Stochastic models*

3.2.1.1. Advanced hidden Markov models. Markov models have been used for different modeling purposes in other domains, including monitoring road traffic management [39], modeling individual's movement patterns [154] or behavior patterns [155], studying human movement behaviors during a disaster evacuation [121,122], etc. Chen et al. [39] proposed a novel approach to monitor traffic congestion based on social media data (Twitter). They have developed a unified statistical framework based on hinge loss Markov random fields. The proposed model was evaluated using Twitter[7] and INRIX[8] probe speed datasets from two major U.S. cities, and results showed that the

[7] https://twitter.com/
[8] https://inrix.com/

Table 1
List of Human behavior modeling methods and sample applications in different domains.

Models	Loss Function	Sample Application(s)
Advanced Hidden Markov Models (HMM)	Action-based cost function [121,122]$cost(\zeta\vert\phi) = \sum_{a\in\zeta}(\phi^{\top}\mathbf{f_a}) = \phi^{\top}\mathbf{f}$	• Traffic management [38,39] • Disaster management [121,122] • Traffic modeling using CDR [134] (https://github.com/morsiee/HMM-based-traffic-using-CDRs) • Predicting traffic congestion [135] (https://github.com/pavansaidiwakar/Vehicular-Traffic-Modelling-using-HMM) • Traffic classification [136]
Generalized models	Parameter estimation using ℓ1-norm orℓ2-norm loss$\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$; $\sum_{i=1}^{N}\vert y_i-\hat{y}_i\vert$	• (https://github.com/JRBANCEL/Traffic-HMM) • Spread of infectious disease [137] • Generalizing human mobility over different spatial scales [105,113,138]
Bayesian Neural Networks (BNN)	Variational free energy$F(\mathscr{D},\theta) = KL[q(\omega\vert\theta)\Vert P(\omega)] - \mathbb{E}_{q(\omega\vert\theta)}[\log P(\mathscr{D}\vert\omega)]$	• Trip purpose prediction [100] • Transportation safety studies [139] • Predicting trip duration [140]
Recurrent Neural Networks (RNN)	Multi-class cross-entropy loss [141]–[143]$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1-y_i)\cdot\log(1-p(y_i)))$Mean Square Error (MSE) loss [144]$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$	• (https://github.com/brendanhasz/tfp-taxi) • Predicting human mobility [141] (https://github.com/vonfeng/DeepMove) • User location prediction [143] (https://github.com/eXascaleInfolab/Flashback_code) • Predicting Citywide Crowd Dynamics at Big Events [144]
Graph Neural Networks (GNNs)	Mean Square Error (MSE) loss [145]$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$Binary Cross Entropy loss [146]$L = -\sum_i y_i \log(\hat{y}_i)$	• (https://github.com/deepkashiwa/DeepUrbanEvent) • Predicting users' next location [142] (https://github.com/yongqyu/STRNN) • Social recommender systems [147] • https://github.com/wenqifan03/GraphRec-WWW19) • Link prediction and community detection [146]
Reinforcement Learning (RL)	$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s^{'})\ U(D)}\left[\left(r+\gamma\max_{a^{'}}Q(s^{'},a^{'};\theta_i^{-}) - Q(s,a;\theta_i)\right)^2\right]$[148]	• (https://github.com/JiaxuanYou/P-GNN) • Online vehicle routing [149] • Traffic signal control [150,151] • EV charging navigation to minimize travel time and charging cost [152]
k-nearest neighbors (kNN), Support Vector Machine (SVM), Random Forest (RF) and Multiple Layered Perceptron (MLP)	Occupancy weighted distance (owd)$owd_{a,b,t,f,tm} = d_{a,b}(1-Cocp_{b,t,f,fm})min(owd_{a,b,tf,tm})$, $b\subseteq i, undelivered$	• Delivery route optimization [153]

proposed Markov approach can achieve accurate predictions. Furthermore, researchers [121,122] have developed Markov based models to understand and predict human mobility during a disaster.

In order to predict individuals' mobility behavior from sparse trajectory data, researchers presented a hierarchical and layered model based on a hidden markov model (HMM) framework [154]. Different from basic HMM, the proposed model is formulated as shown in Eq. (3.1):

$$P(\mathscr{x}^k|\lambda) = \sum_{i=1}^{seq_{max}} p(\mathscr{x}^k|s_i^k)*P(s_i^k) \tag{3.1}$$

where seq_{max} represents the maximum number of hidden states and s_i^k denotes sequences of hidden states within k-length in the proposed movement behavior model. Aggregating the above equation results in Eq. (3.2):

$$\mathrm{P}(\mathscr{x}^k|\lambda) = \sum_{i=1}^{seq_{max}} \left[\prod_{j=1}^{k} \mathrm{P}(\mathrm{x(j)})|\mathrm{s_i(j)})*\mathrm{P}(\mathrm{s_i(j)}|\mathrm{s_i(j-1)}, \mathrm{s_i(j-2)}, \cdots, 1 \right] \tag{3.2}$$

Based on a real-life mobility dataset from the Nokia Mobile Data Challenge, the above proposed HMM framework was evaluated, and results showed that the model outperformed existing studies when analyzing sparse trajectory traces. In another study [156], the author proposed a hybrid Markov-based model to predict human mobility behavior in a large Chinese city using data collected by the Long Term Evolution (LTE) network. The proposed model considers non-Gaussian and Spatio-temporal characteristics of the location data, it includes three components, which are Mobility Pattern Discovery, Variable-order Markov Predictor, and Users Similarity Calculation. Results showed more than 56% accuracy in predicting individuals' future movement.

3.2.1.2. Generalized models. Generalized models describe mobility patterns by fitting universal models or laws to empirical data [157]. These quantifiable models are termed "generalized" because they are formulated to explain and characterize human movements. Examples include Lévy flights, continuous-time random walk (CTRW), and gravity models. Using a comprehensive database of bank note trajectories to capture aspects of human mobility, Brockmann et al. [113] unveiled that the distribution of short time traveling distances decays as a power law, which can be described by Lévy flights. Lévy distribution has been used to model mobility patterns of vector-borne diseases (e.g., dengue outbreaks) in cities [111,112]. Specifically, Barmak et al. [111] modeled human mobility using a truncated Levy distribution to represent the distribution of displacements' lengths. However, Brockman [113] illustrated that describing human mobility as simple Levy flights is incomplete due to the strong spatial inhomogeneity (e.g., people are less likely to leave large cities than suburban areas). CTRW [158] was found to accurately characterize human mobility on large spatial scales, and have found applications in the modeling of human travel and thus the geographical spread of infectious diseases [113]. Unlike standard random walks in which the jumps are made periodically, in CTRW, the waiting times $\Delta t_1, \Delta t_2$ and the jump sizes $\Delta r_1, \Delta r_2, \ldots$ are modeled as mutually independent and identically distributed (i.i.d) random variables. Therefore, the total displacement after time t is $r = \sum_{i=1}^{N} \Delta r_i$ where N is the random number of jumps in the time interval (0, t). Studies [113,159] indicate that the probability density function characterizing human trajectories are fat-tailed and given by $P(\Delta r)\ |\Delta r|^{-1-\alpha}$ and $P(\Delta t)\ |\Delta t|^{-1-\beta}$ where $0 < \alpha \leq 2$ and $0 < \beta \leq 2$.

Gravity models provide insights on commuting flow and are often used to describe a random walk process with time-varying commuting fluxes between different destinations. Balcan et al. [137] analyzed mobility data from 29 different countries and found that a gravity model is able to reproduce commuting patterns up to 300 km. The authors then superimposed epidemic simulations to study commuting networks' effect on the spread of infectious diseases. Yan et al. [138] built on the gravity model to establish a universal model capable of explaining various human mobility behaviors, citing applications that include disease control, social stability, congestion alleviation, information propagation, and e-commerce. The standard gravity model typically assumes a power-law decay with distance

$$\mathrm{W_{ij}} \propto \frac{\mathrm{N_i N_j}}{exp(\mathrm{d_{ij}})} \tag{3.3}$$

where W_{ij} is the commuting flux from i to j N_i is the population at i N_j is the population at j and d_{ij} is the distance between i and j.

3.2.2. Neural networks

3.2.2.1. Bayesian neural network. Classical neural networks show a poor performance in predicting human behavior because of improper network structure and overfitting issues [139,160,161]. As such, the Bayesian neural network (BNN) provides a solution to this problem by introducing connections' weights as probabilistic distributions instead of using the point estimation approach, as shown in Fig. 2 [162]. The variable structure of BNN increased the models prediction accuracy from noisy data and provide results with confidence interval instead of single estimates obtained from the traditional neural network [163,164]. Therefore, the researchers used BNN in predicting travel time [164], secondary incidents [165], incident duration [166], battery status of unmanned vehicles [167], and zone capacity [163].

Bayes' theorem is used to define the posterior distribution of connection weights based on prior distribution and training data as shown in Eq. (3.4) [162]:

$$P(\mathscr{w}|\mathscr{D}) = \frac{P(\mathscr{D}|\mathscr{w})P(\mathscr{w})}{P(\mathscr{D})} \tag{3.4}$$

where $P(\mathscr{w}|\mathscr{D})$ and $P(\mathscr{w})$ are the posterior and prior weights' distribution, $P(\mathscr{D}|\mathscr{w})$ is the likelihood of observations, and $P(\mathscr{D})$ is a normalizing constant.

