Building power consumption datasets: Survey, taxonomy and future directions

Yassine Himeur [a,*], Abdullah Alsalemi [a], Faycal Bensaali [a], Abbes Amira [b]

[a] Department of Electrical Engineering, Qatar University, Doha, Qatar
[b] Institute of Artificial Intelligence, De Montfort University, Leicester, UK

A B S T R A C T

In the last decade, extended efforts have been poured into energy efficiency. Several energy consumption datasets were henceforth published, with each dataset varying in properties, uses and limitations. For instance, building energy consumption patterns are sourced from several sources, including ambient conditions, user occupancy, weather conditions and consumer preferences. Thus, a proper understanding of the available datasets will result in a strong basis for improving energy efficiency. Starting from the necessity of a comprehensive review of existing databases, this work is proposed to survey, study and visualize the numerical and methodological nature of building energy consumption datasets. A total of thirty-one databases are examined and compared in terms of several features, such as the geographical location, period of collection, number of monitored households, sampling rate of collected data, number of sub-metered appliances, extracted features and release date. Furthermore, data collection platforms and related modules for data transmission, data storage and privacy concerns used in different datasets are also analyzed and compared. Based on the analytical study, a novel dataset has been presented, namely Qatar university dataset, which is an annotated power consumption anomaly detection dataset. The latter will be very useful for testing and training anomaly detection algorithms, and hence reducing wasted energy. Moving forward, a set of recommendations is derived to improve datasets collection, such as the adoption of multi-modal data collection, smart Internet of things data collection, low-cost hardware platforms and privacy and security mechanisms. In addition, future directions to improve datasets exploitation and utilization are identified, including the use of novel machine learning solutions, innovative visualization tools and explainable mobile recommender systems. Accordingly, a novel visualization strategy based on using power consumption micro-moments has been presented along with an example of deploying machine learning algorithms to classify the micro-moment classes and identify anomalous power usage.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Contents

---

* Corresponding author.
  E-mail addresses: yassine.himeur@qu.edu.qa (Y. Himeur), a.alsalemi@qu.edu.qa (A. Alsalemi), f.bensaali@qu.edu.qa (F. Bensaali), abbes.amira@dmu.ac.uk (A. Amira).

1. Introduction

Recent studies have shown that buildings are in charge of more than 40% of power consumption demand and greenhouse gas emissions around the world. Indeed, this steady growth in energy consumption has been closely tied to the increasing number of population and the rising levels of prosperity [1,2]. Moreover, even climate conditions in certain regions of the world obliged households to call for more energy for heating, cooling, cooking and refrigeration needs. Business buildings, principally offices and university structures, are also viewed as structures exhibiting high power consumption [3,4]. Consequently, the expansion of power usage and carbon emission as well as expensive energy prices in the above mentioned environments has made energy preserving a vital goal for various public authorities of all governments in order to accomplish efficient energy reduction [5–9].

The building energy sector has recently attract the attention of various public energy efficiency initiatives to accomplish greater energy sustainability. Additionally, a clear correlation is found between household energy consumption and user behavior [10–14]. Every day, new appliances are installed and used in households resulting in an incredible rise of power demand. Monitoring the power usage of these appliances is dependably the initial step towards energy preserving. From this point of view, understanding and controlling user consumption behavior is also a key parameter to help householders reduce energy costs. According to the ongoing evolution of the Internet of Things (IoT), the use of smart meters for monitoring electricity consumption is expanding exponentially [15–17]. The up-and-coming generation of energy saving systems should be more effective, easy to follow and more challenging in order to improve end-user behavior.

As of now, as well as other research areas, energy, environment and sustainable development research topics are encountering the urgency of developing openly accessible datasets. In fact, power consumption datasets progressively come to be more consistent when estimating the precision of power monitoring techniques and perceiving how good they may behave under realistic circumstances. Consequently, checking the precision of outputs in real scenarios is critical in this research area [18–20]. Moreover, simulated database does not reasonably fit realistic datasets as "an experimental database or repository would ordinarily have unpredictable and unexplained complication nature that is laboriously anticipated and most of the time can be laboriously hard to manage [21,22]". On that account, Energy scientists have proved that it is important to have open access databases that provide aggregated power consumption as well as appliance based consumption for the various devices that constitute the overall consumption [23].

Moreover, end-users behavior is responsible of wasting more than 20% of the total energy consumed in buildings, and hence it is a key element in energy consumption [24–26]. Therefore, it is of paramount importance to: (i) design real consumption datasets; (ii) deploy novel platforms and smart-meters to collect granular and appliance-specific data, which have the means or incentive for sharing consumption data with end-users; and (iii) develop tools that help end-users in understanding their energy consumption footprints, such as innovative visualizations, and further implement novel strategies that help them in improving their behavior and reduce wasted energy [27], e.g. via deploying recommender systems. In this context, seeking to study how to save energy and understand power consumption behaviors in buildings, various datasets have been collected globally. They provide a big amount of information and create an immense quantity of readings about daily power usage and user behavior. Therefore, the use of machine learning (ML) algorithms becomes essential for handling large-scale datasets and extracting meaningful features from collected data. This can assist in many applications including forecasting power demand, energy efficiency and electricity preserving, appliance recognition, cost prediction, and other things in relation to energy usage [28–30]. In this respect, any ML algorithm for building power consumption deals with information drawn from smart meters and solar panels during the different periods of the day. This huge quantity of data including multivariate time series is of utmost importance to ML algorithms because future usage can be effectively anticipated [31–33]. In other words, ML algorithms can forecast such data and help in developing energy efficiency frameworks proficiently.

Extremely inspired by the rising relevance of public datasets, we opted to give a general review of up to 31 existing datasets in the field of building power consumption. Presently, householders, firms and public authorities are confronting difficulties to guarantee the energy efficiency and reduce usage costs. The use of a large number of appliances increases the energy demand, cost and carbon emissions as well. In this context, we present in this paper a deep overview of various power consumption datasets, their taxonomy and classification. The taxonomy is adopted to examine existing datasets, what kind of information they provide, their applications, their benefits and their limitations. To the best of our knowledge, this is the first framework that provides a comprehensive and universal survey of building energy consumption datasets, their applications and future trends. Following, discussions and important findings are presented via analyzing and comparing the features of existing datasets, their data collection platforms and related modules. Moreover, a novel dataset, namely Qatar university dataset (QUD) is proposed to answer the challenges and issues raised from the analytical study. Afterward, valuable future orientations to improve the quality of power consumption datasets and enrich their content are discussed, in which a set of recommendations to use novel hardware devices and platforms are described. Moreover, future direction to improve
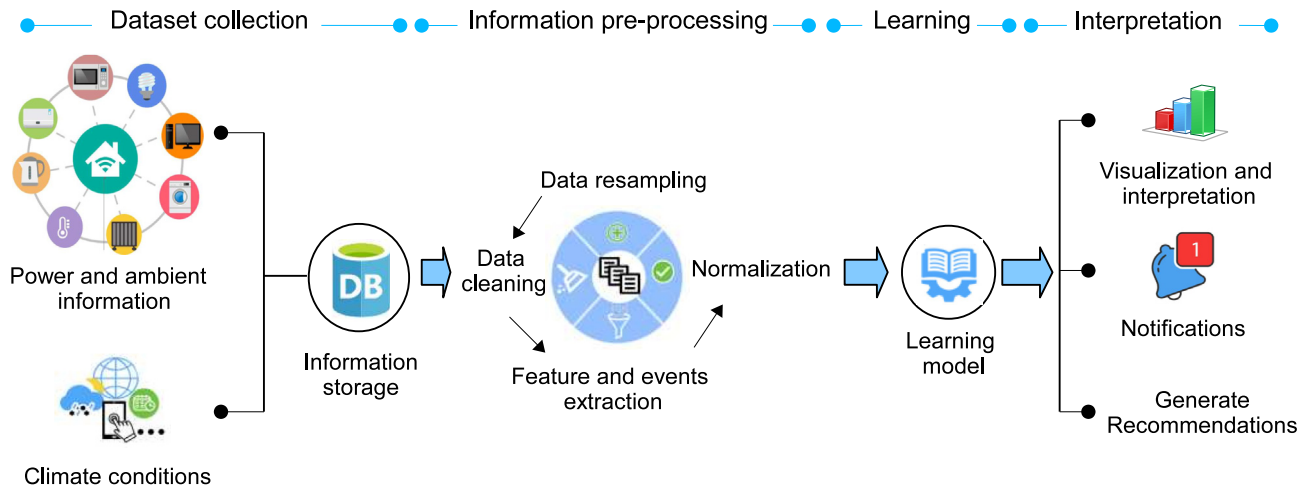
Fig. 1. Standard representation of power consumption dataset collection system with its associated modules.

datasets exploitation and hence improve energy saving are also identified. To summarize, this paper presents a set of novel contributions, which can be listed as follows:

- Reviewing up to 31 building power consumption datasets, describing their properties and highlighting their pros and cons via adopting a multi-perspective comparison based on various parameters.
- Proposing a taxonomy of building power consumption datasets to assess the existing repositories based on their applications an characteristics.
- Analyzing data collection platforms used to record power consumption datasets and related modules used for data transmission, data storage and privacy concerns.
- Presenting a novel dataset called QUD that responds to various issues raised in the analysis of state-of-the-art datasets. QUD can be used for different applications, among them detecting of anomalous power consumption.
- Providing a list of valuable future orientations for (i) improving datasets collection mainly through the use of novel hardware platforms, and (ii) improving datasets exploitation via adopting innovative tools such as visualization strategies and explainable recommender systems.

The rest of this paper has been organized into four sections. Section 2 reviews up to 31 existing building power consumption datasets and describes their usage contexts, properties, advantages and limitations. Section 3 presents a comprehensive discussion about the different characteristics of existing power consumption datasets. In addition, a novel dataset called QUD is presented which presents new functionalities. In Section 4, challenging orientations and future directions that should be followed in order to improve datasets collection and enhance datasets exploitation are described. Section 5 concludes the paper with a set of proposals for improving the quality of power consumption datasets and highlights future works. Finally, a list of abbreviations and nomenclatures used in this paper is presented in the Appendix.

2. Overview of building power consumption datasets

Several datasets can be found in literature and each one has its specific characteristics, making it difficult to select a database for treating energy efficiency issues. To this end, this work makes a deep comparison between all datasets based on various specifications, such as the period and region of collection, sampling rate,

number of monitored houses, number of deployed sub-meters, collected features and release date. As a matter of fact, existing realistic datasets are divided into two major groups; appliance-level datasets versus aggregated-level based databases. The first group class provides sub-meter readings of appliance-by-appliance consumption. This kind of data is used for various applications, including energy saving [34], appliance recognition [35], occupancy detection [36,37] and preference behavior [38,39]. The second class group focuses on collecting overall consumption profiles of different buildings. It can be employed for energy disaggregation, energy efficiency, and further predicting energy consumption. Fig. 1 illustrates a flowchart of a dataset collection process along with its associated modules, required to pre-process, analyze and interpret power consumption patterns. This is a general representation that can be used for different applications.

2.1. State of the art of existing datasets

To fit realistic scenarios of daily power usage and test energy efficiency solutions, scientists and specialists of smart energy monitoring systems need power consumption databases, in which developed algorithms can be evaluated in advance. Different databases have been collected and shared publicly. Under this section we review up to 31 power consumption datasets that are proposed in literature in addition to our novel dataset named QUD. We specify briefly the characteristics of each dataset and registered features in terms of current (I), voltage (V), active power (P), reactive power (Q), apparent power (S), normalized power (Np), energy (E), frequency (f), phase angle (ϕ), power factor (pf), energy cost (EC), weather (Wt), Temperature (T), humidity (H), occupancy (O) and light level (L).

