Towards data-driven energy communities: a review of open-source datasets, models and tools

Hussain Kazmi[a,*], Íngrid Munné-Collado[b], Fahad Mehmood[c], Tahir Abbas Syed[d], Johan Driesen[a]

[a] *Dept. of electrical engineering, KU Leuven, Belgium*
[b] *Universitat Politècnica de Catalunya (UPC-Barcelona Tech), Spain*
[c] *Dubai Business School, University of Dubai, UAE*
[d] *Alliance Manchester Business School, The University of Manchester, UK*

Abstract

Energy communities will play a central role in the sustainable energy transition by helping inform and engage end users to become more responsible consumers of energy. However, the true potential of energy communities can only be unlocked at scale. This scalability requires data-driven solutions that model not just the behaviour of building occupants but also of energy flexible resources in buildings, distributed generation and grid conditions in general. This understanding can then be utilized to improve the design and operation of energy communities in a variety of real-world settings. However, in practice, collecting and analysing the data necessary to realize these objectives forms a large part of such projects, and is often seen as a prohibitive stumbling block. Furthermore, without a proper understanding of the local context, these projects are often at risk of failure due to misplaced expectations. However, this process can be considerably accelerated by utilizing open source datasets and models from related projects, which have been carried out in the past. Likewise, a number of open source, general-purpose tools exist that can help practitioners design and operate LECs in a near-optimal manner. These resources are important because they not only help ground expectations, they also provide LECs and other relevant stakeholders, including utilities and distribution system operators, with much-needed visibility on future energy and cash flows. This review provides a detailed overview of these open-source datasets, models and tools, and the many ways they can be utilized in optimally designing and operating real-world energy communities. It also highlights some of the most important limitations in currently available open source resources, and points to future research directions.

Highlights:

1. The importance of open-source datasets and tools for local energy communities
2. Common use cases for open-source datasets, models and tools for energy communities
3. A thorough review of electricity demand and meteorological datasets and models
4. Most important shortcomings with currently available datasets, models and tools

Keywords: energy communities, open-source, forecasting, optimal control, data-driven analysis

Word count: 9000 words

1. Introduction

Climate change concerns and accelerating digitization are causing arguably the largest shift in energy systems in the last decades [1]. Many countries are facilitating, or plan to facilitate, an increasing proliferation of renewable energy sources (such as wind and solar) by making energy demand more flexible. Digitization is a vital enabling technology in this transition. It allows system operators to better plan and operate the energy grid in light of both demand variations, and supply variability and intermittency introduced by renewable energy sources (RES). Concurrently, bottom-up initiatives, such as energy communities, have emerged as potential solutions to further accelerate the uptake of renewable energy technologies at a local level, while emphasising self-sufficiency, local determination, and engagement and empowerment of energy consumers and prosumers [2].

1.1. Local energy Communities (LECs)

Over the years, many different formulations of energy communities have emerged, including local energy communities, citizen energy communities, market energy communities and renewable energy communities. While a number of definitions have been proposed [3], there remains a tremendous amount of diversity in energy communities in practice. For instance, many energy communities are characterized by a strong emphasis on citizen engagement [4] which allows energy consumers and prosumers to become more involved with how their energy is sourced (or consumed). Often, such energy community projects emphasize innovation in renewable technologies [5] and decentralized ownership through local stakeholders' involvement [6]. These are especially useful in the context of reducing the barriers placed by high investment costs in many community-scale energy projects. In these contexts, there is also a strong element of sharing financial and social benefits for the local community [7].

It must be emphasized here that the *energy community* concept is much more expansive than *community energy* [8]. While community energy is often seen as an instrument to reduce the barriers of entry to local generation (e.g. through renewable sources such as solar PV or wind power etc.), energy communities go much further [9]. For instance, energy communities are characterized by their strong focus on public-private partnerships to further develop local competences. Such energy communities also emphasize the prosumer role, whereby energy consumers in the community are encouraged to take on a more active role by, for instance, installing distributed generation and energy flexible resources such as battery storage. The presence of metering infrastructure is also an important component in modern energy communities, as it allows real-time monitoring of energy flows.