The posterior distribution is obtained through an optimization problem that minimizes the Kullback-Leibler divergence (KL divergence) by changing the parameters θ of weights' distribution $q(\mathscr{w}|\theta)$ as shown in Eqs. (3.5)–(3.7) [162,168]. This results in the variational free energy function shown in Eq. (3.8) [162]. More details regarding BNN and KL divergence are available in [162,169–171].

$$\theta^* = argmin_\theta KL[q(\mathscr{w}|\theta)\|P(\mathscr{w}|\mathscr{D}) \tag{3.5}$$

$$\theta^* = argmin_\theta \int q(\mathscr{w}|\theta) log \frac{q(\mathscr{w}|\theta)}{P(\mathscr{w})P(\mathscr{D}|\mathscr{w})} d_{\mathscr{w}} \tag{3.6}$$

$$\theta^* = argmin_\theta KL[q(\mathscr{w}|\theta)\|P(\mathscr{w})] - \mathbb{E}_{q(\mathscr{w}|\theta)}[logP(\mathscr{D}|\mathscr{w})] \tag{3.7}$$

$$F(\mathscr{D}, \theta) = KL[q(\mathscr{w}|\theta)\|P(\mathscr{w})] - \mathbb{E}_{q(\mathscr{w}|\theta)}[logP(\mathscr{D}|\mathscr{w})] \tag{3.8}$$

The probabilistic weights allow the BNN to successfully overcome overfitting issues. However, this approach considerably increases the number of model parameters and the computational cost [162]. Aiming to tackle this computational problem, Cui et al. [100] used the automatic differential variational inference (ADVI) method to estimate the parameters of a BNN model and predicts trip purposes using Google and Twitter data. The research conducted using the open source PyMC3 package [172]. They also integrated their BNN model with the elastic net method [173] to eliminate insignificant features from the training datasets and decrease the required computational cost by 75%. Furthermore, Park et al [165,166] used the inductive learning algorithm TREPAN to extract a comprehensive and understandable rule from the BNN network and support the decision making in incidents management.

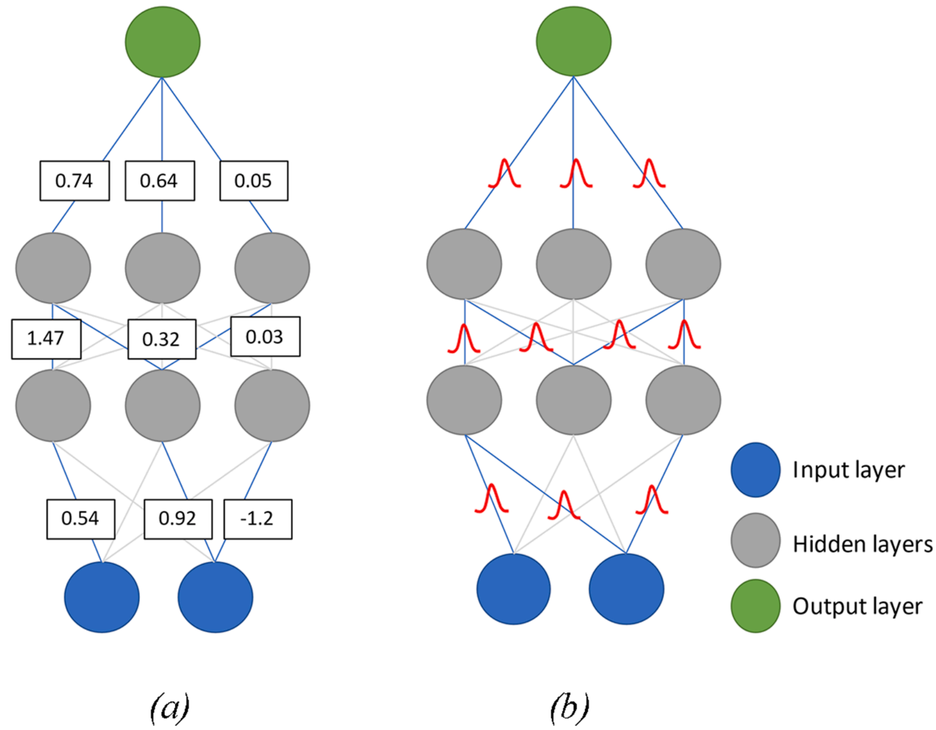

Fig. 2. (a) Neural network with point estimate values, (b) Bayesian neural network with connection weights defined as distributions.

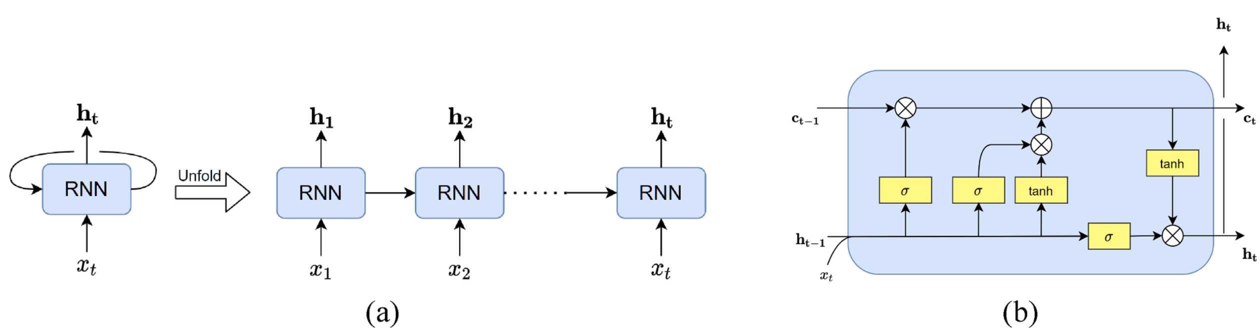

Fig. 3. (a) Illustration of a RNN. (b) Illustration of the LSTM architecture.

3.2.2.2. Recurrent neural networks. Recurrent Neural Networks (RNN) and variants such as Long Short-Term Memory (LSTM) [174] are widely used in modeling human mobility behaviors [141–145,175]. In RNN, the previous outputs are fed as the input to the current step (see Fig. 3a, where x is the input at each time step and h is the corresponding hidden state), making RNN suitable for modeling sequential inputs.

Although RNNs work effectively to model sequential data, they may suffer from the gradient vanishing issue which decreases performance when modeling long sequences. To overcome this issue, LSTM, a variant of RNN, is introduced. An LSTM cell has three gates (forget gate, input gate, and output gate) to control the information flow so that irrelevant information can be forgotten. With this unique memory mechanism, an LSTM is capable of learning long-term dependencies. In a basic LSTM network architecture (Fig. 3b), the hidden state is updated via:

$$\boldsymbol{h}_t = LSTM(\boldsymbol{h}_{t-1}, \mathrm{x}_t; \boldsymbol{W}) \tag{3.9}$$

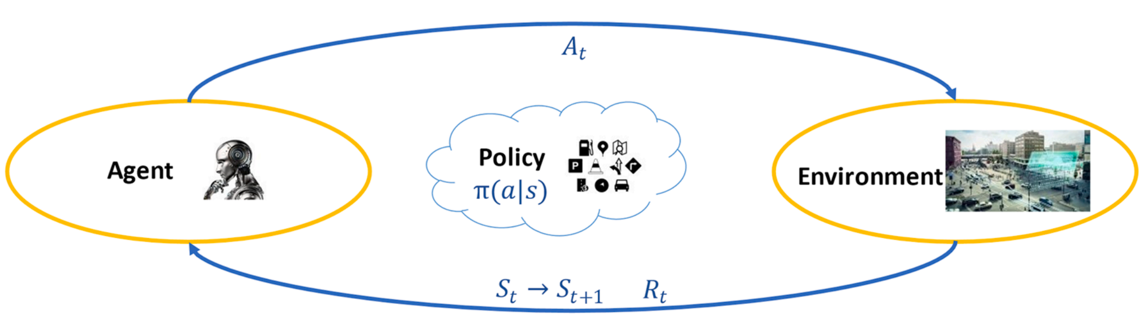

Fig. 4. The Markov decision process for RL agent.

where LSTM(•) represents an LSTM cell and **W** denotes its trainable weight matrices.

Inside each LSTM cell, the hidden state is determined by the input gate vector **i**, forget gate vector **f**, output gate vector **o**, and the cell state vector **c**. The recurrent updating process of the hidden state is described by the equations below:

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_{xi}x_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{ci}\boldsymbol{c}_{t-1} + \boldsymbol{b}_i)$$

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_{xf}x_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{cf}\boldsymbol{c}_{t-1} + \boldsymbol{b}_f)$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t\boldsymbol{c}_{t-1} + \boldsymbol{i}_t tanh(\boldsymbol{W}_{xc}x_t + \boldsymbol{W}_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c)$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_{xo}x_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{co}\boldsymbol{c}_{t-1} + \boldsymbol{b}_o)$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t tanh(\boldsymbol{c}_t) \tag{3.10}$$

where **W** terms are the weight matrices, $\sigma(\cdot)$ denotes the sigmoid activation function, and each **b** term with a subscript is the bias vector for the corresponding gating operation.

Based on RNN, various methods are proposed to model mobility patterns and predict future occupancy states of users based on history mobility data. For example, STRNN [142] is an earlier RNN-based model which introduces time-specific transition matrices and distance-specific transition matrices in a recurrent manner. However, the transition matrices result in hand-crafted functions with too many parameters which makes it difficult to train and apply. To overcome the above shortcoming, DeepMove [141] leverages the attention mechanism to capture the long-range dependencies from lengthy and sparse trajectories for mobility modeling. Furthermore, Flashback [143] is a more recent RNN based method for predicting human mobility, which takes the geographical context into consideration in its attention mechanism. At a city-wide scale, DeepUrbanEvent [144] is proposed to model the crowd dynamics at extreme events such as earthquakes and typhoons. The occupant behaviors (e.g., crowd flow) are described in an analogous manner to videos, which provides a new perspective for processing such behavior data with RNN. These popular RNN-based methods and their implementations are summarized in the above Table 1.

3.2.3. Reinforcement learning

Intelligent mobility systems provide broad and interactive support for other urban subsystems [176]. Some of the tasks for many mobility system analyses are formulated as an optimization problem such as optimal control problems [177]. Among the control strategies, reinforcement learning (RL) method relies on large amounts of data and advanced algorithms to optimize sequential actions with respect to controlling mobility behaviors.