In [40–45], large-scale datasets are formed, namely, HES, IHEPCDS, UMSM, SustData, REFIT and Dataport, respectively. While HES, IHEPCDS, UMSM and Dataport assembled energy consumptions patterns at a minutely level, SustSata and REFIT reported power usage profiles over intervals in seconds. All these databases provide consumption records at the appliance-level for long periods of monitoring. For example; in HES and UMSM data are raised for a period of one year, in REFIT and SustData energy patterns are accumulated for 213 days and 1114 days, respectively. Further, different features are gathered during the experimental campaign, such as I, V, P, Q, S f and T. REFIT has also the particularity of providing EC in $. Dataport repository [45] is also quite similar to UMSM database, since it captures energy usage at the same sampling intervals of 239 households but for a short collection period

of two months. Dataport repository is also quite similar to UMSM database, since it captures energy usage at the same sampling intervals of more than 1200 households for a long collection period, which is more than 4 years.

In [46], OCTES is proposed, which is similar to REFIT. It records P, ϕ and EC ($\$$). In addition, data are collected in a shorter investigation period. A bigger examination size and information are recorded at a comparable rate to REFIT. It lists the power consumption of each house; nonetheless, other pieces of information about the houses are not provided except their geological position. The case study specied in this work depicts the utilization of a sauna in one home; as though, this data isn't shared publicly. Consequently, a presumption should be put with regards to the energy usage. In addition to power consumption, REFIT provides also readings about temperature, light, and motion patterns expanded with dwelling reviews specifying; size, age, warming sort, isolation,fabrication type and details about the tenants or occupants, job description and age.

In [47], Tracebase database includes power consumption patterns of various devices, which enables to examine disaggregation. The readings are collected at a sampling rate of 1 s. This dataset can be utilized for energy efficiency applications. However, it can not be employed for appliance recognition, preference detection or energy disaggregation since no data are provided about the devices being investigated and their properties. It gathers data of 43 distinct appliances, in which every one has various recordings from several days and several households. Furthermore, date and time records, P and Np are provided at a sampling frequency of 8 s.

In [48–51], AMPds1, AMPds2, ECB and PSD are proposed, respectively, which are minutely power datasets. Overall, AMPds1 and AMPds2 repositories are deemed as largely used databases, which compiled information of one and two years, accordingly, with a sampling rate of 1 min. In fact, energy consumption of 11 appliances is observed using 21 sub-meters. On the other side, ECB that provides electricity consumption benchmarks of 25 domestic residents located in Victoria State in south-eastern Australia is released. Consumption patterns were extracted from the aggregated circuit and for individual appliances over a duration of two years and at a sampling rate of 30 min. Further, consumption footprints of device-event labels from 10 homes in Austin, USA, were assembled.

In [52,53] authors released MEULPv.1 and MEULPv.2 datasets, respectively. MEULPv.1 gives energy consumption readings of 12 Canadian households. Data were recorded at 1-min sampling rates at both the aggregated and appliance levels. A total of 8 appliances are monitored during the data collection process. Meanwhile, MEULPv.2 provides one year monitoring of 23 households using a sampling rate of 1 min that designates aggregated and appliance-based consumptions as well.

In [21,54], RAE and GREEND databases are proposed, in which data are collected at a frequency of 1 Hz. The RAE is the initial version of an energy consumption repository that includes 1 Hz recordings for aggregated and sub-metered levels of two households. Besides power information, T and H records from a house's indoor regulator are incorporated. On the other side, GREEND is proposed to describe detailed energy consumption patterns collected through an experimental campaign via assessing electricity usage of various individual appliances in Austria and Italy. During the collection campaign, eight households are monitored, where each one contains up to nine different individual devices. The power usage patterns at a device-level are gleaned at a resolution of 1 Hz through a period of six months.

In [55–57], ECO, IWAE and DRED that capture energy information at 1 Hz sampling intervals are nominated, accordingly. ECO is an entire measurement campaign managed in order to collect comprehensive information of consumption patterns in six Swiss homes through an eight months duration. During the collection campaign, I, V, and π are collected from aggregated circuits and a set selected appliances at a sampling frequency of 1 Hz. Through the IAWE campaign, measurements were performed in a pilot household with three floors in Delhi in order to measure power, water and environmental profiles. Data are collected for a duration of 73 days from May to August 2013. In addition, 33 sub-meters are deployed through the whole house. DRED is publicly launched to capture energy, occupancy patterns and environmental data of one pilot house in the Netherlands. Sensor units are installed to measure aggregated energy consumption and appliance level electricity usage. In fact, 12 different domestic appliances are sub-metered at sampling intervals of 1 min while 1 Hz sampling rates are used to gather aggregated consumption.

In [58], DISEC is launched, in which various data are collected for 19 apartments at an Indian faculty housing complex during 284 days. Different features, such as P and Wt, are collected in a 30 s sampling intervals and then aggregated to 15 min, 30 min and 60 min intervals. As well, Wt variations are updated through measuring atmospheric conditions from nearly station measurements.

In [59,60], two hourly electricity consumption datasets are proposed. The first one called CRHLP includes energy patterns of 16 residential and commercial buildings monitored at every hour for a period of one year. Additionally, solar radiation and meteorological records are also collected. The second one, namely HUE, captures long-term energy usage profiles from five households with a sampling frequency of 1 h. Furthermore, while device-level consumptions from house 1 are collected for a period of two years with sampling intervals of one minute, data from house 2 are extracted for a one year period with a resolution rate of 1 Hz. In [61], UK-DALE is proposed, which summarizes the current and voltage profiles of three houses at sampling intervals of 16 kHz and two houses at sampling frequencies of 1 Hz. Moreover, patterns of individual devices of five other households are collected at a sampling rate of 6 s for various periods varying from 39 to 655 days.

In [62–64], REDD, BLUED and BLOND datasets are proposed. Energy consumption records are captured at a sampling frequency of more than 10 kHz. The monitoring process, by contrast, is conducted for only a few weeks. For example, in REDD, six households are monitored, where the aggregated electricity consumption is measured at a sampling rate (15 kHz). Also, electricity consumption reviews of up to 24 devices are monitored at sampling intervals of 0.5 Hz. Furthermore, load patterns of other 20 appliances are observed at a frequency of 1 Hz while BLUED resumes the current and voltage readings of an individual household in Pittsburgh, Pennsylvania, USA. Data are listed at a sampling frequency of 12 kHz over a period of one week. For BLOND, it aims to capture continuous power consumption data. It delivers voltage and current records at the aggregated and device levels. This database includes data from 53 devices that represent 16 appliance groups. It englobes two main repositories; (i) BLOND-50 that in turn has consumption data obtained at sampling intervals of 50 kSps for grouped circuits and 64 kSps for individual devices; and (ii) BLOND-250 that entails usage patterns for a period of 50 days gathered using sampling rates of 250 kSps at the aggregated-level and 50 kSps at the appliance-level.

In [65], PLAID expresses power consumption profiles for more than 56 specific domestic equipments that represent about 11 appliance categories. Data are captured at a sampling frequency of 30 kHz that is judged among the highest resolution frequency used in existing building power consumption datasets when collecting load profiles. In addition, energy consumption information is captured for a period of three months during the summer of
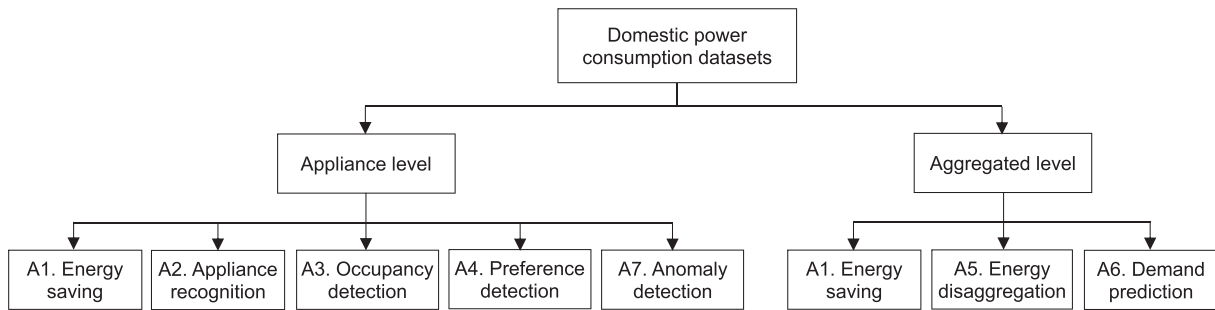
Fig. 2. General taxonomy of existing building power consumption datasets.

2013 and the measurement campaign has been carried on in Pittsburgh, Pennsylvania, USA.

ACS-F1 [66] and BERDS [67] datasets that monitor load patterns at a comparable sampling rate are proposed. ACS-F1 records the amount of energy used in a set of households at an appliance-level. In this context, electricity sub-meters were employed to measure the energy consumption of 100 house devices that represent 10 appliance classes. Power sub-metering is managed at sampling intervals of 10 s for a period of only one hour. This database is especially suitable for appliance recognition applications. On the other side, BERDS collects energy consumption outlines at 20 s sampling rates for a period of one year.

In addition, QUD is presented in this framework, which is based on an appliance-based collection campaign. It can be used for different purposes, such as the energy saving, anomaly detection and energy demand prediction. QUD is collected using a system that incorporates sub-metering modules registering power consumption footprints in terms of P and other indoor climate conditions, including O, T, H and L. The data are recorded with sampling intervals ranging from 3 s to 30 min. The collection process will be spread over a one year period, while three months of data recording have been already completed.

2.2. Taxonomy

Power consumption datasets are split into two main groups: Appliance-level versus aggregated-level. The first one traces power consumption arrangements of individual devices. The second one provides the whole power consumption of households. Datasets can also be classified based on different aspects including application purposes or the nature of buildings, where data are acquired among which households, commercial buildings, academic buildings, industrial, etc. Fig. 2 details the global taxonomy of various building power consumption datasets found in the literature.

2.3. Applications (A)

Using detailed power consumption readings and based on the nature of data collection procedures at appliance or aggregated levels, existing datasets could be exploited for various applications including, but not restricted to, energy saving, appliance recognition, occupancy detection, user preference detection, abnormal detection, energy disaggregation and energy demand prediction.

A1. Energy saving: Investigating the building sector in terms of energy saving which is a principal element of its environmental and financial effects is of utmost importance. Consequently, energy saving is the most popular application of building power consumption datasets [68–70]. It can effectively reduce energy bills and decrease carbon dioxide emissions. It is made out of the following four stages: (1) the dataset collection stage, in which information is reaped from various sources, including energy sub-meters, ambi-

ent condition sensors and climate sources. The information gathered from these heterogeneous sources is saved in a specific dataset; (2) the pre-processing step, in which the information stored in the first step is pre-processed before utilizing various ML strategies. the pre-processing includes data cleaning, data resampling, features and events extraction and normalization; (3) The learning stage, in which ML algorithms are utilized to learn functions and models; and (4) The adoption of visualizations and recommendations phase, in which visualization tools are first adopted to provide end-users with interpretation of their consumption patterns. Following, specific recommendations or directives are derived in order to promote energy efficiency behaviors. Since the energy saving application is very relevant, we focus in this paper on studying how to improve systems developed in this direction along with related applications.

A2. Appliance recognition: Appliance recognition systems can help detecting operating conditions of devices using collected power usage patterns, and thoroughly recognizing the nature of each appliance [71]. In [72], a model was designed to detect the device activity and then to associate activities with devices using collected data. Analyzing power signals and checking relations among activities can assist detecting unattended devices, which use energy power without taking part the domestic's activities. In [73], in order to fit realistic conditions, experiments are usually conducted on a set of building power consumption databases, such as ACS-F1, PLAID, BLUED and UK-DALE.

A3. Occupancy detection: Solutions presented in this area detect individuals' occupancy in each specific part of a building based on power consumption profiles, as well as other environmental specifications, such as the temperature, humidity, luminosity and carbon dioxide emissions [74]. Dataset patterns are inspected before using ML approaches to derive the occupancy of the monitored part. Generally, occupancy is detected in two stages; (i) the presence or absence of individuals is investigated; and (ii) the number of individuals in the monitored building/room is then calculated [75–77]. In [78], a set of ML models as well as their boosting forms are developed and tested to detect occupancy using collected data from the AMPds2 measurement campaign.