A number of North-Western European countries (such as Sweden, Denmark and the Netherlands) have been forerunners in the development of energy communities emphasizing renewable energy. These efforts have been largely driven by conducive regulatory conditions

*Corresponding author details, hussainsyed.kazmi@kuleuven.be

leading to citizen-led decentralized renewable energy projects. For instance, the Tvindkraft Project[1] developed in Denmark in 1978 is an example of an early local energy community. On the EU level, the recent Recast Renewable Energy Directive (RED-II) builds on the existing regulatory framework for climate and energy to further stimulate the development of renewable energy communities (RECs) in Member States [10]. As of April 2020, the European Federation of Citizen Energy includes a network of 1500 European energy cooperatives and about one million citizens active in the energy transition[2]. Some important energy community projects in Europe are detailed in [11].

Historically, energy communities have also been used as an effective means of providing affordable access to energy to people in rural and low population density regions, especially in the USA. As a result, these co-operatives are responsible for 11% of the total energy sold, and manage around 13% of the overall connections in the USA [9]. Over the past few years, the same ideas have also been extended to electrify off-grid communities in developing countries. The unavailability or unreliability of the national grid means that the self-sufficiency aspect of energy communities fits in very well with the issues faced by people in such communities. For example, use of community solar and micro-hydro plants has become widespread in South and South-East Asia [12].

It is important to point out that the RECs highlighted above are meant to foster community engagement and incentivise local renewable energy projects, and are separate from Citizen Energy Communities (CECs), as postulated by another recent European directive. Unlike RECs, CECs are meant to be technology-neutral and not renewable-specific. Furthermore, they are not necessarily geographically constrained, and allow large corporations to participate in the energy community project, albeit with some restrictions [9]. This reflects the broader picture where, over time, energy communities have come to encompass not just residential communities, but also other energy consumers in neighborhoods, ranging from tertiary buildings and local industry to public electric vehicle charging infrastructure.

In practical terms, energy communities have a long track record of promoting renewable energy and managing energy consumption, while empowering end users and increasing their energy awareness. Some of the most interesting practical use cases are discussed in subsequent sections. Going further, we will use local energy communities (LECs) as an umbrella term to refer to the overarching concepts of energy communities in general, and RECs in particular. We use LECs to denote that our focus in this paper is primarily on the promotion of renewable generation and the prosumer mentality in such energy communities. However, the definition we employ is broader than the EU REC formulation in the sense that we do not impose any limitations on who can be part of the energy community, thereby incorporating some elements of the CECs.

1.2. Digitization

As mentioned above, a key enabling technology to operationalize energy communities in practice is by increasing digitization. The most critical component of this is the advanced

[1] https://www.tvindkraft.dk/en/
[2] https://www.rescoop.eu/

metering infrastructure, which provides a way to record energy demand and generation values in (near-)real time, thereby paving the way for monitoring, user engagement, peer to peer trading concepts and downstream optimization [13]. However, accessing this data as well as processing it in a way that creates additional value can often be prohibitively expensive for bottom-up citizen initiatives. This is an especially hard problem, as the value of data for downstream services only becomes visible once it has been collected and analysed. The large upfront costs to collect and process this data can therefore dampen the enthusiasm of such community initiatives.

Pre-existing open sourced data and software projects offer an attractive solution to this problem. Such datasets, when collected previously in similar settings, can unlock a number of use cases. For instance, they can be used in an ex ante analysis to estimate the economic feasibility of setting up a new energy community. This open sourcing of datasets is, in many cases, mandated by project funding bodies, which may require or request that data collected using public money should be made available for future scientific research. This includes the emphasis placed on open access and data management in Horizon 2020 projects by the European Commission [14]. Besides compliance with grant agreements, open sourcing datasets can also benefit companies, research organizations and society at large by helping advance scientific state of the art and enabling meta-analysis.

Once relevant data has been acquired (either within the LEC, or using a historically similar project), a number of open-source software projects exist already to analyze it, which can quickly help LECs estimate the additional value that can be created with the data. The motivation for software developers participating in developing these open source projects is often multi-fold [15]. On the one hand, there are intrinsic rewards for participation, such as altruism and personal fulfilment. On the other, extrinsic rewards such as expected future returns can also play a large role in developer motivation. More recently, a number of large companies, including Google, Facebook and Uber, have started open sourcing their internal tools (often focused on analysing data at different stages of a project's life), with the aim to accelerate scientific development and stimulate community participation in software systems development [16].