An RL agent learns how to map situations to actions so as to maximize a numerical reward signal [178]. As shown in Fig. 4, an RL agent sequentially interacts with a system by taking an action A_t and receiving a reward signal $R_{t+1}$for some discrete time t. The agent chooses a deterministic or stochastic action that tries to maximize future returns under:

$$p(s^{'}, r|s, a) = Pr(S_{t+1} = s^{'}, R_{t+1} = r|S_t = s, A_t = a) \tag{3.11}$$

which is known as Markov decision process (MDP). The state S_t, a specific condition of the environment, is transferred to a subsequent state S_{t+1} with the transition probability:

$$p(s^{'}|s, a) = \sum_r p(s^{'}, r|s, a) = Pr(S_{t+1} = s^{'}|S_t = s, A_t = a) \tag{3.12}$$

A policyπ, a distribution over actions given states, defines the learning agent's way of behaving at a given time:

$$\pi(a|s) = Pr(A_t = a|S_t = s) \tag{3.13}$$

For value-based algorithms, a state-action value is usually estimated by a deep Q-network (DQN) [179]. A world leading school, DeepMind[9], continually pushes the boundaries of the technique, advancing the field of artificial intelligence. On the other hand, policy-based algorithms build a representation of a policy by mapping s to a. Policy-based methods have better convergence properties and are more effective in high-dimensional action space. For a parametrized policy, the objective for policy-based algorithms is to find the optimal policy parameters:

$$\theta^{*} = argmax_{\theta}J(\theta) = argmax_{\theta}Pr(\tau;\theta)R(\tau) \tag{3.14}$$

where $R(\tau)$is the sum of rewards over a trajectory $\tau$and $Pr(\tau;\theta)$ is the probability over trajectories when executing policy. Policy-based methods give more attention to sample efficiency. One toolkit for developing and comparing algorithms was developed by OpenAI Gym[10].

A recent review work has demonstrated the state-of-the-art for RL applied in mobility [180]. In addition to combinatorial vehicle routing optimization [149], there are several scenarios where deep RL can outperform other methods when the problem is complex. For example, Deep Q-network (DQN) delivers optimal performance in a single-intersection scenario [150]. Intelligent traffic light settings have also been modeled by multi-agent RL [151], and the controller was able to make improvements in time delay and speed in urban transportation. Another example concerns EV optimal charging navigation [152]. Deep RL is usually conducted for minimizing the total travel time and the charging cost by considering actual battery consumption and driving cost. Generally, RL provides sustainable dynamic routing solutions for urban planners, individual car users and the transport sector. The method integrates a number of infrastructures data sources to improve the transport policies at both district level and urban level. Occupancy schedule in buildings will be more predictable with growing realization of autonomous electric vehicles and seamless mobility. The advantage of RL method is that the agent can handle complex environment by quickly developing adaptive policies. However, data acquisition at large scale for training RL agent is still a challenging task.

3.2.4. Network modeling

The modeling of social and mobility networks continues to gain importance in a variety of fields ranging from epidemiology [116,181], to social and community detection [59,60,146,182], to user movement and behavior understanding [52,183]. In the Simultaneous Extraction of Context and Community (SECC) model proposed by Nguyen et al. [60], each context is represented as a multinomial distribution, which indicates the participating level of the users to that context. To detect communities, SECC then computes the clusters of multinomial distribution to discover proximity contexts and users. Recently, Graph Neural Networks (GNNs), which have shown great ability in learning and modeling graph data, are widely applied in social network modeling. For example, GraphRec [147] is designed for social recommender systems to learn user representations from different perspectives. Two graphs, namely user-user graph and user-item graph, are built to model users from the social perspective and understand the interactions between users and items. Consequently, GraphRec considers heterogeneous strengths of social relations by modeling the two graphs coherently. For community detection, Position-aware Graph Neural Networks (PGNNs) are proposed to compute position-aware node embeddings in [146]. By capturing locations of nodes with respect to the selected anchor nodes, this special class of GNNs is able to incorporate node positional information, while retaining inductive capability and utilizing node features.

GNNs, first proposed in [184], are an extension of neural networks

[9] https://github.com/deepmind

[10] https://github.com/openai/gym Some other useful repositories can be found at https://lilianweng.github.io/lil-log/2018/04/08/policy-gradient-algorithms and https://github.com/cuhkrlcourse/RLexample

that allows for processing of data represented in graph domains. A graph is made up of nodes, which are defined by unique features, and edges, which represent the relationships between nodes. Thus a node is defined not only by its own features but also by the nodes surrounding it. A GNN aims to learn a state embedding x_n for each node $n$based on the information contained in the neighborhood of n. This state can be represented by a parametric function f_w, called the local transition function, which expresses the dependence of x_n on its neighborhood. The output on produced by node n can then be expressed by a local output function g_w as follows:

$$\boldsymbol{x}_n = \boldsymbol{f}_w\left(\boldsymbol{l}_n, \boldsymbol{l}_{co[n]}, \boldsymbol{x}_{ne[n]}, \boldsymbol{l}_{ne[n]}\right)$$

$$\boldsymbol{o}_n = \boldsymbol{g}_w(\boldsymbol{s}_n, \boldsymbol{l}_n) \tag{3.15}$$

where l_n, $l_{co[n]}$, $x_{ne[n]}$, and $l_{ne[n]}$ are the label of n, the labels of its edges, the states, and the labels of the nodes in the neighborhood of n, respectively. The outputs and states can be computed iteratively using

$$\boldsymbol{x}_n(t+1) = \boldsymbol{f}_w\left(\boldsymbol{l}_n, \boldsymbol{l}_{co[n]}, \boldsymbol{x}_{ne[n]}(t), \boldsymbol{l}_{ne[n]}\right)$$

$$\boldsymbol{o}_n(t) = \boldsymbol{g}_w(\boldsymbol{x}_n(t), \boldsymbol{l}_n), n \in \boldsymbol{N}. \tag{3.16}$$

The computation of f_w and g_w can be interpreted as the workings of a neural network; therefore, each node of the graph can be replaced with a neural network unit. In [47], the authors created a graph in which nodes represent individual locations and edges represent different relationships depending on their domain: edges in the spatial domain represent location-to-location movement and edges in the temporal domain represent connections to past days. The spatiotemporal GNN, trained to minimize mean squared logarithmic error, outperforms LSTM and Seq2Seq baselines in predicting COVID-19 caseloads for 20 counties in the U.S. The authors of [118] use a GNN in which nodes correspond to a country's regions and edges represent interregional movement; they then apply a model-agnostic *meta*-learning approach to transfer knowledge from one country's model to another. This approach is shown to outperform other COVID-19 forecasting techniques in four European countries.

3.2.5. Agent-based modeling

An agent-based approach, such as those as used by [185,98], or [46], can be used to model and simulate human mobility and interaction by representing each individual as a software agent. Each agent is characterized by a number of attributes which inform its movement and behavior. The types of agents used in the model and their attributes vary widely depending on the application. In [98], there are two classes of agents: occupants and buildings. Occupant-class agents are assigned to a host building and characterized by their occupant energy conservation index (OECI), susceptibility to peer pressure, and zealotry (probability of changing behavior based on feedback). Building-class agents are characterized by their energy intensity (EI), elasticity (percentage of their energy use which can vary depending on occupant behavior), and building energy conservation index (BECI). Both occupants and buildings have their own social network (other agents to which they are connected). The actions of the occupant-agents inside each building directly impact the building's attributes over time as follows:

$$BECI_i^{(t)} = \sum_{j=1}^{N}\left(OECI_j^t / \mathrm{N}\right) \tag{3.17}$$

$$\mathrm{EI}_i^{(t)} = \mathrm{EI}_i^{(0)}\left[1 + 2\cdot elasticity_i\left(0.5 - BECI_i^t\right)\right] \tag{3.18}$$

Additionally, each occupant-agent's behavior is influenced by the behaviors of the agents in its social network and by feedback from its host building's social network:

$$OECI_i^{(t+1)} = (1 - Zealot_i \times Susceptibility_i) \times OECI_i^{(t)} + Zealot_i \times Susceptibility_i \times \sum_{j=1}^{N}\left(\frac{BECI_j^t}{\mathrm{N}}\right)_i \tag{3.19}$$

In [46], ABM is used to simulate the progression of the H1N1 virus outbreak in Mexico. In this model, each agent represents an individual and has its behavior defined by three models: a mobility model, a social network model, and a disease progression model. The mobility model is formed by dividing each day into a set of non-overlapping equal-length time slots and assigning the agent to a specific location at each time slot using CDR data. The social network is defined as the set of agents with whom a particular agent has reciprocal contact at least once. The

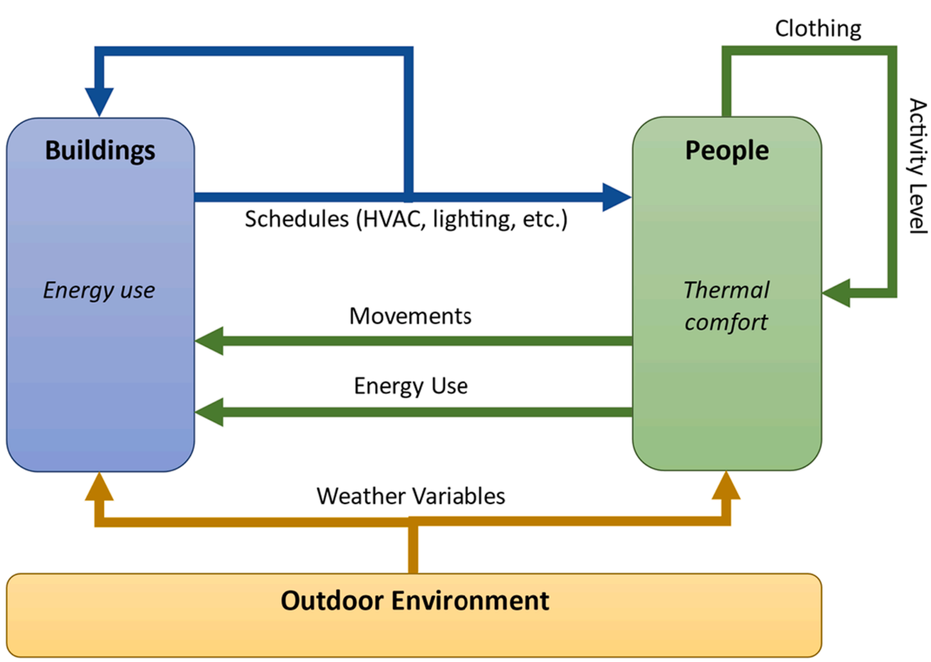

Fig. 5. Agent interactions. (Italicized statements indicate quantities of interest for each type of agent; arrows indicate how each type of agent affects others through various actions.)