A4. Preference detection: Methods described in this class deal with evaluating individual preferences through analyzing energy usage profiles. Most approaches treat the thermal comfort, although there are other arrangements that address visual comfort. Works released in this area investigate information-driven methodologies from an ML point of view and yielded arrangements that determine the preferences (e.g. the habits related to appliance usage) even through getting reports from individuals, i.e. information labeling or via observing the historic behavior of end-users to construe (in a straightforward manner) their consumption priorities or contexts that satisfy their well-being [38,79].

A5. Energy disaggregation: Energy disaggregation is the issue of segregating the overall power consumption record into particu-

lar signals, in which each one represents an individual consumption of each electrical device [80–83]. This is valuable since getting separated power consumption of each appliance helps individuals to save energy and provide consumers with indexes on how to make appropriate actions [84]. Most of existing energy disaggregation frameworks resolving the problem of non-intrusive load monitoring (NILM) attempt to segregate the overall energy consumption without utilizing separate meters for each appliance [85–89]. For this specific application, REDD, BERDS, REFIT, AMPds1 and AMPds2 datasets are reputed among the famous repositories used for energy disaggregation.

A6. Demand prediction: ML algorithms generate precise power demand forecasts and they can be selected by public authorities and project managers instrumenting energy-efficiency procedures [90–94]. For domestic households, academic and industrial buildings, if the power demand could be predicted using ML strategies, directives and mechanisms that should be followed in advance can be established with a view of reducing load consumption of equipments and appliances inside these infrastructures[95–98]. Moreover, even if most the above presented databases (Section 2.1) are used for energy forecasting, we can find in the literature other datasets that are only designed for the specific problem of load and energy price forecasting, such as GEFCom2012 [99] and GEFCom2014 [100].

A7. Anomaly detection: With the progressive widespread use of smart-meters and smart sensors to monitor load usage in house-holds, the utilization of power consumption observations as a solution to detect abnormal usage of energy is absolutely fascinating. Specifically, early detection approaches can be deployed to identify a large set of failures. In addition, recent works illustrate that for example, anomalous in lighting appliances can be responsible of 2–11 % of the whole power consumption of households and commercial structures [101]. Furthermore, detecting faults or anomalies can permit analysts to comprehend energy consumption behavior of end-users and to be conscious of unpredictable energy usage values [102,103]. Various data mining approaches have been explored and deployed to detect anomalous events during energy usage process [104–109]. In addition, it is worthy to mention that there is an absence of annotated datasets dedicated to power consumption anomaly detection.

However, in order that a dataset could be correctly and efficiently used for a specific application, it should respect some specific requirements. For energy disaggregation, datasets should include both aggregated and appliance-level consumption fingerprints to compare the results obtained from disaggregation solutions with individual patterns. To conduct a user preference detection or even an occupancy detection, datasets should encompass appliance-level power consumption because it is difficult even impossible to infer user preferences from aggregated data. In addition, for occupancy detection, it is also required that consumption and ambient condition should be gleaned from individual appliances and from various parts of the building. For

Table 1
Features comparison of existing building power consumption datasets.

#	Acronym	Country	Period	#Homes	#sub-meters	Features	Sampling rate	Applications	Release
1	REDD [62]	Massachusetts, USA	119 days	6	24	I, V, P	3 s	A1,A5	2011
2	HES [40]	England, UK	1 year	26/251	23	I, V, P, T	10 min	A1,A6	2011
3	IHEPCDS [41]	Paris, France	47 months	1	3	I, V, P, Q	1 min	A1,A6	2012
4	UMSM [42]	Massachusetts, USA	1 year	400	8	I, V, P, f, S	1 min	A1	2012
5	Dataport [45]	USA	4 years	+1200	70	P	1 min	A1	/
6	MEULPv1 [52]	Canada	1 year	11	8	P	1 min	A1,A5	2012
7	BLUED [63]	Pennsylvania, USA	1 week (Oct)	1	Agg	I, V, switch events	12 kHz	A2,A5	2012
8	TraceBase [47]	Darmstadt, Germany	N/A	15	158 (43 classes)	P, Np	1–8 s	A1,A	2012
9	PSD [51]	Austin, USA	1 week	10	/	P	1 min	A2	2012
10	CRHLL [59]	USA	1 year	16	10	P	1 h	A1,A6	2013
11	IAWE [56]	New delhi, India	73 days (May–Aug)	1	33 (10 classes)	I, V, P, f, S, E, Φ	1 Hz	A1,A6	2013
12	ACS-F1 [66]	Switzerland	1 h (2 sessions)	/	100 (10 types)	I, V, P, Q, f, Φ	10 s	A2	2013
13	AMPds1 [48]	Vancouver, Canada	1 year	1	21	I, V, P, Q, S, pf, F	1 min	A1,A2,A5	2013
14	BERDS [67]	Berkely, USA	1 year	/	4 groups	P, Q, S	20 s	A1	2013
15	ECODS [55]	Switzerland	8 months	6	/	I, V, Φ	1 Hz	A1	2014
16	ECB [50]	Australia	2 years	25	Aggregated	P	1 Hz	A5,A6	2014
17	PLAID [65]	USA	3 months (Summer)	11	60	I, V	30 kHz	A2	2014
18	SustData [43]	Portugal	1144	50	24	I, V, P, Q, S	2 s/ 10 s	A1	2014
19	AMPds2 [49]	Vancouver, Canada	730 days	1	21	I, V, P, S, F, pf	1 min	A1,A2,A5	2014
20	UK-DALE [61]	England, UK	655 days	4	5 (H4), 53 (H1)	P, Aggregated P	6 s/ 6 kHz (Agg)	A1	2015
21	DRED [57]	Netherland	6 months (Jul–Dec)	1	13	P, T, H, Ws, Pr, Agg	1 min/ 1 Hz (for Agg)	A1,A3,A4	2015
22	GREEND [54]	Italy & Austria	6 months	8	9	P	1sec Hz	A1,A5	2015
23	REFIT [44]	England, UK	213 days	20	9, Agg	P, pf, T, O, L, EC ($)	8 s	A5	2015
24	OCTES [46]	Scotland, UK	4–13 months	33	Agg	P, EC ($)	7 s	A5	2015
25	COOLL [110]	France	2 h	1	46 (12 groups)	I, V	100 kHz	A2	2016
26	MEULPv2 [53]	Canada	1 year	23	5 groups	P	1 min	A1	2017
27	RAE [21]	Canada	72 days	1	24	O, V, P, Q, S, f, E	1 Hz	A1,A6	2018
28	DISEC [58]	New Delhi, India	284	19	/	P, Wt	30 s/15, 30, 60 min (Agg)	A1,A5	2018
29	BLOND [64]	Germany	213 days	1	53 (16 groups)	I, V, P	6.4 kSps/54 kSps (Agg)	A1,A2	2018
30	HUE [60]	B. Columbia, Canada	3 years	5	/	p	1 h, H1(1 min), H2 (1 Hz)	A1	2019
31	ENERTALK [111]	Seoul, South Korea	122 days	22	1–7 (Agg)	p	1 Hz, 15 Hz	A1,A5	2019
32	QUD	Doha, Qatar	3 months–1 year	3	4	P, H, T, O	3 s–30 min	A1,A3,A4,A7	2019

anomaly detection, it is of utmost importance that it includes labels annotating normal and anomaly consumption footprints to train developed algorithms. Lastly, for energy demand prediction, collecting power consumption at appliance-level or aggregated level will be appreciated, however, the collection period should be long to be useful.

2.4. Characteristics comparison of existing datasets

Aiming to extract representative outputs and relevant interpretations, a deep comparison study of existing building power consumption datasets is conducted in this section. Various dataset properties are investigated, which have a great importance when collecting data for developing energy efficiency solutions. Table 1 presents a comparative investigation of existing power consumption datasets. The analysis is built based on various characteristics that were collected in each dataset, including the region and period of collection, number of monitoring houses, number of monitoring appliances per house, collected features, sampling rate and release year. Additionally, we check and compare collected features for each database.

2.5. Data collection platforms

Data collection platforms used to glean big energy consumption fingerprints are significantly impacting the energy efficiency systems. Specifically, sensing devices and attached platforms have a big role in gathering and safely storing data in appropriate databases. In this line, in this subsection, we focus on inspecting different architecture platforms used in the literature to collect energy consumption datasets and their properties, including wireless capability, data logging process and data storage. In addition, because of the nature of collected data and their public access capability, privacy concerns are of utmost importance when producing datasets. Specifically, transmitting and sharing individuals' real-time power usage footprints and further their identities are probably quite harmful. To that end, it is important to investigate if the connections to the servers are secure or not in the presented dataset platforms. It is worthy to mention that in this section we focus on analyzing hardware architectures and related modules for only the datasets from Table 1, which present a description of their implemented platforms.

In [43], a power consumption monitoring and feedback platform is deployed, which is based on the use of sensors and a notebook for recordings data, storing them on MongoDB database, performing calculations and providing feedback to end-users. In [44], readings from several smart appliances are collected and transmitted using a commercial communication gateway called Vera3 smart home controller. The latter uses an encryption protocol to transmit data before their storage in a MySQL database. In [49], electricity footprints are gleaned using industrial meters and transmitted using a commercial platform named Obvius AcquiSuite EMB A8810, which includes many feature, among them the security provision. After that, they are stored offsite on a MySQL database server. In [54], platforms based on Raspberry Pi or BeagleBone along with a Plugwise Basic kit4 are used, in which collected data from sensing outlets are transmitted via a Zigbee network. Collected data are then stored on via a remote storage on a MySQL server without considering privacy concerns. In [62], a wireless plug monitoring device with an off-the-shelf system are used to collect power consumption data before transmitting them to central server. To keep the privacy of end-users, REDD dataset has focused only on hiding the identity of end-users and without deploying any secure protocol for data transmission.

In [21], power consumption readings of several appliances are wirelessly gathered using a data acquisition platform based on a Raspberry Pi 2B. Then, data are locally stored on an USB drive. In [61], a Nanode platform is used to wirelessly collect consumption data from individual appliance monitors and current transformers. Following, gleaned data are stored in a Nanode base station. It is worthy to mention privacy issues have not been considered. In [111], consumption records are acquired using a commercial plug, namely ENERTALK PLUG, which includes a microcontroller unit to process and save them in a device storage unit. After that, they are wirelessly transmitted to a data collector server. Finally, data are saved on a NoSQL Hadoop database server.

In this framework power consumption is measured using sub-meters components such as NodeMCU and SEN-11005 current transformer. Furthermore, occupancy patterns, luminosity, temperature and humidity data are also recorded using smart-sensors and then transmitted wirelessly using Raspberry Pi 4 Model B platform. The latter includes a No SQL CouchDB server that is used to store the gathered data using the JavaScript Object Notation (JSON). JSON represents a vastly used text format for data exchange, which keeps data structure without adding notation overhead. Table 2 summarizes the properties of hardware platforms used to collect different datasets, including wireless capability, data logging process, data storage and privacy consideration.

3. Discussion and important findings

3.1. Discussion

Under this framework, a large number of building power consumption databases have been described, reviewed and evaluated according to different parameters as indicated in Table 1. In what follows, we derive pros and cons of each dataset, based on what has been discussed in the previous lines. This can adequately guides us to map recommendations for enriching and improving energy consumption databases.

Table 2
Example of data collection platforms and their properties used in different datasets.

Dataset	Platfom	Wireless capability	Data logging	Data storage	Privacy consideration
REDD	Laptop + smart meters	yes	–	Hard drive	no
GREEND	Raspberry/BeagleBone	yes	JSON	MySQL server	no
SustData	Laptop + smart meters	non	JSON	MongoDB server	no
REFIT	Vera3 smart home controller Vera3 + smart plugs	yes	JSON	MySQL server	yes
AMPDS	Obvius AcquiSuite EMB A8810	non	SQL	MySQL server	yes
RAE	Raspberry Pi 2B + sub-meters	yes	XML	Local storage (USB drive)	no
UK-DALE	Nanode (Atmel ATmega328P)	yes	JSON	Nanode base station	no
ENERTALK	ENERTALK PLUG	yes	Hadoop	NoSQL Hdoop database	no
QUD	NodeMCU + Raspberry Pi 4B + smart sensors	yes	JSON	No-SQL CouchDB server	no

- The biggest databases in terms of length and period of study are UMSM, HES, SustData and REFIT. Otherwise, for the case of HES, the observing period is too short and the sampling frequency of 2 min is a bit big. The same for UMSM, where data are gathered at a sampling rate of 1 min. Therefore, these datasets are inadequate for energy disaggregation as it will be difficult to differentiate between individual devices and occurrences. In contrast, these two repositories provide properties data about monitored homes, among others, the nature of building, size and rooms number and occupants number. Moreover, even if SustData and REFIT use a sampling rate of 8 s and 10 s, this is still not enough when conducting a real-time monitoring.
- In some databases, e.g. PLAID, REDD and BLUED, high frequency monitoring is proceeded for only a few number of houses. This draws upon the prerequisites of energy disaggregation, where comprehensive characteristics catching transitory behavior can be extracted when high frequency collection is explored.
- The majority of databases were gathered in the USA and Canada, under a 120 V voltage and European nations under 230 V. It can be deduced from Table 1 that the existing databases are collected in 13 different countries which are located in four continents; inter alia, America, Europe, Asia and Australia. In this context, these real databases have been produced in distinct climate zones, which cover humid regions (UMSM, REDD, BLUED), humid semitropical (IAWE), marine west coast atmosphere (UK-DALE,HES, AMPds1, REFIT, Tracebase, BLOND, IHEPCDS and OCTES), Mediterranean weather (BERDS) and arid zone (ECB ad QUD).

  However, no databases from Africa countries have been gathered under this investigation since there is no work in the literature who treat this topic in such countries. Moreover, to the best of the authors knowledge, QUD is the first dataset in the Middle East, where ordinarily 240 V voltage is used. Also, some collected particularities; for instance, the climate and environmental data depend on the location of the monitoring campaign.
- The number and nature of monitored appliances, just like the number of observed houses essentially restrain the final usage of databases. In particular, a high number of houses and appliances is required for statistical inspections. In this case, UMSM, Dataport, OCTES, TraceBase, REFIT and HES are the most suitable databases. Plus, some datasets supervise various houses through multiple time intervals leading to difficulty and even impracticality while comparing between different homes. More than that, the setting under which domestic equipments are employed throughout the day is a basic operator for analyzing the complexity of the usage. This way, experimental campaign should be conducted in real conditions such as households, laboratories, or offices as opposed to simulated environments.
- Some databases collect short-term energy consumption and only deliver records of real power, this is the case of COOLL, PSD, ACS-F1 and BLUED. Eventually, seasonal energy usage attitude can not be captured for short-term periods. In this aspect, making use of these databases to track power consumption behavior of end-users is not suitable.
- A number of databases, among them UMSM, IAWE, ACS-F1, AMPds1, RAE and DRED have furnished a set of electric parameters, including I, V, P, Q, E, f and ϕ. Additionally to these records, other conditions such as T, O and L are also reported in QUD and REFIT datasets. The latter provides also analytic information notably related to the monitored electrical appliances and integrates statistics about daily activities in dwelling and residential environments, as well. This endows better a interpretive depth comparable to identical repositories (REDD, BLUED, GREEND).

- Most of the studied datasets did not capture the exogenous conditions, such as the weather temperature, humidity, which can affect effectively the energy consumption. However, while the REFIT dataset has identical properties to OCTES and ACS-F1 datasets, it is also different because it adds other environmental data including the temperature, light, and motion patterns. In addition, household reviews are also reported, which include the surface, age, heating system, insulation, nature of buildings along with other data specifying the number of individuals, job quality and age. In this context, quantitative statistics gleaned from the reviews with occupants' statistics offer more possibilities to researchers to study the influence of other parameters on energy consumption.
- There is a lack of available publicly annotated power consumption datasets to train/learn anomaly detection algorithms, in which power consumption variables are clearly labeled as normal or anomalous. Specifically, all the investigated datatsets in this framework except QUD do not encompass labels that identify normal or abnormal consumption, and thereby they can only be used to train unsupervised anomaly detection algorithms because they do not require annotated datasets.
- Privacy and security concerns have not seriously been considered in most of the existing datasets. This is due to the fact that conventional meters required to be physically accessed and they registered power consumption for longer time periods (i.e. the real-time monitoring was not considered).

3.2. Qatar university dataset (QUD)

Using the pros and cons of the state-of-the-art datasets presented in the previous section, a measurement campaign has been conducted in the Qatar university energy lab to glean QUD repository. Specifically, in order to compensate the undersupply of appliance-level datasets dedicated for energy efficiency and anomaly detection in power consumption, a real-time micro-moment laboratory has been developed to gather accurate power usage footprints. Put simply, QUD is a set of consumption records from different installed electrical devices (e.g. air conditioner, heating system, desktop and light lamps) in addition to contextual data; including humidity, temperature, room occupancy and ambient light intensity. To the best of our knowledge, QUD is the first dataset in the Middle East, in which consumption data are collected at an ordinarily 240 V voltage. This dataset have multiple usage scenarios such as detecting consumption abnormalities, testing recommender systems and assessing innovative visualization tools. Moreover, it is worth noting that QUD is among the first annotated repositories dedicated for anomaly detection in power consumption.

Therefore, the time-series data representing power consumption footprints for two appliances are registered along with corresponding cubicle occupancy, indoor temperature, indoor humidity, and luminosity. In order to label QUD consumption observations, the micro-moment paradigm is used which helps in identifying the moments of good or anomalous usage. Specifically, the

Table 3
Micro-moments assumption and labeling

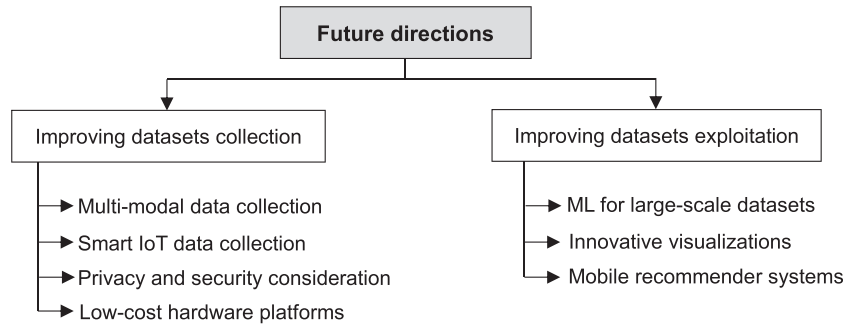| Micro-moment	Label	Description
Good usage	0	Non-excessive usage
Turn on	1	Switching on a device
Turn off	2	Switching off a device
Excessive consumption	3	Consumption >95% of device's maximum active power consumption level
Consumption when outside	4	Device consumption without the presence of the end-user

Fig. 3. Future directions to improve both datasets collection and exploitation.
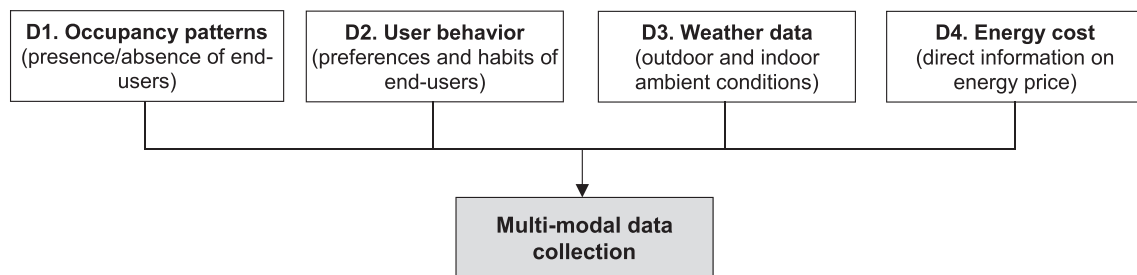

Fig. 4. Principal factors impacting the power consumption in buildings and contributing in the multi-modal data collection.

micro-moments are deployed to come up with accurate statistics about consumers [10,34]. Using this dataset, the power consumption observations are labeled via the use of five micro-moment classes according to a set of standards out of the yielded appliance. These five micro moments are defined as; "good usage", "turn on", "turn off", "excessive power consumption", and "consumption when outside". The last two micro-moments represent anomalous consumption behaviors that are leading to much wasted energy. Table 3 describes the micro-moment classes and labels used in QUD (QUD can be accessed through: http://em3.i-know.org/data-sets/). In addition, it is worthy to mention that the micro-moment "consumption when outside" is limited to a set of appliances, such as air conditioners, televisions, light lamps, desktops/laptops, and fans, in which the end-user should be present during their operation to not be considered as an anomalous consumption [112].

4. Future directions

After analyzing, comparing and capturing pros and cons of existing datasets, a set of important orientations that can improve data collection and enrich datasets' content are identified. In addition, other directions to improve datasets exploitation are described as well. Fig. 3 summarizes the future directions that are identified to improve both datasets collection and exploitation.

4.1. Improving the dataset collection

In order to develop powerful energy efficiency systems, it is of paramount importance to improve dataset collection procedures and hence enhance the content of collected data. In this respect, the following recommendations and directions can be establish:

4.1.1. Multi-modal data collection

Multi-modal data collection means merely collecting more than one type of data to accomplish an efficient energy saving task or other related applications. Specifically, power consumption in

buildings depends on multiple factors, which should be gleaned together power consumption footprints in order to design comprehensive datasets [113,114]. Fig. 4 summarizes the principal parameters impacting the power consumption in buildings and contributing in the multi-modal data collection.

D1. Occupancy patterns: Domestic residents utilize more energy when they're occupied. Even this may appear glaringly evident, collecting occupancy data is a serious matter that must be inspected when searching for wasteful energy aspects in households. Specifically, we ensure that these structures consume less power when unoccupied. Individuals in households influence power consumption for the most part via lighting, cooling, heating and other plug loads. Analyzing power consumption for the duration of the day demonstrates an immediate relationship amongst occupancy and power usage. For the moments when individuals are in a household, different rooms are conditioned or heated to an agreeable temperature. Of course, normal day-by-day activities need also power usage. The effect of individuals utilizing energy in a household is the reason we underscore the relevance of individuals turning off unused appliances or other devices in unoccupied rooms [115,116]. In this regard, the use of occupancy sensors is highly recommended in households or other buildings such as offices, laboratories or campus buildings to sense when someone is present or not and then turn off the appliances accordingly. By this way, a lot of energy can be preserved when the absence of individuals is confirmed.

D2. User behavior: Comprehending and improving individual power consumption behavior is among the successful approaches to reduce energy demand and encourage energy preserving. In fact, user behavior can be responsible of about 20–50% of the consumption level [117,118]. Therefore, collecting and inserting data about end-users' behaviors in the energy efficiency model can significantly decrease wasted energy. This can be done through gathering information related to their preferences and habits [119–121].

D3. Weather data: Relation between weather circumstances and power consumption has been proved in several works [122–124]. As a matter of exemplification, peak energy demand during

heat waves is widely seen in so many hot countries. For that reason, gathering weather data is regarded as crucial while investigating user behavior. Over and above, existing and newly built households will certainly undergo the impact of climate change. Accordingly, collection and measurement of new energy consumption databases of these houses ought to consider weather patterns that integrate certain repercussions of climate change, rather than only considering historical climate information [125–127].