Table 2
Common performance metrics, adapted from [186–188,195].

Metric (R: Regression, C: Classification, P: Probabilistic)	Description and/or formula
R - MSE (Mean Squared Error) and RMSE (Root Mean Squared Error)	$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2, RMSE = \sqrt{MSE}$
R - RMSLE (Root Mean Squared Logarithmic Error) and CVRMSE (Coefficient of Variation of Root Mean Squared Error)	$RMSLE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\log\frac{1+y_i}{1+\hat{y}_i}\right)^2}, CVRMSE = \frac{RMSE}{\bar{\bar{y}}}$
R - MAE (Mean Absolute Error) and MBE (Mean Bias Error)	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i-\hat{y}_i\rvert, MBE = \frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)$
R - Coefficient of determination R-Squared (R^2)	$R^2 = 1-\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y}_i)^2}, where\bar{y} = \frac{1}{N}\sum_{i=1}^{N}y_i$
C - Accuracy and error rate	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, Errorrate = 1-Accuracy$
C - Precision and recall	$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}$
C - F1-Score and FBeta-Score	$F_\beta = \frac{(1+\beta)\times Precision\times Recall}{\beta^2\times Precision+Recall}, F_1 = F_{\beta=1} = \frac{2\times Precision\times Recal}{Precision+Recal}$
C - AUC (Area Under the Curve) and ROC (Receiver Operating Characteristic) curve	ROC is a plot of the true positive (*TP*) rate against the false positive (*FP*) rate for various threshold values. AUC is the area under the ROC curve.
C - Confusion matrix	Tabular representation of the predictions for each class against the ground truth values.
P - Prediction Interval Coverage Probability (PICP)	$PICP = \frac{1}{N}\sum_{i=1}^{N}c_i, C_i = \begin{cases} y_i = 1, if y_i \in [L_i, U_i] \\ y_i = 0, otherwise\end{cases}$ PICP gives an indication of the number of measurements that fall within the prediction intervals.
P - Coverage Width Criterion (CWC)	$CWC = PINMW + \gamma(PICP)e^{-\eta(PICP-\mu)} PINMW = \frac{1}{N\cdot R}\sum_{i=1}^{N}(U_i-L_i)^2, \gamma(PICP) = \begin{cases}0, if PICP \geq \mu \\ 1, otherwise\end{cases}$ μ is the nominal confidence level and its value can be determined based on the confidence level $(1-\alpha)$% associated with the *PIs.CWC* provides an evaluation of the prediction intervals between two conflicting viewpoints, correctness represented by the prediction interval coverage probability (*PICP*) and sharpness represented by the prediction interval normalized mean width (*PINMW*)
P-Continuous Ranked Probability Score (CRPS)	$CPRS(F,y) = -\int_{-\infty}^{\infty}(F(Y)-\mathbb{1}\{Y\geq y\})^2 d_Y CPRS(F,y) = \frac{1}{2}E_F\lvert Y-Y'\rvert - E_F\lvert Y-y\rvert$ *CRPS* is defined as the integral of the Brier scores for the associated binary probability forecasts at all real-valued thresholds. It provides a measure of the discrepancy between the predictive distribution and a single observation taking into account the distribution's sharpness. *Y* and *Y'* are independent copies of a random variable with distribution function *F* and finite first moment.
y Observed value *N* Number of observations *R* Range of observed values *FN* False negative *TN* True negative	$\hat{y}$ Predicted value *U*, *L* Lower and upper bound of prediction interval *FP* False positive *TP* True positive *F* Cumulative distribution function (CDF) of random variable*Y*

probability that two agents who are in the same location at the same time and are part of the same social network can transmit the virus to each other is computed as follows:

$$p_i = 1 - exp\left(\tau\sum_{r\in R}N_r ln(1 - rs_i\rho)\right) \quad (3.20)$$

where τ is the duration of exposure, R is the set of infective agents, N_r is the number of agents with infectivity r, s_i is the agent's susceptibility, and ρ is the disease's basic transmissibility.

In [185], the authors simulated the movement of people on a campus, their thermal indoor and outdoor comfort levels, and the energy performance of the buildings they occupy. There are two types of agents in the proposed ABM, "people" and "buildings", interacting with the outdoor environment and impacting each other's behavior as summarized in Fig. 5. The execution of the proposed model occurs along three submodels that cover people movement, thermal comfort calculation, and building energy estimation using surrogate models of building energy performance.

In [43], the authors introduced a graph-coupled hidden Markov model (GCHMM), a discrete-time multi-agent model in which multiple HMM agents are linked graphically with edges representing interactions between agents at time *t*. Each agent, which represents an individual, has an attribute (or state) associated with their stage of disease progression (susceptible/infectious) and multiple attributes describing their current symptoms. The GCHMM formulation introduced in the study allows for fitting a wide range of agent-based models to large social network datasets.

3.3. Performance evaluation

Various metrics exist and can be used to evaluate the accuracy of models [186]–[188]. A summary of metrics is presented in Table 2, with a distinction made between metrics used for regression, classification, and probabilistic problems. It is important to note that most of the metrics are analytical and can be computed with specific formulas, as shown in the table (e.g., mean squared error, accuracy, and F1-score). Others, such as ROC curve and the confusion matrix, are rather graphical and require interpretation to extract knowledge about the performance of the models. Metrics evaluating the performance of probabilistic predictions should take into account both its reliability and sharpness. Examples of metrics that consider both aspects are the coverage width criterion (CWC) and the continuous ranked probability score (CRPS).

Most of the metrics listed in Table 2 have been used for the evaluation of different occupant behavior modeling methods across the disciplines covered in the previous sections. For instance, Jiefan et al. [79] used the MBE and the CVRMSE metrics to compare actual and predicted HVAC energy consumption generated using real-time occupancy data. Using CWC and PICP, Chong et al. [189] compared different spatial resolutions of occupancy data by evaluating prediction intervals of building energy consumption. Karimzadeh et al. [190] applied MAE to test their predictions of mobile users' future locations. Ghosh and Ghosh [154] used accuracy, precision, and recall to test the predictive capabilities of a model that predicts people movements based on sparse trajectory traces. In a similar application of mobility prediction, Klous et al. [191] used the R^2 metric when comparing predictions to ground truth data. Li and Dong [131] combined accuracy and ROC to evaluate short-term predictions of occupancy in residential buildings. Likewise, F1-score has been used to evaluate predictions of human mobility prediction [192], user's next point of interest [193], and dwelling times [194].

Although Probabilistic predictive models of occupancy are often proposed to more realistically account for occupant behavior's stochastic nature, it is somewhat surprising that probabilistic predictions are often simplified and evaluated similarly to deterministic forecasts. Ideally, the probabilistic predictions should have a high coverage of measured values (correctness) but still correctly represent measurement variability (informativeness). The CWC and the CRPS are examples of two metrics illustrated in Table 2 that considers both the correctness and informativeness of the prediction intervals.

4. Future opportunities

This section answers the following research question: What are the potential future research directions for building design, operation, and policies at a community level, with enhanced data sources and modeling methods from other domains? In other words, how can we bridge the data source and methodology gap between building science and beyond. We will discuss this in the following 3 opportunities.

4.1. Opportunity 1. District heating and cooling system design and operation

From current review, occupant behavior has been introduced for district heating and cooling system design and operation. However, current research mainly focuses on community scale, with a simplified occupancy profile. With the increasing needs of energy efficiency/flexibility, and the desire to integrate low-temperature heat resources, 5th generation district heating and cooling (5GDHC) technology has been recently proposed. This technology is characterized by low temperature supply (i.e. close to ground temperature), bi-directional operation (i.e. providing simultaneous heating and cooling services), decentralized energy flows (i.e. enabling multiple heat sources and heat sinks in the network), and heat sharing (i.e. recovering waste heat and sharing with different users [196]). 5GDHC technology has moved to a consumer/prosumer-centric perspective, where occupant behavior plays an important role. However, at the moment, it rarely considers different occupant behavior profiles in the modeling of energy flows of different players/agents (e.g. data centers, shopping malls, and residential buildings). The implementation of 5GDHC will trigger the heating/cooling sharing in a local energy market, and this will influence occupant behavior and thus buildings demand profiles, when energy needs of one building can be balanced with the surplus of another. Further, the local energy supply will be influenced, because less heating or cooling will be needed from the infrastructure. Potential agent-based models and reinforcement learning models can be applied for the simulation of 5GDHC's design and operation. What's more, the condition of climate change has higher requirement for the energy use and carbon emission, where the insight of variety of occupant behavior and the prediction result of occupant behavior are significant for system design and operation to combine with the renewable resources. This can be foreseen as an essential opportunity for future research.

4.2. Opportunity 2. Grid-interactive community energy planning and management

Traditional community energy planning and management considers urban land use and infrastructure-level energy management [96]. Future smart communities, with connected buildings, increasing penetration of distributed energy resources (DERs), electric vehicles (EVs), battery energy storage systems (BESS), and blockchain-enabled peer-to-peer (P2P) energy trading at the individual building level, require an optimal and distributed coordination of a cluster of buildings, DERs, and smart grid while considering human behavior and mobility. In addition, such communities should provide load flexibility and demand side management to the grid. Unfortunately, existing measurement of occupant behavior is not cost effective and scalable at a community and urban scale. Lack of knowledge and capability of modeling occupant behavior impedes further load flexibility and demand benefits from the aforementioned urban scale energy systems. The development and engagement of smart meters and social media have opened up a new paradigm for community energy planning and underscore the need for a holistic engineering framework to model a new energy infrastructure of a community. However, optimal energy planning studies at the community level that take into account smart meters, connected buildings, energy trading, social media, and human behavior data are still rare, and could be future research opportunities.