D4. Energy cost: Estimating the cost of household power consumption and providing this data to end-users can motivate them improving their behavior [128]. Forecasting the user's electricity bill and integrating energy price signals in energy efficiency applications can effectively increase power saving [129–131] since it helps the consumer to cognitively bridge the gap between consumption and cost. Moreover, collecting energy price profiles at the appliance level makes it unambiguous for the user which appliance raises more the cost. As a result, the consumer can relatively behave in order to reduce wasted energy.

4.1.2. Smart IoT data collection

Conventional meters are not able to gather the type of granular and device-level data, however, this becomes possible today with smart meters. In this line, in order to achieve target requirements in relation to data accuracy and further supporting real-time data collection and analysis, deploying smart meters and Internet of things (IoT) sensors to ensure a smart IoT data collection strategy is of paramount importance [132]. This helps in optimizing the communication, storage and computing resources.

4.1.3. Low-cost hardware platforms

In order to reduce the cost of datasets collection, the use of hardware platforms, enabling more cost-effective and powerful alternatives to process and transmit collected data is a high priority, such as the Raspberry PI 4 (RPI4) model B [133], ODROID-XU4 [134] and Jetson TX1 [135]. Those platforms can monitor energy consumption data along with collecting other essential contextual information, which ultimately results in a larger pool of data.

4.1.4. Privacy and security consideration

To preserve the end-users' privacy, power consumption footprints and end-users personal information should be protected. Personal data related to end-user specific power consumption patterns can be exploited to identify and supervise behavior patterns inside buildings (households or public structures). This is possible since electrical devices e.g. the microwave, air condition, washing machine, dishwasher, etc. can be detected and recognized from their power consumption fingerprints [2]. Therefore, personal data related to consumption signatures may be deployed to carry out real-time surveillance of end-users. In this regard, the data collection process must encourage producing challenging datasets and make power consumption statistics available to end-users and energy providers while respecting end-users' personal privacy and security. To that end, adopting robust techniques to remove personal information is a must, including encryption, steganography and aggregation.

4.2. Improving the dataset exploitation

Almost energy efficiency systems are built and validated using energy consumption datasets, which make them very important. Further, with the increasing amount of data collected in each database, the need for challenging solutions that can extract comprehensive information is becoming inevitable [136]. In this section we present three main directions, which can be investigated to ameliorate energy saving initiatives. It is worthy to mention that although the following directions are from a consumer's perspective, however, they are valuable for both consumers and energy providers. Specifically, they are generally developed by the latter and deployed to the benefit of end-users to help them in optimizing their energy usage. In addition, as discussed in Section 1, consumers are responsible for wasting more than 20% of the total energy consumed in buildings [24–26].

4.2.1. Mobile recommender systems

Lastly, it is noticed that mobile smart devices are becoming an indispensable part of our daily life. Unlike earlier mobile phones that provide limited functionality, smart phones can do a variety of very useful jobs. With the widespread usage of smartphones and the fast growing of the internet and network facilities, a massive amount of data is produced. Consequently, modern societies have started the age of Big Data through successfully discovering users' possible demand and preferences. This has raised the necessity for data scientists and energy management stakeholders to conduct studies on mobile recommender systems for controlling users' energy efficiency [137].

Recommendation systems are commonly deployed to polish the use of smartphones and to assist in dealing with the large amount of data through establishing appropriate advices using recommendation schemes and contextual information. In this regards, the role of recommender systems will be essential to promote energy efficiency and help end-users understanding and improving their consumption footprints [138]. More specifically, every particular recommender application is generally elaborated with an explicit context in mind with the aim of solving in some sense the data overloading issue due to the large-scale datasets of power consumption. The effectiveness of a mobile recommender system has been demonstrated through real-use applications in academic buildings [137], in which a context-aware based recommender app is developed to help in supporting end-users to transform their energy consumption habits.

Furthermore, the architecture of recommendation systems that is usually based on interactional models, graphic user interfaces and recommendation engines makes them productive and useful to deal with energy efficiency applications [139]. To summarize, the use of mobile recommender systems is recommended to improve power consumption datasets exploitation via:

- Developing explainable recommender systems can be very supportive to improve data exploitation with a view to replacing inefficient energy habits with efficient ones. An explainable recommender system aims at providing end-users with tailored recommendations, followed by explanations about them [140]. Explanations refer to the motivations behind the recommendation or to the benefits from providing the recommended action or advice. They can enhance the persuasiveness of the system, end-users' understanding and satisfaction and provide an immediate reward to them.
- Developing intelligent mobile home monitoring systems using collected data to provide information and monitoring options to the end-user to help him control its load usage, visualize consumption statistics and compare them to those of other users, and further predict the overall charge of monthly bills [141].

4.2.2. Visualization for understanding user behavior

Visualization is seen to be the most effective way to assimilate increasingly large datasets with the aim of interactively and perfectly conveying insights to end-users, consumers, and stakeholders in general. Recent tools, methods, and softwares leveraged for visualization of energy consumption require further improvements to remain more important in a planet with larger low-carbon emissions. Moreover, they are required to sensitize energy-consuming behavior in an approachable and stimulating way. In this context,
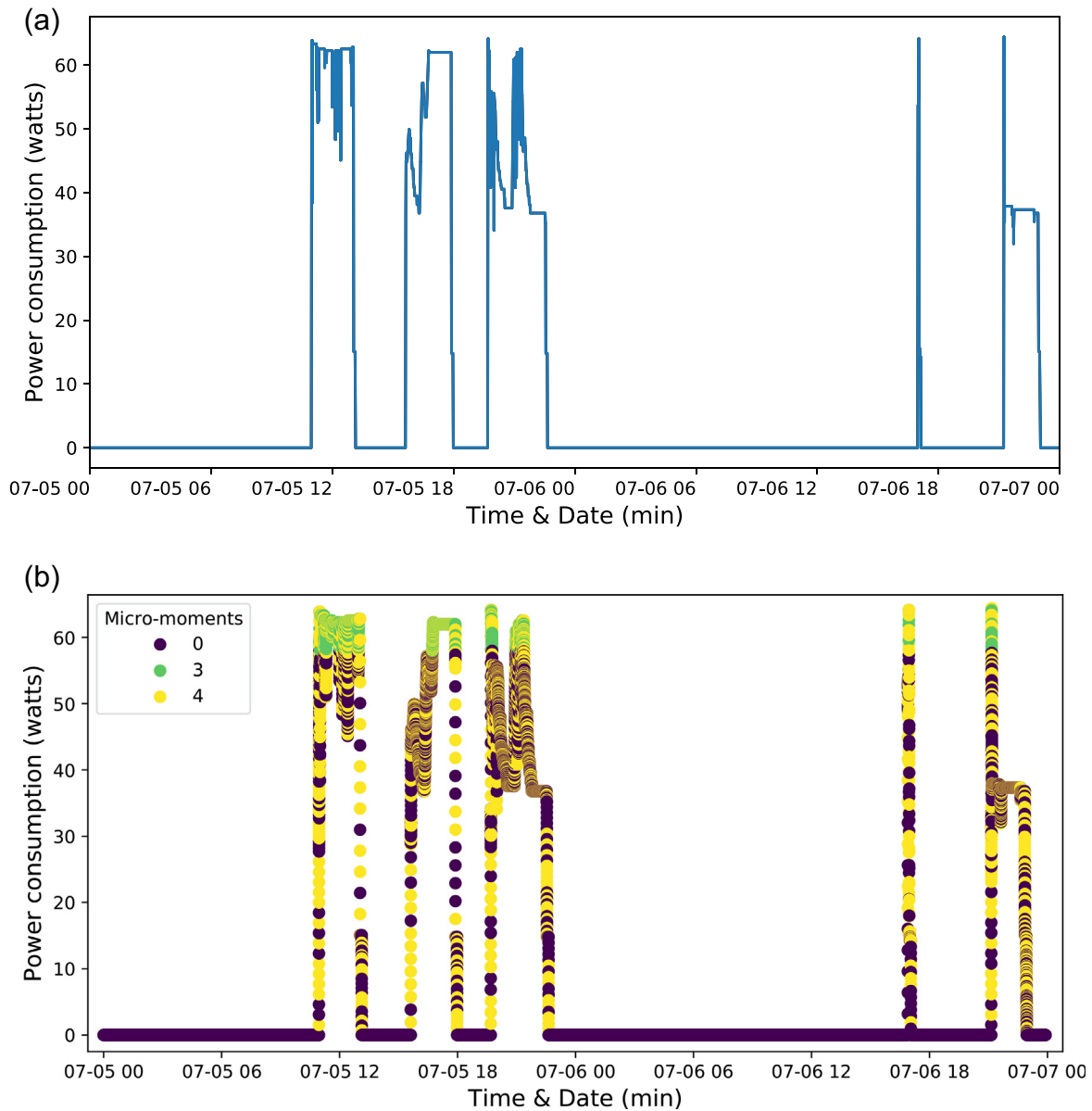
(a)

(b)

Fig. 5. Time-series power consumption of a television and its micro-moments scatter plot from DRED: top) time-series power consumption, and bottom) micro-moments scatter plot at a sampling rate of 3 min.

we present in this section an example of a novel visualization approach based on micro-moments analysis. Fig. 5 displays a time-series energy consumption of a television and its micro-moments scatter plot at sampling intervals of 3 min, recorded in DRED dataset. This novel visualization strategy is presented as an example, in which energy usage micro-moment classes of 2 days are captured and plotted, defined as: good usage (class 0), excessive usage (class 3) and consumption while outside (class 4). Users can seamlessly get the plots at different sampling rate starting from the milliseconds.

As it can be deduced, tracing micro-moments through time patterns facilitates identifying moments of abnormal consumption and then makes it easy to establish precise guidelines helping to reduce energy waste. Moreover, this helps end-users understanding their consumption footprints, increasing their awareness, and hence triggering them to improve their behavior through the use of tailored recommendations. In addition anomalous consumption behaviors can be identified when an adequate visualization tool is

adopted, e.g. the micro-moments visualization, and hence end-users can improve their behaviors based on the detected anomaly. Moreover, it is worth noting that the use of the micro-moments paradigm to detect anomalous consumption can be enlarged to identify other kinds of anomalies, e.g. detecting abnormal consumption of an air conditioner while doors/windows are open via considering other information sources. Therefore, end-users will be provided with the appropriate notifications and advices, i.e. close doors/windows to reduce wasted energy.

In addition, a set of valuable recommendations and future directions towards designing effective visualizations aiming to increase end-users energy awareness is summarized as follows:

- Visualizations need to catch the attention of their users via using bright colors, contrasts and varied views, where somethings are changing constantly. More importantly, they should implement colors that are legible for people with color vision deficiencies [142].

Table 4
List of abbreviations and nomenclatures used in this paper.

Acronym	Description	Nomenclature	Description
UMSM	UMass Smart* Microgrid	I	current
IHEPCDS	Individual household electric power consumption data set	V	voltage
IHEPCD	Household electricity survey	P	active power
REFIT	Personalized retrofit decision support tools for UK homes	Q	reactive power
ORBEET	Opportunities for community groups through energy storage	S	apparent power
AMPds1	Almanac of minutely power dataset version 1	Np	normalized power
ECB	Electricity consumption benchmarks	E	energy
PSD	Pecan street dataset	f	frequency
MEULPv.1	Measured end-use electric Load profiles version 1	ϕ	phase angle
RAE	Rainforest automation energy	pf	power factor
GREEND	Green dataset	EC	energy cost
ECO	Electricity consumption and occupancy	Wt	weather
RDED	Dutch residential energy dataset	T	temperature
IWAE	Indian dataset for ambient water and energy	H	humidity
DISEC	Dataset on information strategies for energy conservation	O	occupancy
CRHLP	Commercial and residential hourly load profiles dataset	L	light level
HUE	Hourly usage energy dataset		
UK-DALE	UK domestic appliance-level electricity		
REDD	Reference energy disaggregation dataset		
BLUED	Building-level fully labeled electricity disaggregation dataset		
BLOND	Building-level office environment dataset		
PLAID	Plug Load Appliance Identification Dataset		
ACS-F1	Appliance consumption signatures version 1		
ENERTALK	Energy talks		
DRED	Dutch residential energy dataset		
REFIT	Retrofit decision support tools for UK homes		
OCTES	Opportunities for community groups through energy storage		
COOLL	Controlled On/Off Loads Library		
MEULPv.2	Measured end-use electric Load profiles version 2		
BERDS	Berkeley energy disaggregation dataset		
CRHLP	Commercial and residential hourly load profiles dataset		
IoT	Internet of things		
LDA	Linear discriminant analysis		
SVM	Support vector machine		
KNN	K-nearest neighbors		
DT	Decision tree		
EBT	Ensemble bagged tree		
DNN	Deep neural networks		
MLP	Multilayer perceptron		
LR	Logistic regression		
NILM	Non-intrusive load monitoring		
JSON	JavaScript Object Notation		
SQL	Structured query language		
ML	Machine learning		
GANs	Generative adversarial networks		
$(EM)^3$	Endorsing energy efficiency with micro-moments		

- Visualizations also need to employ comparisons between different time period consumptions to stimulate energy users' reflection and learning. Individuals are highly motivated in preserving power usage through making comparison of their current consumption to their own previous consumptions.
- Fragile long time-series aggregated systems for visualizing energy consumption are better to be avoided. By nature, the human being's mind can not always memorize all what they do in each second or which appliance do they use more in their daily routine. Hence, it was concluded by many studies that:
  a. Collecting appliance-level consumption data is highly needed for consumers to cognitively decode their consumption data and to finally make a decision towards changing their behavior [143].
  b. Instantaneous short-time intervals are the best way to aid the consumer to easily figure out at what instance their energy cost increased [144]. This makes it possible for the users to ameliorate their behavior and to notice the direct effect from it.
- 3D visualizations of households with real-time energy consumption statistics can impart better contextual understanding to homeowners and help them control practices shaping their consumption. In relation, it is also crucial to support that by exploring the deployment of other emerging technologies, including virtual and augmented reality, interactive visualizations and wall sized presentations.

4.2.3. ML algorithms of large-scale datasets

Another important direction that can improve the quality and exploitation of power consumption datasets relies on deploying ML algorithms, which can help significantly in reducing energy consumption and thereby decreasing energy costs and carbon dioxide emissions [145,146]. Therefore, boosting novel progresses and challenges with regard to ML models is of paramount importance. In this context, several directions could be identified in which ML play a major role when shifting to more sustainable end energy efficiency environment, among them:

- The use of generative models, such as generative adversarial networks (GANs), which can tremendously improve the quality of collected data via completing incomplete power consumption signals (due to data loss occurred during the collection step) and hence leading to a better exploitation of the produced datasets in different applications [147].
- The use of deep learning models to identify consumption anomalies via classifying the micro-moment classes of end-users' power consumption [34]. These algorithms assist in analyzing consumption footprints to detect abnormal consumption related to "excessive consumption" and "consumption when outside". After defining the micro-moments described in Section 4.2.2, Deep neural networks (DNN) and other ML algorithms with different configuration parameters could be deployed, including logistic regression, linear discriminant analysis (LDA), support vector machine (SVM), K-nearest neighbors (KNN), decision Tree (DT) and ensemble bagged tree (EBT. The selection of the appropriate ML model is mainly based on ensuring the best compromise between the identification performance and computational complexity. Therefore, this results in a better exploitation of the collected datasets for identifying abnormal power consumption.

5. Conclusion

Existing building power consumption datasets have been reviewed together with their merits and drawbacks within the lines of this work. Comprehensive comparisons between these datasets have been conducted in terms of different factors and data collection platforms that categorize each of these. Based on the fruitful analysis gleaned from these comparisons, a novel annotated dataset has been presented, namely QUD, which can be very useful for power consumption anomaly detection since it includes labels of good and anomalous usage. Moreover, we came with recommendations and future directives for improving the quality and the content of future power consumption datasets. Thus, a set of relevant orientations have been identified as follows:

- Adopting a multi-modal data collection, which is based on gathering various data sources because the energy consumption is affected by different factors. Therefore, data should be gathered from several geographical positions with regard to ambient conditions, atmospheric and environmental resources, occupation patterns, user preferences and energy cost.
- Adopting smart IoT data collection strategies to collect data from various IoT sensors at a low sampling rate, which leads therefore to developing real-time and scalable energy saving systems.
- Collecting more annotated anomaly detection datasets in order to encourage testing and developing anomaly detection algorithms. Overall, detecting anomalous power consumption plays a major role in reducing wasted energy.
- Producing comprehensive databases by reference to the level of consumption (especially at the appliance-level) and the duration of the collection campaign (i.e. collecting data for the entire seasonal periods of the year) while respecting end-users personal privacy and security
- Formulating protocols and standards to characterize building power consumptions datasets that can help make unified meta-data strategies and terminologies, facilitating the comparisons, and rigorously understanding the state-of-the-art.

In addition, genuine initiatives to improve datasets exploitation and utilization have been identified in this framework. Consequently, a novel visualization strategy has been presented based on the micro-moments analysis, which enables people to comprehend their own power usage footprints, and accordingly interpret their electricity consuming behavior. Moreover, it helps them easily getting statistics on their actual power consumption. Moreover, another example of using ML algorithms has been introduced to classify power consumption micro-moments and detect anomalous usage, such as "excessive consumption" or "consumption while outside". These two behaviors are responsible for wasting a large amount of energy. Consequently, it will be part of our future work to improve datasets exploitation and help promoting energy efficiency through:

- The use of novel ML algorithms, including deep learning and generative adversial networks (GAN), which can effectively deal with imbalanced and large-scale datasets.
- The development of innovative visualization tools that can cognitively improve end-users comprehension of their consumption behavior. Therefore, interactive visualization tools will be integrated into smart power consumption dashboards that permit individuals to engage with, apprehend their energy usage and translate them into positive actions throughout their everyday life.

- The deployment of explainable recommender systems in order to trigger action recommendations at the correct moment, especially if appropriate hardware materials are used to collect and analyze data in a real-time manner.

Declaration of Competing Interest

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

Acknowledgements

Appendix A

Abbreviation description of the power consumption datasets along with a description of the nomenclatures considered in this paper are summarized in Table 4.

References

[1] T. Chaudhuri, Y.C. Soh, H. Li, L. Xie, A feedforward neural network based indoor-climate control framework for thermal comfort and energy saving in buildings, Applied Energy 248 (2019) 44–53.

[2] Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, Robust event-based non-intrusive appliance recognition using multi-scale wavelet packet tree and ensemble bagging tree, Applied Energy 267 (2020) 114877.

[3] M.S. Gul, S. Patidar, Understanding the energy consumption and occupancy of a multi-purpose academic building, Energy and Buildings 87 (2015) 155–165.

[4] X.M. Zhang, K. Grolinger K, M.A.M. Capretz, L. Seewald, Forecasting residential energy consumption: single household perspective, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 110–117..

[5] M. Brogger, K.B. Wittchen, Estimating the energy-saving potential in national building stocks – A methodology review, Renewable and Sustainable Energy Reviews 82 (2018) 1489–1496.

[6] M. Brogger, P. Bacher, H. Madsen, K.B. Wittchen, Estimating the influence of rebound effects on the energy-saving potential in building stocks, Energy and Buildings 181 (2018) 62–74.

[7] M.A.A Mohamed, Saving energy through using green rating system for building commissioning, Energy Procedia 162 (2019) 369–378. Emerging and Renewable Energy: Generation and Automation..

[8] F. Sher, A. Kawai, F. Gulec, H. Sadiq, Sustainable energy saving alternatives in small buildings, Sustainable Energy Technologies and Assessments 32 (2019) 92–99.

[9] T. Jafarinejad, A. Erfani, A. Fathi, M.B. Shafii, Bi-level energy-efficient occupancy profile optimization integrated with demand-driven control strategy: University building energy saving, Sustainable Cities and Society 48 (2019) 101539.

[10] A. Alsalemi, C. Sardianos, F. Bensaali, I. Varlamis, A. Amira, G. Dimitrakopoulos, The role of micro-moments: a survey of habitual behavior change and recommender systems for energy saving, IEEE Systems Journal (2019) 1–12.

[11] W. Al-Marri, A. Al-Habaibeh, M. Watkins, An investigation into domestic energy consumption behaviour and public awareness of renewable energy in Qatar, Sustainable Cities and Society 41 (2018) 639–646.

[12] S.A. Sarkodie, A.O. Crentsil, P.A. Owusu, Does energy consumption follow asymmetric behavior? An assessment of Ghana's energy sector dynamics, Science of The Total Environment 651 (2019) 2886–2898.

[13] K. Zhou, S. Yang, Understanding household energy consumption behavior: The contribution of energy big data analytics, Renewable and Sustainable Energy Reviews 56 (2016) 810–819.

[14] M.H. Ishak, I. Sipan, M. Sapri, A.H.M. Iman, D. Martin, Estimating potential saving with energy consumption behaviour model in higher education institutions, Sustainable Environment Research 26 (6) (2016) 268–273.

[15] T. Csoknyai, J. Legardeur, A.A. Akle, M. Horvath, Analysis of energy consumption profiles in residential buildings and impact assessment of a serious game on occupants' behavior, Energy and Buildings 196 (2019) 1–20.

[16] P.V. Aubel, E. Poll, Smart metering in the Netherlands: What, how, and why, International Journal of Electrical Power & Energy Systems 109 (2019) 719–725.

[17] D.B. Avancini, J.J.P.C. Rodrigues, S.G.B. Martins, R.A.L. Rabelo, J. Al-Muhtadi, P. Solic, Energy meters evolution in smart grids: A review, Journal of Cleaner Production 217 (2019) 702–715.

[18] S. Latif, A. Shabani, A. Esser, A. Martkovich, Analytics of residential electrical energy profile, in: 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017, pp. 1–4.

[19] P.P. Moletsane, T.J. Motlhamme, R. Malekian, D.C. Bogatmoska, Linear regression analysis of energy consumption data for smart homes, in: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 0395–0399..

[20] A. Muhammad Mehar, A. Qumer Gill, K. Matawie, Analytical model for residential predicting energy consumption, in: 2018 IEEE 20th Conference on Business Informatics (CBI), vol. 02, 2018, pp. 82–88..

[21] S. Makonin, Z.J. Wang, C. Tumpach, RAE: The rainforest automation energy dataset for smart grid meter data analysis, Data 3 (1) (2018) 1–9.

[22] J.L. Ramirez-Mendiola, P. Grunewald, N. Eyre, The diversity of residential electricity demand – A comparative analysis of metered and simulated data, Energy and Buildings 151 (2017) 121–131.

[23] A. Alsalemi, Y. Himeur, F. Bensaali, A. Amira, C. Sardianos, I. Varlamis, G. Dimitrakopoulos, Achieving domestic energy efficiency using micro-moments and intelligent recommendations, IEEE Access 8 (2020) 15047–15055.

[24] K. White, R. Habib, D.J. Hardisty, How to SHIFT consumer behaviors to be more sustainable: A literature review and guiding framework, Journal of Marketing 83 (3) (2019) 22–49.

[25] D. Ürge Vorsatz, L.F. Cabeza, S. Serrano, C. Barreneche, K. Petrichenko, Heating and cooling energy trends and drivers in buildings, Renewable and Sustainable Energy Reviews 41 (2015) 85–98.

[26] W. Al-Marri, A. Al-Habaibeh, H. Abdo, Exploring the relationship between energy cost and people's consumption behaviour, Energy Procedia vol. 105, 2017, 3464–3470. 8th International Conference on Applied Energy, ICAE2016, 8–11 October 2016, Beijing, China..

[27] A. Paone, J.P. Bacher, The impact of building occupant behavior on energy efficiency and methods to influence it: a review of the state of the art, Energies 11 (4) (2018).