4.3. Opportunity 3: Resilient building and community design

The concept of resilience is gaining increasing interest in the built environment in response to disaster management and the recent COVID-19 pandemic, which has transformed the manner in which we live and interact in the built environment [197]. While still filled with uncertainties, the pandemic has highlighted once more the need to incorporate resilience concepts into urban planning and management which are crucial in mitigating the impacts of disasters and unforeseeable events. Originating from the study of ecological systems, resilience is defined as a system's ability to persist and absorb changes and disturbances [198]. For technical systems, it is the ability of a system to withstand and fulfill its functional requirements and recover to the expected performance requirements during and after the occurrence of an unforeseeable and disruptive event [199]. The wealth of available data and the many data-driven models are proving to be useful tools to supervise the functioning of an urban system (e.g., [200]), to fast detect occurrence of unforeseeable events (e.g., a terrorist attack [201]), and monitor the propagation of disruptive events (e.g., the spread of infectious diseases [202]). However, it is difficult for these models to consider the spatial–temporal relationships between humans and buildings [203]. For example, models considering human behavior and infectious disease transmission have mostly been for a single space within a building [202]. To improve resilience, cities need systems that can fast detect anomalies due to disruptive events, flexibly adapt to new conditions, and provide effective emergency response to recover a minimum functionality level. Knowledge of real-time human mobility, building occupancy, and energy use provides the ability to simulate different interventions and evaluate their effectiveness in reducing the impact of disruptive events like infectious disease outbreaks, heatwaves, flooding, and earthquakes. From this review, we found significant differences in modeling methods and data requirements in the field of building science compared to disciplines such as transportation, epidemiology, disaster management and social recommendation systems. Integrating measurements of human mobility, building occupancy, and energy uses with mathematical models of infection spread or local climate alteration shows promise and presents future research opportunities.

To reinvigorate life in cities post-pandemic, a major overhaul in designing and repurposing buildings and public spaces is required. Indeed, the traditional spaces inside a house showed to be unsuitable and even the role and functions of an apartment in a post-pandemic society need to be rediscussed to increase resilience by properly integrating smart working and educational needs, spaces for physical exercise, and outdoor or semi-outdoor spaces, like balconies and terraces, as buffer zones for mental and social well-being [204]. Furthermore, the citizens as a key stakeholder in cities need to be involved in the decision-making process. Participatory design and large scale citizen science projects [205] could be explored to involve humans in the loop, both to empower citizens as well as to involve the "wisdom of the crowds" in the algorithmic-driven society [206]. Taken together, this review suggests that the integration between different modeling methods taking into consideration data sources at varying spatial and temporal resolutions would be a fruitful area for future work.

5. Conclusion

Based on a comprehensive review of urban scale occupant behavior modeling methods used in Building Science and other domains, it can be concluded that: 1) urban scale building applications still rely on modeling occupant behavior at the individual building level, not at the urban scale; 2) emerging data sources and methodologies in transportation, epidemiology, disaster management, and marketing/promotion domains can potentially meet the modeling requirement of building applications discussed in this paper.

Advanced modeling approaches employed in other domains such as reinforcement learning, social/network modeling, and agent-based modeling have proven to be capable of learning and predicting the behavior of people across physical and digital (cyber) spaces. These methods may be adapted to building scale applications to learn not only human mobility patterns, but also their behaviors and interactions with buildings and each other within buildings. Among those modeling methods, neural networks and graphical network analysis have gained much attention and shown promising results. This paper also summarized common performance evaluation metrics that can be adopted to evaluate the accuracy of models in the building science.

Given the capabilities of new modeling methods and approaches, this paper discusses three research opportunities including new district heating and cooling system design and operation, grid-interactive community energy planning and management, and resilient building and community design. All three areas conclude that a more accurate and detailed modeling of occupant behavior that considers demographic information, behavior changes, and social network analysis will transform traditional building design and operation, and open up a new paradigm for future research.

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.

References

[1] Mahdavi A. People in building performance simulation. Book chapter. In: Hensen J, Lamberts R, E. editors. Building performance simulation for design and operation, London ; New York: Spon Press, 1996, p. 507.
[2] Ren X, Yan D, Wang C. Air-conditioning usage conditional probability model for residential buildings. Build Environ Oct. 2014;81:172–82.
[3] Yasue R, Habara H, Nakamichi A, Shimoda Y. Modeling the occupant behaviour relating to window air conditioner operation based on survey results. 13th Conf Int Build Perform Simul Assoc., 2013.
[4] Chang W-K, Hong T. Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data. Build Simul Oct. 2013;6 (1):23–32.
[5] Dietz T, Gardner GT, Gilligan J, Stern PC, Vandenbergh MP. Household actions can provide a behavioral wedge to rapidly reduce US carbon emissions. Proc Natl Acad Sci Oct. 2009;106(44):18452–6.
[6] Carlucci S, et al. Modeling occupant behavior in buildings. Build Environ May 2020;174:106768.
[7] Mirakhorli A, Dong B. Occupancy behavior based model predictive control for building indoor climate—A critical review. Energy Build Oct. 2016;129:499–513.
[8] Park JY, et al. A critical review of field implementations of occupant-centric building controls. Build Environ Oct. 2019;165:106351.
[9] Xu X, Culligan PJ, Taylor JE. Energy Saving Alignment Strategy: Achieving energy efficiency in urban buildings by matching occupant temperature preferences with a building's indoor thermal environment. Appl Energy Oct. 2014;123:209–19.
[10] Happle G, Fonseca JA, Schlueter A. A review on occupant behavior in urban building energy models. Energy Build. vol. 174. Elsevier Ltd, pp. 276–292, 01-Sep-2018.
[11] Zhang L, Jing J, Duan M, Qian M, Yan D, Zhang X. Analysis of district cooling system with chilled water thermal storage in hot summer and cold winter area of China. Build Simul 2020;13(2):349–61.
[12] Noussan M, Jarre M, Poggio A. Real operation data analysis on district heating load patterns. Energy Oct. 2017;129:70–8.
[13] Qian M, Yan D, Liu H, Berardi U, Liu Y. Power consumption and energy efficiency of VRF system based on large scale monitoring virtual sensors. Build Simul Oct. 2020;13(5):1145–56.
[14] Gang W, Augenbroe G, Wang S, Fan C, Xiao F. An uncertainty-based design optimization method for district cooling systems. Energy Oct. 2016;102:516–27.
[15] Hu S, Yan D, Guo S, Cui Y, Dong B. A survey on energy consumption and energy usage behavior of households and residential building in urban China. Energy Build Aug. 2017;148:366–78.
[16] Zhou X, Yan D, Feng X, Deng G, Jian Y, Jiang Y. Influence of household air-conditioning use modes on the energy performance of residential district cooling systems. Build Simul Oct. 2016;9(4):429–41.
[17] An J, Yan D, Hong T, Sun K. A novel stochastic modeling method to simulate cooling loads in residential districts. Appl Energy Oct. 2017;206:134–49.
[18] Dong B, Yan D, Li Z, Jin Y, Feng X, Fontenot H. Modeling occupancy and behavior for better building design and operation—A critical review. Build Simul. vol. 11, no. 5. Tsinghua University Press, pp. 899–921, 01-Oct-2018.
[19] Qian M, Yan D, An J, Hong T, Spitler JD. Evaluation of thermal imbalance of ground source heat pump systems in residential buildings in China. Build Simul 2020;13(3):585–98.
[20] Barbour E, Davila CC, Gupta S, Reinhart C, Kaur J, González MC. Planning for sustainable cities by estimating building occupancy with mobile phones. Nat Commun Dec. 2019;10(1):1–10.
[21] Dong B, Wu W, Wang Q, Abdelmutaal S, Prakash V. Derive Urban Scale Occupant Behavior Profiles From Mobile Position Data: A Pilot Study. Build Simul 2019; 2020:113–20.
[22] Wu W, et al. A novel mobility-based approach to derive urban-scale building occupant profiles and analyze impacts on building energy consumption. Appl Energy Nov. 2020;278:115656.
[23] Kang X, Yan D, Sun H, Jin Y, Xu P. An Approach For Obtaining And Extracting Occupancy Patterns In Buildings Based On Mobile Positioning Data. Build Simul 2019;2020:2364–70.
[24] Dong B, Lam KP. A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. Build Simul Feb. 2014;7(1):89–106.
[25] Salim FD, et al. Modelling urban-scale occupant behaviour, mobility, and energy in buildings: A survey. Build Environ Oct. 2020;183:106964.
[26] Taha AF, Gatsis N, Dong B, Pipri A, Li Z. Buildings-to-Grid Integration Framework. IEEE Trans Smart Grid 2019;10(2):1237–49.
[27] Liu Y, et al. Coordinating the operations of smart buildings in smart grids. Appl Energy Oct. 2018;228:2510–25.
[28] Mirakhorli A, Dong B. Model predictive control for building loads connected with a residential distribution grid. Appl Energy Oct. 2018;230:627–42.
[29] Jin Y, et al. Appliance use behavior modelling and evaluation in residential buildings: A case study of television energy use. Build Simul 2020;13(4):787–801.
[30] Chen Y, et al. Quantification of electricity flexibility in demand response: Office building case study. Energy Dec. 2019;188:116054.
[31] Hu M, Xiao F. Quantifying uncertainty in the aggregate energy flexibility of high-rise residential building clusters considering stochastic occupancy and occupant behavior. Energy Mar. 2020;194:116838.
[32] Sun K, Hong T. A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures. Energy Build Jul. 2017;146: 383–96.
[33] Lovati M, Zhang X, Huang P, Olsmats C, Maturi L. Optimal Simulation of Three Peer to Peer (P2P) Business Models for Individual PV Prosumers in a Local Electricity Market Using Agent-Based Modelling. Buildings Oct. 2020;10(8):138.
[34] Lovati M, Dallapiccola M, Adami J, Bonato P, Zhang X, Moser D. Design of a residential photovoltaic system: the impact of the demand profile and the normative framework. Renew Energy Oct. 2020;160:1458–67.
[35] Hu S, Yan D, Azar E, Guo F. A systematic review of occupant behavior in building energy policy. Build Environ Oct. 2020;175:106807.
[36] Pan B, Zheng Y, Wilkie D, Shahabi C. Crowd sensing of traffic anomalies based on human mobility and social media. In: in the 21st ACM SIGSPATIAL International Conference; 2013. p. 344–53.
[37] Itoh M, Yokoyama D, Toyoda M, Tomita Y, Kawamura S, Kitsuregawa M. Visual fusion of mega-city big data: An application to traffic and tweets data analysis of Metro passengers. In: Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014, 2014, p. 431–440.
[38] Lin L, Li J, Chen F, Ye J, Huai J. Road Traffic Speed Prediction: A Probabilistic Model Fusing Multi-Source Data. IEEE Trans Knowl Data Eng Jul. 2018;30(7): 1310–23.
[39] Chen PT, Chen F, Qian Z. Road Traffic Congestion Monitoring in Social Media with Hinge-Loss Markov Random Fields. In: Proceedings - IEEE International Conference on Data Mining, ICDM, 2014, vol. 2015-January, no. January, p. 80–89.
[40] Jiang J, Lin F, Fan J, Lv H, Wu J. A Destination Prediction Network Based on Spatiotemporal Data for Bike-Sharing. Complexity Oct. 2019;2019:1–14.
[41] Tang J, Liang J, Zhang S, Huang H, Liu F. Inferring driving trajectories based on probabilistic model from large scale taxi GPS data. vol. 506, p. 566–577, Sep. 2018.
[42] Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. Proc R Soc London Ser A, Contain Pap a Math Phys Character, Oct 1927;115(772):700–21.
[43] Dong W, Pentland A, Heller KA. Graph-Coupled HMMs for Modeling the Spread of Infection. arXiv:1210.4864 [physics, stat], Oct. 2012.

[44] Dong W, Heller K, Pentland A. (Sandy). Modeling Infection with Multi-agent Dynamics. 2012, p. 172–9.
[45] Madan A, Cebrian M, Lazer D, Pentland A. Social sensing for epidemiological behavior change. In: Ubicomp '10: The ACM Conference on Ubiquitous Computing, 2010; 2010. p. 291–300.
[46] Frias-Martinez E, Williamson G, Frias-Martinez V. An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing; 2011. p. 57–64.
[47] Kapoor A, et al. Examining COVID-19 Forecasting using Spatio-Temporal Graph Neural Networks. arXiv2007.03113 [cs], Oct. 2020.
[48] Wang Q, Taylor JE. Patterns and Limitations of Urban Human Mobility Resilience under the Influence of Multiple Types of Natural Disaster. PLoS ONE Oct. 2016;11 (1):e0147299.
[49] Wang Q, Taylor JE. Process map for urban-human mobility and civil infrastructure data collection using geosocial networking platforms. J Comput Civ Eng 2016;30(2):4015004.
[50] Wilson R, et al. "Rapid and Near Real-Time Assessments of Population Displacement Using Mobile Phone Data Following Disasters: The 2015 Nepal Earthquake. PLoS Curr 2016;8:Oct.
[51] Chen Y, Wang Q, Ji W. Rapid Assessment of Disaster Impacts on Highways Using Social Media. J Manag Eng Oct. 2020;36(5):4020068.
[52] Ren Y, Tomko M, Salim FD, Chan J, Sanderson M. Understanding the predictability of user demographics from cyber-physical-social behaviours in indoor retail spaces. EPJ Data Sci Oct. 2018;7(1):1–21.
[53] Kaur M, Salim FD, Ren Y, Chan J, Tomko M, Sanderson M. Shopping intent recognition and location prediction from cyber-physical activities via Wi-Fi logs. 2018, p. 130–9.
[54] Kaur M, Salim FD, Ren Y, Chan J, Tomko M, Sanderson M. Joint modelling of cyber activities and physical context to improve prediction of visitor behaviors. ACM Trans Sens Networks 2020;16(3):1–25.
[55] Ren Y, Tomko M, Salim FD, Chan J, Clarke CLA, Sanderson M. A Location-Query-Browse Graph for Contextual Recommendation. IEEE Trans Knowl Data Eng Feb. 2018;30(2):204–18.
[56] Kim S, Lee JG. Utilizing In-store Sensors for Revisit Prediction. In: Proceedings - IEEE International Conference on Data Mining, ICDM, 2018, vol. 2018-November, p. 217–26.
[57] Hang M, Pytlarz I, Neville J. Exploring student check-in behavior for improved point-of-interest prediction. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2018. p. 321–30.
[58] Wang R, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014, p. 3–14.
[59] Nguyen T, Nguyen V, Salim FD, Phung D. SECC: Simultaneous extraction of context and community from pervasive signals. 2016, p. 1–9.
[60] Nguyen T, Nguyen V, Salim FD, Le DV, Phung D. A Simultaneous Extraction of Context and Community from pervasive signals using nested Dirichlet process. Pervasive Mob Comput Oct. 2017;38:396–417.
[61] Li W, Tian Z, Lu Y, Fu F. Stepwise calibration for residential building thermal performance model using hourly heat consumption data. Energy Build Oct. 2018; 181:10–25.
[62] Zhang Q, Tian Z, Ma Z, Li G, Lu Y, Niu J. Development of the heating load prediction model for the residential building of district heating based on model calibration. Energy Oct. 2020;205:117949.
[63] Kazas G, Fabrizio E, Perino M. Energy demand profile generation with detailed time resolution at an urban district scale: A reference building approach and case study. Appl Energy Oct. 2017;193:243–62.
[64] van den Brom P, Hansen AR, Gram-Hanssen K, Meijer A, Visscher H. Variances in residential heating consumption – Importance of building characteristics and occupants analysed by movers and stayers. Appl Energy Oct. 2019;250:713–28.
[65] Qian M, Yan D, Zhang L, Zhou X. Influence of occupant behaviour on oversizing issue of heat pumps for residential district in Hot Summer and Cold Winter zone of China. Procedia Eng Oct. 2017;205:2434–41.
[66] Huchuk B, O'Brien W, Sanner S. A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data. Build Environ Oct. 2018;139:199–210.
[67] Gianniou P, Reinhart C, Hsu D, Heller A, Rode C. Estimation of temperature setpoints and heat transfer coefficients among residential buildings in Denmark based on smart meter data. Build Environ Oct. 2018;139:125–33.
[68] Gianniou P, Liu X, Heller A, Nielsen PS, Rode C. Clustering-based analysis for residential district heating data. Energy Convers Manag Oct. 2018;165:840–50.
[69] Calikus E, Nowaczyk S, Sant'Anna A, Gadd H, Werner S. A data-driven approach for discovering heat load patterns in district heating. Appl Energy, vol. 252, p. 113409, Oct. 2019.
[70] Huchuk B, Sanner S, O'Brien W. Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. Build Environ Oct. 2019;160:106177.
[71] Peng C, Yan D, Wu R, Wang C, Zhou X, Jiang Y. Quantitative description and simulation of human behavior in residential buildings. Build Simul Oct. 2012;5 (2):85–94.
[72] Tureczek AM, Nielsen PS, Madsen H, Brun A. Clustering district heat exchange stations using smart meter consumption data. Energy Build Oct. 2019;182: 144–58.
[73] Ma Z, Yan R, Nord N. A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings. Energy Oct. 2017;134:90–102.
[74] Lu Y, Tian Z, Peng P, Niu J, Li W, Zhang H. GMM clustering for heating load patterns in-depth identification and prediction model accuracy improvement of district heating system. Energy Build Oct. 2019;190:49–60.
[75] Yuan J, et al. Identification heat user behavior for improving the accuracy of heating load prediction model based on wireless on-off control system. Energy Oct. 2020;199:117454.
[76] Ruan Y, Cao J, Feng F, Li Z. The role of occupant behavior in low carbon oriented residential community planning: A case study in Qingdao. Energy Build Oct. 2017;139:385–94.
[77] Hu S, Yan D, An J, Guo S, Qian M. Investigation and analysis of Chinese residential building occupancy with large-scale questionnaire surveys. Energy Build Oct. 2019;193:289–304.
[78] Jiang S, Yang Y, Gupta S, Veneziano D, Athavale S, González MC. The TimeGeo modeling framework for urban mobility without travel surveys. Proc Natl Acad Sci 2016;113(37):E5370–8.
[79] Jiefan G, Peng X, Zhihong P, Yongbao C, Ying J, Zhe C. Extracting typical occupancy data of different buildings from mobile positioning data. Energy Build Oct. 2018;180:135–45.
[80] Happle G, Fonseca JA, Schlueter A. Context-specific urban occupancy modeling using location-based services data. Build Environ Oct. 2020;175:106803.
[81] Parker J, Hardy A, Glew D, Gorse C. A methodology for creating building energy model occupancy schedules using personal location metadata. Energy Build Oct. 2017;150:211–23.
[82] Yoshida T, Yamagata Y, Murakami D. Energy demand estimation using quasi-real-time people activity data. Energy Procedia 2019;158:4172–7.
[83] Mohammadi N, Taylor JE. Urban energy flux: Spatiotemporal fluctuations of building energy consumption and human mobility-driven prediction. Appl Energy Oct. 2017;195:810–8.
[84] Lu X, Feng F, Pang Z, Yang T, O'Neill Z. Extracting typical occupancy schedules from social media (TOSSM) and its integration with building energy modelling; 2020.
[85] Miller C, Meggers F. Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings. Energy Build Oct. 2017;156:360–73.
[86] Yan L, Liu M, Xue K, Zhang Z. A study on temperature-setting behavior for room air conditioners based on big data. J Build Eng Oct. 2020;30:101197.
[87] Razavi R, Gharipour A, Fleury M, Akpan IJ. Occupancy detection of residential buildings using smart meter data: A large-scale study. Energy Build Oct. 2019; 183:195–208.
[88] Westermann P, Deb C, Schlueter A, Evins R. Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data. Appl Energy Oct. 2020;264:114715.
[89] Wang C, et al. Dynamic occupant density models of commercial buildings for urban energy simulation. Build Environ Oct. 2020;169:106549.
[90] Hazyuk I, Ghiaus C, Penhouet D. Optimal temperature control of intermittently heated buildings using Model Predictive Control: Part I - Building modeling. Build Environ Oct. 2012;51:379–87.
[91] Baran ME, Wu FF. Optimal capacitor placement on radial distribution systems. IEEE Trans Power Deliv 1989;4(1):725–34.
[92] Fontenot HTA, Ayyagari KS, Dong B, Gatsis N. Buildings-to-distribution network integration for coordinated voltage regulation and building energy management via distribute resource flexibility. Under Rev. 2020.
[93] Fontenot HC, Dong B, Aradillaz K, Pineda G, Li Z, Jiang T. Nationwide Evaluation of Potential Energy Savings and Payback of Integrated Building and Battery Energy Storage System through Model Predictive Controls. Build Simul 2019; 2020:1659–66.
[94] Wang W, Chen J, Hong T, Zhu N. Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology. Build Environ Oct. 2018;138:160–70.
[95] Mirakhorli A, Dong B. Market and behavior driven predictive energy management for residential buildings. Sustain Cities Soc Oct. 2018;38:723–35.
[96] Dong B, Li Z, Taha A, Gatsis N. Occupancy-based buildings-to-grid integration framework for smart and connected communities. Appl Energy Oct. 2018;219: 123–37.
[97] Chen Y, Hong T, Luo X. An agent-based stochastic Occupancy Simulator. Build Simul Oct. 2018;11(1):37–49.
[98] Azar E, Al Ansari H. Multilayer agent-based modeling and social network framework to evaluate energy feedback methods for groups of buildings. J Comput Civ Eng 2017;31(4):4017007.
[99] Kaffash S, Nguyen AT, Zhu J. Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis. Int J Prod Econ 2020; 231:107868.
[100] Cui Y, Meng C, He Q, Gao J. Forecasting current and next trip purpose with social media data and Google Places. Transp Res Part C Emerg Technol 2018;97:159–74.
[101] Neilson A, Daniel B, Tjandra S. Systematic review of the literature on big data in the transportation domain: Concepts and applications. Big Data Res 2019;17: 35–44.
[102] Gundlegaard D, Rydergren C, Breyer N, Rajna B. Travel demand estimation and network assignment based on cellular network data. Comput Commun 2016;95: 29–42.
[103] Harrison G, Grant-Muller SM, Hodgson FC. New and emerging data forms in transportation planning and policy: Opportunities and challenges for 'Track and Trace' data. Transp Res Part C Emerg Technol 2020;117:102672.
[104] Fekih M, Bellemans T, Smoreda Z, Bonnel P, Furno A, Galland S. A data-driven approach for origin–destination matrix construction from cellular network