[28] X. Liu, N. Iftikhar, H. Huo, R. Li, P.S. Nielsen, Two approaches for synthesizing scalable residential energy consumption data, Future Generation Computer Systems 95 (2019) 586–600.

[29] Y. Guo, Z. Tan, H. Chen, G. Li, J. Wang, R. Huang, et al., Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving, Applied Energy 225 (2018) 732–745.

[30] N.T. Ngo, Early predicting cooling loads for energy-efficient design in office buildings by machine learning, Energy and Buildings 182 (2019) 264–273.

[31] X. Xu, W. Wang, T. Hong, J. Chen, Incorporating machine learning with building network analysis to predict multi-building energy use, Energy and Buildings 186 (2019) 80–97.

[32] J.S. Chou, D.S. Tran, Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders, Energy 165 (2018) 709–726.

[33] W. Wang, T. Hong, X. Xu, J. Chen, Z. Liu, N. Xu, Forecasting district-scale energy dynamics through integrating building network and long short-term memory learning algorithm, Applied Energy 248 (2019) 217–230.

[34] A. Alsalemi, M. Ramadan, F. Bensaali, A. Amira, C. Sardianos, I. Varlamis, et al., Endorsing domestic energy saving behavior using micro-moment classification, Applied Energy 250 (2019) 1302–1311.

[35] A.G. Ruzzelli, C. Nicolas, A. Schoofs, G.M.P. O'Hare, Real-time recognition and profiling of appliances through a single electricity sensor, in: 2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2010, pp. 1–9..

[36] Y. Gao, A. Schay, D. Hou, Occupancy detection in smart housing using both aggregated and appliance-specific power consumption data, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1296–1303.

[37] E. Sala, D. Zurita, K. Kampouropoulos, M. Delgado, L. Romeral, Occupancy forecasting for the reduction of HVAC energy consumption in smart buildings, in: IECON 2016–42nd Annual Conference of the IEEE Industrial Electronics Society, 2016, pp. 4002–4007.

[38] S. Ahmadi-Karvigh, A. Ghahramani, B. Becerik-Gerber, L. Soibelman, One size does not fit all: Understanding user preferences for building automation systems, Energy and Buildings. 145 (2017) 163–173.

[39] C. Franco, K. Nielsen, P.J. Kerstens, Uncertainty management for classification and benchmarking of energy-use preference profiles, in: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2018, pp. 1–8.

[40] N. Terry, J. Palmer, Household electricity survey, UK Data Archive Study (2012) 1–31.

[41] K. Bache, M. Lichman, Individual Household Electric Power Consumption Dataset, University of California, School of Information and Computer Science, CA, 2013.

[42] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, J. Albrecht, Smart*: An open data set and tools for enabling research in sustainable homes, in: Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability (SustKDD 2012), 2012, pp. 1–6..

[43] L. Pereira, F. Quintal, R. Goncalves, Nunes NJ. SustData, A Public Dataset for ICT4S Electric Energy Research, in: ICT for Sustainability 2014, (ICT4S-14), Atlantis Press, 2014/08..

[44] D. Murray, J. Liao, L. Stankovic, V. Stankovic, R. Hauxwell-Baldwin, C. Wilson, et al., A data management platform for personalised real-time energy feedback, in: Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting, 2015.

[45] O. Parson, G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, et al., Dataport and NILMTK: A building data set designed for non-intrusive load monitoring, in: 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2015, pp. 210–214.

[46] European Union. Opportunities for Community Groups Through Energy Storage (OCTES), 2013. Accessed: 2019-05-03. URL:http://octes.oamk.fi/final/..

[47] A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, et al., On the accuracy of appliance identification based on distributed load metering data, in: 2012 Sustainable Internet and ICT for Sustainability (SustainIT), 2012, pp. 1–9.

[48] S. Makonin, F. Popowich, L. Bartram, B. Gill, I.V. Bajic, AMPds: A public dataset for load disaggregation and eco-feedback research, in: 2013 IEEE Electrical Power Energy Conference, 2013, pp. 1–6..

[49] I.V.B. Stephen Makonin, Bradley Ellert, F. Popowich, Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014, Scientific Data 3 (180048) (2016) 1–12.

[50] Australian Energy Regulator. Electricity consumption benchmarks; 2014. Data retrieved from data.gov.au, URL:www.energymadeeasy.gov.au..

[51] C. Holcomb, Pecan Street Inc.: a test-bed for NILM, in: International Workshop on Non-Intrusive Load Monitoring, ACM, New York, NY, USA, 2012, pp. 3:1–3:8.

[52] N. Saldanha, I. Beausoleil-Morrison, Measured end-use electric load profiles for 12 Canadian houses at high temporal resolution, Energy and Buildings 49 (2012) 519–530.

[53] G. Johnson, I. Beausoleil-Morrison, Electrical-end-use data from 23 houses sampled each minute for simulating micro-generation systems, Applied Thermal Engineering 114 (2017) 1449–1456.

[54] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, A.M.G.R.E.E.N.D. Tonello, An energy consumption dataset of households in Italy and Austria, in: 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), 2014, pp. 511–516.

[55] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, S. Santini, The ECO data set and the performance of non-intrusive load monitoring algorithms, in: Proceedings of the 1st ACM International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys 2014), Memphis, TN, USA. ACM, 2014, pp. 80–89.

[56] N. Batra, M. Gulati, A. Singh, M.B. Srivastava, It's different: insights into home energy consumption in India, in: Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings, BuildSys'13, New York, NY, USA, ACM, 2013, pp. 3:1–3:8..

[57] A.S.N. Uttama Nambi, A. Reyes Lua, Prasad VR. LocED, Location-aware energy disaggregation framework, in: Proceedings of the 2Nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments. BuildSys '15, ACM, New York, NY, USA, 2015, pp. 45–54.

[58] V.L. Chen, M.A. Delmas, S.L. Locke, A. Singh, Dataset on information strategies for energy conservation: A field experiment in India, Data in Brief 16 (2018) 713–716.

[59] Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States, Accessed: 2019-05-30, URL:https://openei.org/datasets/files/961/pub/..

[60] S. Makonin, HUE: The hourly usage of energy dataset for buildings in British Columbia, Data in Brief 23 (2019) 103744.

[61] J. Kelly, W. Knottenbelti, The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes, Scientific Data 2 (150007) (2015) 1–14.

[62] J.Z.R.E.D.D. Kolter, A public data set for energy disaggregation research, in: Proceedings of the 1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD), ACM, San Diego, CA, USA, 2011.

[63] K. Anderson, A. Ocneanu, D.R. Carlson, A. Rowe, M. Bergés, BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research, in: Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD), ACM, Beijing, China, 2012.

[64] T. Kriechbaumer, H.A. Jacobsen, BLOND, a building-level office environment dataset of typical electrical appliances, Scientific Data 5 (180048) (2018) 1–14.

[65] J. Gao, S. Giri, E.C. Kara, M.P.L.A.I.D. Bergés, A public dataset of high-resolution electrical appliance measurements for load identification research: demo abstract, in: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings. BuildSys '14, ACM, New York, NY, USA, 2014, pp. 198–199.

[66] C. Gisler, A. Ridi, D. Zufferey, O.A. Khaled, J. Hennebert, Appliance consumption signature database and recognition test protocols, in: 2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA), 2013, pp. 336–341..

[67] M. Maasoumy, B.M. Sanandaji, K. Poolla, A.S. Vincentelli, in: BERDS-BERkeley EneRgy Disaggregation Data Set, University of California, Berkeley, 2013.

[68] A. Alsalemi, F. Bensaali, A. Amira, N. Fetais, C. Sardianos, I. Varlamis, Using micro-moments to visualize domestic energy usage, in: Intelligent Systems Conference (IntelliSys-2019), London, UK, 2019.

[69] C. Sardianos, I. Varlamis, G. Dimitrakopoulos, D. Anagnostopoulos, A. Alsalemi, F. Bensaali et al. I want to.... change' Micro-moment based recommendations can change users' energy habits, in: 8th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS 2019), Crete, Greece, 2019..

[70] Y. Himeur, A. Elsalemi, F. Bensaali, A. Amira, Improving in-home appliance identification using fuzzy-neighbors-preserving analysis based QR-decomposition, in: International Congress on Information and Communication Technology (ICICT), 2020, pp. 1–8.

[71] F. Rossier, P. Lang, J. Hennebert, Near real-time appliance recognition using low frequency monitoring and active learning methods, Energy Procedia 122 (2017) 691–696.

[72] S.C. Lee, G.Y. Lin, W.R. Jih, J.Y.J. Hsu, Appliance recognition and unattended appliance detection for energy conservation, in: Proceedings of the 5th AAAI Conference on Plan, Activity, and Intent Recognition, AAAIWS'10-05, AAAI Press, 2010, pp. 37–44..

[73] M. Kahl, A. Ul Haq, T. Kriechbaumer, H.A.A. Jacobsen, Comprehensive feature study for appliance recognition on high frequency energy data, in: Proceedings of the Eighth International Conference on Future Energy Systems. e-Energy '17, ACM, New York, NY, USA, 2017, pp. 121–131.

[74] C. Sardianos, I. Varlamis, C. Chronis, G. Dimitrakopoulos, Y. Himeur, A. Alsalemi, et al. A model for predicting room occupancy based on motion sensor data, in: 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 394–399..

[75] Y. Wei, L. Xia, S. Pan, J. Wu, X. Zhang, M. Han, et al., Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks, Applied Energy 240 (2019) 276–294.

[76] J. Ahmad, H. Larijani, R. Emmanuel, M. Mannion, A. Javed, Occupancy detection in non-residential buildings – A survey and novel privacy preserved occupancy monitoring solution, Applied Computing and Informatics (2018).

[77] Z. Chen, C. Jiang, L. Xie, Building occupancy estimation and detection: A review, Energy and Buildings 169 (2018) 260–270.

[78] T. Vafeiadis, S. Zikos, G. Stavropoulos, D. Ioannidis, S. Krinidis, D. Tzovaras, et al., Machine learning based occupancy detection via the use of smart meters, in: 2017 International Symposium on Computer Science and Intelligent Controls (ISCSIC), 2017, pp. 6–12.

[79] S. Khashe, A. Heydarian, D. Gerber, B. Becerik-Gerber, T. Hayes, W. Wood, Influence of LEED branding on building occupants' pro-environmental behavior, Building and Environment 94 (2015) 477–488.

[80] G. Tang, Z. Ling, F. Li, D. Tang, J. Tang, Occupancy-aided energy disaggregation, Computer Networks 117 (2017) 42–51, Cyber-physical systems and context-aware sensing and computing..

[81] V. Breschi, D. Piga, A. Bemporad, Kalman filtering for energy disaggregation, IFAC-PapersOnLine 51 (5) (2018) 108–113. 1st IFAC Workshop on Integrated Assessment Modelling for Environmental Systems IAMES 2018..

[82] M. Aiad, P.H. Lee, Energy disaggregation of overlapping home appliances consumptions using a cluster splitting approach, Sustainable Cities and Society 43 (2018) 487–494.

[83] A. Miyasawa, Y. Fujimoto, Y. Hayashi, Energy disaggregation based on smart metering data via semi-binary nonnegative matrix factorization, Energy and Buildings 183 (2019) 547–558.

[84] Y. Himeur, A. Elsalemi, F. Bensaali, A. Amira, Efficient multi-descriptor fusion for non-intrusive appliance recognition, in: The IEEE International Symposium on Circuits and Systems (ISCAS), 2020, pp. 1–5.

[85] Y. Liu, X. Wang, L. Zhao, Y. Liu, Admittance-based load signature construction for non-intrusive appliance load monitoring, Energy and Buildings 171 (2018) 209–219.

[86] A.L. Wang, B.X. Chen, C.G. Wang, D. Hua, Non-intrusive load monitoring algorithm based on features of V-I trajectory, Electric Power Systems Research 157 (2018) 134–144.