signalling data: a case study of Lyon region (France). Transportation (Amst). p. 1–32, 2020.
[105] Song C, Koren T, Wang P, Barabási A-L. Modelling the scaling properties of human mobility. Nat Phys 2010;6(10):818–23.
[106] Kung KS, Greco K, Sobolevsky S, Ratti C. Exploring universal patterns in human home-work commuting from mobile phone data. PLoS ONE 2014;9(6):e96180.
[107] Pappalardo L, Simini F, Rinzivillo S, Pedreschi D, Giannotti F, Barabási A-L. Returners and explorers dichotomy in human mobility. Nat Commun 2015;6(1):1–8.
[108] Milne D, Watling D. Big data and understanding change in the context of planning transport systems. J Transp Geogr 2019;76:235–44.
[109] Li J, Xiang T, He L. Modeling epidemic spread in transportation networks: A review. J Traffic Transp Eng (English Ed, Jan) 2021.
[110] Otero M, Barmak DH, Dorso CO, Solari HG, Natiello MA. Modeling dengue outbreaks. Math Biosci Oct. 2011;232(2):87–95.
[111] Barmak DH, Dorso CO, Otero M, Solari HG. Dengue epidemics and human mobility. Phys Rev E Oct. 2011;84(1):11901.
[112] Massaro E, Kondor D, Ratti C. Assessing the interplay between human mobility and mosquito borne diseases in urban environments. Sci Rep Oct. 2019;9(1):16911.
[113] Brockmann D, Hufnagel L, Geisel T. The scaling laws of human travel. Nature 2006;439(7075):462–5.
[114] Dong W, Lepri B, Pentland A. (Sandy). Modeling the co-evolution of behaviors and social relationships using mobile phone data. 2011, p. 134–43.
[115] Chernozhukov V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. J Econom Jan. 2021;220(1):23–62.
[116] Chang SY, et al. Mobility network modeling explains higher SARS-CoV-2 infection rates among disadvantaged groups and informs reopening strategies. Epidemiology 2020.
[117] Ramchandani A, Fan C, Mostafavi A. DeepCOVIDNet: An Interpretable Deep Learning Model for Predictive Surveillance of COVID-19 Using Heterogeneous Features and Their Interactions. IEEE Access 2020;8:159915–30.
[118] Panagopoulos G, Nikolentzos G, Vazirgiannis M. Transfer Graph Neural Networks for Pandemic Forecasting Computer Graphics View project Web mining View project George Panagopoulos École Polytechnique Transfer Graph Neural Networks for Pandemic Forecasting; 2020.
[119] Hu S, Xiong C, Yang M, Younes H, Luo W, Zhang L. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. Transp Res Part C Emerg Technol Mar. 2021;124:102955.
[120] Wang Y, Taylor JE. Coupling sentiment and human mobility in natural disasters: a Twitter-based study of the 2014 South Napa Earthquake. Nat Hazards Oct. 2018;92(2):907–25.
[121] Song X, Zhang Q, Sekimoto Y, Horanont T, Ueyama S, Shibasaki R. Modeling and probabilistic reasoning of population evacuation during large-scale disaster. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13, 2013, vol. Part F128815, p. 1231.
[122] Song X, Zhang Q, Sekimoto Y, Shibasaki R. Prediction of human emergency behavior and their mobility following large-scale disaster. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14, 2014, pp. 5–14.
[123] Solmaz G, Turgut D. Modeling pedestrian mobility in disaster areas. Pervasive Mob Comput Sep. 2017;40:104–22.
[124] Eagle N, Pentland AS. Reality mining: sensing complex social systems. Pers ubiquitous Comput 2006;10(4):255–68.
[125] Pietiläinen A-K, Oliver E, LeBrun J, Varghese G, Diot C. MobiClique: middleware for mobile social networking. 2009;49–54.
[126] Shao W, Salim FD, Song A, Bouguettaya A. Clustering big spatiotemporal-interval data. IEEE Trans Big Data 2016;2(3):190–203.
[127] Rumi SK, Deng K, Salim FD. Crime event prediction with dynamic features. EPJ Data Sci 2018;7(1):43.
[128] ASHRAEAS. Standard 90.1-2004, Energy standard for buildings except low rise residential buildings. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 2004.
[129] Abushakra B, Sreshthaputra A, Haberl JS, Claridge DE. Compilation of Diversity Factors and Schedules for Energy and Cooling Load Calculations. ASHRAE Research Project 1093-RP, Final Report; Oct. 2001.
[130] DOE, "EnergyPlus". United States Department of. Energy 1997.
[131] Li Z, Dong B. A new modeling approach for short-term prediction of occupancy in residential buildings. Build Environ Oct. 2017;121:277–90.
[132] Baetens R, Saelens D. Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour. J Build Perform Simul Oct. 2016;9(4):431–47.
[133] Mirakhorli A, Dong B. Occupant-behavior driven appliance scheduling for residential buildings. Build Simul Oct. 2017;10(6):917–31.
[134] Algizawy E. HMM-based-traffic-using-CDRs. Github, 2019.
[135] Diwakar PS. Vehicular Traffic Modelling using HMM. Github 2018.
[136] Bancel J-R. Traffic HMM. Github 2013.
[137] Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci 2009;106(51):21484–9.
[138] Yan X-Y, Wang W-X, Gao Z-Y, Lai Y-C. Universal model of individual and population mobility on diverse spatial scales. Nat Commun 2017;8(1):1–9.
[139] Xie Y, Lord D, Zhang Y. Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. Accid Anal Prev Oct. 2007;39(5):922–33.
[140] Hasz B. tfp taxi. Github 2019.
[141] Feng J, et al. Deepmove: Predicting human mobility with attentional recurrent networks. 2018;1459–68.
[142] Liu Q, Wu S, Wang L, Tan T. Predicting the next location: A recurrent model with spatial and temporal contexts. 2016.
[143] Yang D, Fankhauser B, Rosso P, Cudre-Mauroux P. "Location Prediction over Sparse User Mobility Traces Using RNNs. Flashback in Hidden States!" 2020:2184–90.
[144] Jiang R, et al. Deepurbanevent: A system for predicting citywide crowd dynamics at big events. 2019:2114–22.
[145] Sun K, Qian T, Chen T, Liang Y, Nguyen QVH, Yin H. Where to Go Next: Modeling Long- and Short-Term User Preferences for Point-of-Interest Recommendation. Proc AAAI Conf Artif Intell Oct. 2020;34(01):214–21.
[146] You J, Ying R, Leskovec J. Position-aware graph neural networks. arXiv Prepr. arXiv1906.04817, 2019.
[147] Fan W, et al. Graph Neural Networks for Social Recommendation. The World Wide Web Conference 2019:417–26.
[148] Mnih V, et al. Human-level control through deep reinforcement learning. Nature. 2015:518(7540);529–533.
[149] James JQ, Yu W, Gu J. Online vehicle routing with neural combinatorial optimization and deep reinforcement learning. IEEE Trans Intell Transp Syst 2019;20(10):3806–17.
[150] Liu X-Y, Ding Z, Borst S, Walid A. Deep reinforcement learning for intelligent transportation systems. arXiv Prepr. arXiv1812.00979, 2018.
[151] Balaji PG, German X, Srinivasan D. Urban traffic signal control using reinforcement learning agents. IET Intell Transp Syst 2010;4(3):177–88.
[152] Qian T, Shao C, Wang X, Shahidehpour M. Deep Reinforcement Learning for EV Charging Navigation by Coordinating Smart Grid and Intelligent Transportation System. IEEE Trans Smart Grid 2019;11(2):1714–23.
[153] Ohsugi S, Koshizuka N. Delivery route optimization through occupancy prediction from electricity usage. 2018:1;842–49.
[154] Ghosh SK, Ghosh S. Modeling Individual's Movement Patterns to Infer Next Location from Sparse Trajectory Traces. In: Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018, 2019, p. 693–98.
[155] Zhou X, Liu T, Yan D, Shi X, Jin X. An action-based Markov chain modeling approach for predicting the window operating behavior in office spaces. Build Simul Apr. 2020;14(2):301–15.
[156] Qiao Y, Si Z, Zhang Y, Ben Abdesslem F, Zhang X, Yang J. A hybrid Markov-based model for human mobility prediction. Neurocomputing 2017;278:99–109.
[157] Toch E, Lerner B, Ben-Zion E, Ben-Gal I. Analyzing large-scale human mobility data: a survey of machine learning methods and applications. Knowl Inf Syst 2019;58(3):501–23.
[158] Montroll EW, Weiss GH. Random Walks on Lattices. II. J Math Phys Oct. 1965;6(2):167–81.
[159] Gonzalez MC, Hidalgo CA, Barabasi A-L. Understanding individual human mobility patterns. Nature Oct. 2008;453(7196):779–83.
[160] Akoush S, Sameh A. Movement Prediction Using Bayesian Learning for Neural Networks. In: 2007 Second International Conference on Systems and Networks Communications (ICSNC 2007), 2007, p. 6.
[161] Fan C, Yan D, Xiao F, Li A, An J, Kang X. Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches. Build Simul. 01-Feb-2021;14(1). Tsinghua University, p. 3–24.
[162] Blundell C, Cornebise J, Kavukcuoglu K, Wierstra D. Weight Uncertainty in Neural Networks. arXiv1505.05424 [cs, stat], Oct. 2015.
[163] Bian Z, Ozbay K. Estimating Uncertainty of Work Zone Capacity using Neural Network Models. Transp Res Rec J Transp Res Board Feb. 2019;2673(2):49–59.
[164] Van Hinsbergen CPIJ, Hegyi A, Van Lint JWC, Van Zuylen HJ. Bayesian neural networks for the prediction of stochastic travel times in urban networks. IET Intel Transport Syst 2011;5(4):259–65.
[165] Park H, Haghani A. Real-time prediction of secondary incident occurrences using vehicle probe data. Transp Res Part C Emerg Technol Sep. 2016;70:69–85.
[166] Park H, Haghani A, Zhang X. Interpretation of Bayesian neural networks for predicting the duration of detected incidents. J Intell Transp Syst Jul. 2016;20(4):385–400.
[167] Eleftheroglou N, et al. Intelligent data-driven prognostic methodologies for the real-time remaining useful life until the end-of-discharge estimation of the Lithium-Polymer batteries of unmanned aerial vehicles with uncertainty quantification. Appl Energy Nov. 2019;254:113677.
[168] Kullback S, Leibler RA. On Information and Sufficiency. Ann Math Stat Oct. 1951;22(1):79–86.
[169] Hinton GE, van Camp D. Keeping the neural networks simple by minimizing the description length of the weights. 1993, p. 5–13.
[170] Aggarwal CC. An Introduction to Neural Networks; p. 535.
[171] Lampinen J, Vehtari A. Bayesian approach for neural networks—review and case studies. Neural Networks Oct. 2001;14(3):257–74.
[172] PyMC, Inference — PyMC3 3.9.3 documentation. The PyMC Development Team, 2020.
[173] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B (Statistical Methodol, Oct) 2005;67(2):301–20.
[174] Hochreiter S, Schmidhuber J, Short-Term Long. Memory 1997.
[175] Chen Y, Long C, Cong G, Li C. Context-aware Deep Model for Joint Mobility and Time Prediction. In: WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining; 2020. p. 106–14.
[176] Su K, Li J, Fu H. Smart city and the applications. In: in *2011 International Conference on Electronics, Communications and Control (ICECC)*; 2011. p. 1028–31.