[87] C. Liu, A. Akintayo, Z. Jiang, G.P. Henze, S. Sarkar, Multivariate exploration of non-intrusive load monitoring via spatiotemporal pattern network, Applied Energy 211 (2018) 1106–1122.

[88] S. Henriet, U. Simsekli, B. Fuentes, Richard G.A generative model for non-Intrusive load monitoring in commercial buildings, Energy and Buildings 177 (2018) 268–278.

[89] S.S. Hosseini, K. Agbossou, S. Kelouwani, A. Cardenas, Non-intrusive load monitoring through home energy management systems: A comprehensive review, Renewable and Sustainable Energy Reviews 79 (2017) 1266–1274.

[90] Z.X. Wang, L.Y. He, H.H. Zheng, Forecasting the residential solar energy consumption of the United States, Energy 178 (2019) 610–623.

[91] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, P. Chatellier, Modeling and forecasting building energy consumption: A review of data-driven techniques, Sustainable Cities and Societ 48 (2019) 101533..

[92] T. Ahmad, H. Chen, Deep learning for multi-scale smart energy forecasting, Energy 175 (2019) 98–112.

[93] T. Hong, P. Pinson, Energy forecasting in the big data world, International Journal of Forecasting (2019).

[94] G.P. Herrea, M. Constantino, B.M. Tabak, H. Pistori, J.J. Su, A. Naranpanawa, Data on forecasting energy prices using machine learning, Data in Brief 104122 (2019).

[95] J. Li, R.E. Just, Modeling household energy consumption and adoption of energy efficient technology, Energy Economics 72 (2018) 404–415.

[96] M. Villca-Pozo, J.P. Gonzales-Bustos, Tax incentives to modernize the energy efficiency of the housing in Spain, Energy Policy 128 (2019) 530–538.

[97] L.M. Lopez-Ochoa, J. Las-Heras-Casas, L.M. Lopez-Gonzalez, P. Olasolo-Alonso, Towards nearly zero-energy buildings in Mediterranean countries: energy performance of buildings directive evolution and the energy rehabilitation challenge in the Spanish residential sector, Energy 176 (2019) 335–352.

[98] A. Thonipara, P. Runst, C. Ochsner, K. Bizer, Energy efficiency of residential buildings in the European Union – An exploratory analysis of cross-country consumption patterns, Energy Policy 129 (2019) 1156–1167.

[99] T. Hong, P. Pinson, S. Fan, Global energy forecasting competition 2012, International Journal of Forecasting 30 (2) (2014) 357–363.

[100] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, R.J. Hyndman, Probabilistic energy forecasting: global energy forecasting competition 2014 and beyond, International Journal of Forecasting 32 (3) (2016) 896–913.

[101] W. Cui, H. Wang, Anomaly detection and visualization of school electricity consumption data, in: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), 2017, pp. 606–611.

[102] J.E. Seem, Using intelligent data analysis to detect abnormal energy consumption in buildings, Energy and Buildings 39 (1) (2007) 52–58.

[103] I. Khan, A. Capozzoli, S.P. Corgnati, T. Cerquitelli, Fault detection analysis of building energy consumption using data mining techniques, Energy Procedia 42 (2013) 557–566. Mediterranean Green Energy Forum 2013: Proceedings of an International Conference MGEF-13..

[104] H. Janetzko, F. Stoffel, S. Mittelstadt, D.A. Keim, Anomaly detection for visual analytics of power consumption data, Computers & Graphics 38 (2014) 27–37.

[105] Z. Ma, J. Song, J. Zhang, A real-time detection method of abnormal building energy consumption data coupled POD-LSE and FCD, Procedia Engineering 205 (2017) 1657–1664. 10th International Symposium on Heating, Ventilation and Air Conditioning, ISHVAC2017, 19–22 October 2017, Jinan, China..

[106] D.B. Araya, K. Grolinger, H.F. ElYamany, M.A.M. Capretz, G. Bitsuamlak, An ensemble learning framework for anomaly detection in building energy consumption, Energy and Buildings 144 (2017) 191–206.

[107] C. Nordahl, M. Persson, H. Grahn, Detection of residents' abnormal behaviour by analysing energy consumption of individual households, in: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), 2017, pp. 729–738.

[108] H. Qiu, Y. Tu, Y. Zhang, Anomaly detection for power consumption patterns in electricity early warning system, in: 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI), 2018, pp. 867–873.

[109] Y. Weng, N. Zhang, C. Xia, Multi-agent-based unsupervised detection of energy consumption anomalies on smart campus, IEEE Access 7 (2019) 2169–2178.

[110] T. Picon, M.N. Meziane, P. Ravier, G. Lamarque, C. Novello, J.L. Bunetel, et al., COOLL: controlled on/off loads library, a public dataset of high-sampled electrical signals for appliance identification, CoRR (2016), abs/1611.05803.

[111] C. Shin, E. Lee, J. Han, J. Yim, W. Rhee, H. Lee, The ENERTALK dataset, 15 Hz electricity consumption data from 22 houses in Korea, Scientific Data 12 (2019) 6.

[112] Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks, Cognitive Computation (2020) 1–23.

[113] E. Fotopoulou, A. Zafeiropoulos, F. Terroso, A. Gonzalez, A. Skarmeta, U. Simsek, et al., Data aggregation, fusion and recommendations for strengthening citizens energy-aware behavioural profiles, 2017 Global Internet of Things Summit (GIoTS) (2017) 1–6.

[114] Y. Himeur, A. Alsalemi, A. Al-Kababji, F. Bensaali, A. Amira, Data fusion strategies for energy efficiency in buildings: Overview, challenges and novel orientations, Information Fusion (2020) 1–36.

[115] A. Capozzoli, M.S. Piscitelli, A. Gorrino, I. Ballarini, V. Corrado, Data analytics for occupancy pattern learning to reduce the energy consumption of HVAC systems in office buildings, Sustainable Cities and Society 35 (2017) 191–208.

[116] H. Kang, M. Lee, T. Hong, J.K. Choi, Determining the optimal occupancy density for reducing the energy consumption of public office buildings: A statistical approach, Building and Environment 127 (2018) 173–186.

[117] E. Delzendeh, S. Wu, A. Lee, Y. Zhou, The impact of occupants' behaviours on building energy analysis: A research review, Renewable and Sustainable Energy Reviews 80 (2017) 1061–1071.

[118] S. Ge, J. Li, H. Liu, X. Liu, Y. Wang, H. Zhou, Domestic energy consumption modeling per physical characteristics and behavioral factors, Energy Procedia 158 (2019) 2512–2517. Innovative Solutions for Energy Transitions..

[119] Y. Ding, X. Ma, S. Wei, W. Chen, A prediction model coupling occupant lighting and shading behaviors in private offices, Energy and Buildings 216 (2020) 109939..

[120] X. Jiang, L. Wu, A residential load scheduling based on cost efficiency and consumer's preference for demand response in smart grid, Electric Power Systems Research 186 (2020) 106410..

[121] S. Lefkeli, E. Manolas, K. Ioannou, G. Tsantopoulos, Socio-cultural impact of energy saving: studying the behaviour of elementary school students in Greece, Sustainability 10 (3) (2018).

[122] J. Koci, V. Koci, J. Madera, R. Cerny, Effect of applied weather data sets in simulation of building energy demands: Comparison of design years with recent weather data, Renewable and Sustainable Energy Reviews 100 (2019) 22–32.

[123] Y. Liu, R. Stouffs, A. Tablada, N.H. Wong, J. Zhang, Comparing micro-scale weather data to building energy consumption in Singapore, Energy and Buildings 152 (2017) 776–791.

[124] Y. Geng, W. Ji, B. Lin, J. Hong, Y. Zhu, Building energy performance diagnosis using energy bills and weather data, Energy and Buildings 172 (2018) 181–191.

[125] S. Farah, D. Whaley, W. Saman, J. Boland, Integrating climate change into meteorological weather data for building energy simulation, Energy and Buildings 183 (2019) 749–760.

[126] G. Lupato, M. Manzan, Italian TRYs: New weather data impact on building energy simulations, Energy and Buildings 185 (2019) 287–303.

[127] S. Erba, F. Causone, R. Armani, The effect of weather datasets on building energy simulation outputs, Energy Procedia 134 (2017) 545–554. Sustainability in Energy and Buildings 2017: Proceedings of the Ninth KES International Conference, Chania, Greece, 5–7 July 2017..

[128] R. Antonietti, F. Fontini, Does energy price affect energy efficiency? Cross-country panel evidence, Energy Policy 129 (2019) 896–906.

[129] A. Satchwell, P. Cappers, C. Goldman, Customer bill impacts of energy efficiency and net-metered photovoltaic system investments, Utilities Policy 50 (2018) 144–152.

[130] D. Eryilmaz, S. Gafford, Can a daily electricity bill unlock energy efficiency? Evidence from Texas, The Electricity Journal 31 (3) (2018) 7–11.

[131] M.G. Fikru, Electricity bill savings and the role of energy efficiency improvements: A case study of residential solar adopters in the USA, Renewable and Sustainable Energy Reviews 106 (2019) 124–132.

[132] N. Hossein Motlagh, M. Mohammadrezaei, J. Hunt, B. Zakeri, Internet of Things (IoT) and the Energy Sector, Energies 13 (2) (2020).

[133] Raspberry Pi 4 Model B. Accessed: 2020-05-04. URL:https://www.raspberrypi.org/products/raspberry-pi-4-model-b/..

[134] ODROID-XU4. Accessed: 2020-05-04. URL:https://www.hardkernel.com/shop/odroid-xu4-special-price/..

[135] Jetson TX1 Developer kit. Accessed: 2020-05-04. URL:http://www.nvidia.com/object/jetson-TX1-dev-kit.htmll..

[136] A. Alsalemi, Y. Himeur, F. Bensaali, A. Amira, C. Sardianos, C. Chronis, et al., A micro-moment system for domestic energy efficiency analysis, IEEE Systems Journal (2020) 1–8.

[137] C. Sardianos, I. Varlamis, G. Dimitrakopoulos, D. Anagnostopoulos, A. Alsalemi, F. Bensaali, et al., REHAB-C: recommendations for energy HABits change, Future Generation Computer Systems 112 (2020) 394–407.

[138] F. Ricci, L. Rokach, B. Shapira, P.B. Kantor, Recommender Systems Handbook, first ed., Springer-Verlag, Berlin, Heidelberg, 2010.

[139] A. Starke, M. Willemsen, C. Snijders, Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system, in: Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17, ACM, New York, NY, USA, 2017, pp. 65–73.

[140] Y. Zhang, X. Chen, et al., Explainable recommendation: A survey and new perspectives. Foundations and Trends®, Information Retrieval 14 (1) (2020) 1–101.

[141] Neurio: Intelligently managing the home's energy. Accessed: 2019-07-04. https://www.neur.io/..

[142] S.T. Moghadam, S. Coccolo, G. Mutani, P. Lombardi, J.L. Scartezzini, D. Mauree, A new clustering and visualization method to evaluate urban heat energy planning scenarios, Cities 88 (2019) 19–36..

[143] M.R. Herrmann, D.P. Brumby, T. Oreszczyn, X.M.P. Gilbert, Does data visualization affect users' understanding of electricity consumption?, Building Research & Information 46 (3) (2018) 238–250

[144] A. Spence, M. Goulden, C. Leygue, N. Banks, B. Bedwell, M. Jewell, et al., Digital energy visualizations in the workplace: the e-Genie tool, Building Research & Information 46 (3) (2018) 272–283.

[145] G. Li, C. Kou, H. Wang, Estimating city-level energy consumption of residential buildings: A life-cycle dynamic simulation model, Journal of Environmental Management 240 (2019) 451–462.

[146] M. Zekic-Susac, S. Mitrovic, A. Has, Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities, International Journal of Information Management 102074 (2020).

[147] M. Fekri, A.M. Ghosh, K. Grolinger, Generating energy data for machine learning with recurrent generative adversarial networks, Energies 12 (2019) 13.