[177] Schultz L, Sokolov V. Deep Reinforcement Learning for Dynamic Urban Transportation Problems. arXiv1806.05310 [cs, stat], Oct. 2018.
[178] Sutton RS, Barto AG. Reinforcement Learning, second edition: An Introduction. MIT Press; 2018.
[179] Mnih V, et al. Playing Atari with Deep Reinforcement Learning. arXiv1312.5602 [cs], Oct. 2013.
[180] Haydari A, Yilmaz A. Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey. arXiv2005.00935 [cs, eess, stat], Oct. 2020.
[181] El-Sayed AM, Scarborough P, Seemann L, Galea S. Social network analysis and agent-based modeling in social epidemiology. Epidemiol Perspect Innov Oct. 2012;9:1.
[182] Chen Z, Li X, Bruna J. Supervised Community Detection with Line Graph Neural Networks. arXiv:1705.08415 [stat], Oct. 2020.
[183] Cho E, Myers SA, Leskovec J. Friendship and mobility: user movement in location-based social networks. 2011;1082–90.
[184] Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The Graph Neural Network Model. IEEE Trans Neural Networks 2009;20(1):61–80.
[185] Azar E, Nikolopoulou C, Papadopoulos S. Integrating and optimizing metrics of sustainable building performance using human-focused agent-based modeling. Appl Energy Oct. 2016;183:926–37.
[186] Heaton J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 p. ISBN: 0262035618. Genet. Program. Evolvable Mach., vol. 19, 2017.
[187] Bishop C. Pattern Recognition and Machine Learning. New York: Springer-Verlag; 2006.
[188] Friedman J, Hastie T, Tibshirani R. The Elements Of Statistical Learning. Aug, Springer Ser. Stat., vol. 1, no. 10, 2001.
[189] Chong A, Augenbroe G, Yan D. Occupancy data at different spatial resolutions: Building energy performance and model calibration. Appl Energy Mar. 2021;286:116492.
[190] Karimzadeh M, Zhao Z, Gerber F, Braun T. Mobile Users Location Prediction with Complex Behavior Understanding. In: Proceedings - Conference on Local Computer Networks, LCN, 2019, vol. 2018-October, p. 323–6.
[191] Klous G, Kretzschmar MEE, Coutinho RA, Heederik DJJ, Huss A. Prediction of human active mobility in rural areas: development and validity tests of three different approaches. J Expo Sci Environ Epidemiol Nov. 2020;30(6):1023–31.
[192] Miao C, Luo Z, Zeng F, Wang J. Predicting Human Mobility via Attentive Convolutional Network Human Mobility Prediction; Convolutional Neural Network; At-tention Mechanism ACM Reference Format.
[193] Xu J, Xing C, Zhang Y. Predicting Next Points of Interests Based on a Markov Model. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, 2018, vol. 252, p. 264–74.
[194] Yan C, Wang P, Pang H, Sun L, Yang S. CELoF: WiFi dwell time estimation in free environment. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, vol. 10132 LNCS, p. 503–14.
[195] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc Mar. 2007;102(477):359–78.
[196] Buffa S, Cozzini M, D'Antoni M, Baratieri M, Fedrizzi R. 5th generation district heating and cooling systems: A review of existing cases in Europe. Renew Sustain Energy Rev Oct. 2019;104:504–22.
[197] Guo X, et al. Survey of COVID-19 Disease Among Orthopaedic Surgeons in Wuhan, People's Republic of China. J Bone Joint Surg Am 2020.
[198] Holling CS. Resilience and Stability of Ecological Systems. Annu Rev Ecol Syst Nov. 1973;4(1):1–23.
[199] Moazami A, Carlucci S, Geving S. Robust and resilient buildings: A framework for defining the protection against climate uncertainty. IOP Conf Ser Mater Sci Eng Oct. 2019;609:72068.
[200] Bibri SE. On the sustainability of smart and smarter cities in the era of big data: an interdisciplinary and transdisciplinary literature review. J Big Data, vol. 6, no. 1. SpringerOpen, p. 1–64, 01-Dec-2019.
[201] Ding F, Ge Q, Jiang D, Fu J, Hao M. Understanding the dynamics of terrorism events with multiple-discipline datasets and machine learning approach. PLoS ONE Oct. 2017;12(6):e0179057.
[202] Vuorinen V, et al. Modelling aerosol transport and virus exposure with numerical simulations in relation to SARS-CoV-2 transmission by inhalation indoors. Saf Sci Oct. 2020;130:104866.
[203] Zhang N, Huang H, Su B, Ma X, Li Y. A human behavior integrated hierarchical model of airborne disease transmission in a large city. Build Environ Oct. 2018;127:211–20.
[204] Megahed NA, Ghoneim EM. Antivirus-built environment: Lessons learned from Covid-19 pandemic. Sustain Cities Soc Oct. 2020;61:102350.
[205] Salim F, Haque U. 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 Oct. 2015;81:31–48.
[206] Zambonelli F, Salim F, Loke SW, De Meuter W, Kanhere S. Algorithmic Governance in Smart Cities: The Conundrum and the Potential of Pervasive Computing Solutions. IEEE Technol Soc Mag 2018;37(2):80–7.