1.3. Challenges

Even though open source tools and datasets can theoretically be used to estimate or even improve the economic viability of LECs, a number of challenges remain. For instance, the energy sector has lagged behind in both publicly available datasets of energy demand and production, and specialized open source tools to analyze them [17]. Some of the reasons for this delay include large institutional inertia, privacy concerns and the slow pace of digitization. Legislation requiring metering of energy data has also hit roadblocks in many countries, with even European Union member states showing a fractured landscape. Furthermore, the disparate nature of many different energy vectors (including electricity, gas, heat etc.) complicate efforts to measure demand and supply in a holistic manner. Finally, the data being recorded often leads to loss of privacy and can leave customers vulnerable to security exploits - which means data availability and privacy concerns go hand in hand [18].

Beyond these challenges, a number of operational issues also persist before LECs can achieve mainstream adoption. Foremost amongst these is the fact that LECs rely on either direct user engagement or automation to elicit energy flexibility that can be used to improve energy usage or grid reliability in some way. However, in practice, both of these instruments can be rather limited in scope. For instance, automation-driven flexibility requires potentially expensive instrumentation to enable remote control, and is then limited only to the few devices that have been instrumented in this way. Similar issues persist with direct user engagement: providing feedback to users can lead to information fatigue, especially when the financial rewards are low. Further, convincing users to install more energy efficient appliances can have a rebound effect on usage, implying that gains in efficiency do not always translate to an equivalent reduction in emissions [19].

Therefore, in the business as usual scenario, LECs face a number of operational hurdles, such as limited flexibility in the case of automation, and user indifference in the case of engagement. Even in the best case scenario when sufficient flexibility is available, the lack of availability of data and tools necessary to analyze the collected data remain stumbling blocks. Indeed, the lack of real data means that planners often have to work in the dark while planning and operating their LECs.

1.4. Contributions

This paper presents a detailed overview of the open source tools, datasets and models that can be used to operationalize LECs in practice. By compiling state of the art, open-source datasets, tools and models in one place, this paper will form a quick reference guide of available resources, and help identify any potential shortcomings. More concretely, in this paper, we focus on LECs from the perspective of electrical energy demand and generation using solar PV (although the same data can be extended to work with solar thermal systems as well). We also explore available ambient conditions data and provide a brief overview of thermal and electrical storage. This information is relevant not just to prospective energy communities, but also practitioners and system operators that might help with the organization and setting up of such communities. Furthermore, electricity utilities and aggregators of distributed energy resources (DERs), in addition to the wider community of architects, building engineers and policy makers, may find it useful as well. A more detailed overview of these use cases in presented in the next section. Finally, even though the paper cannot directly influence the amount of energy flexibility available in an LEC, it provides the tools and data necessary to estimate and leverage the available flexibility in the most optimal manner.

The remainder of the paper is organized as follows: in section 2, we highlight key use cases which openly available data can enable for different stakeholders in a LEC. Section 3 discusses the terminology employed in this paper. Section 4 provides a compilation of open-source datasets that can be leveraged in LECs, including demand side data in the built environment, generation data, as well as relevant markets and weather data. It also identifies some key shortcomings in existing datasets. Section 5 discusses existing models that have been built using either human domain expertise or using collected data that could not be made available due to privacy concerns. These models can, in turn, be used to generate data that can then be used for further analysis in real world LECs. Related to these software models, section 6 provides a review of some general purpose open source software projects that can be utilized to gain insights from energy-related data in LECs. Section 7 provides a conclusion, as well as some future directions for research in and operationalization of LECs.

2. Data use cases in LECs

Data availability is critical for the successful operationalization of LECs, both during the design and the operation phase. During the design phase, it is necessary to estimate the optimal dimensions of local energy flexible resources (EFRs) such as battery-inverter systems and distributed or community generation. On the other hand, in the operation phase, it is necessary to manage any local flexibility in a way that helps achieve community goals such as maximal energy efficiency, self consumption etc. Figure 1 presents an overview of these use cases.

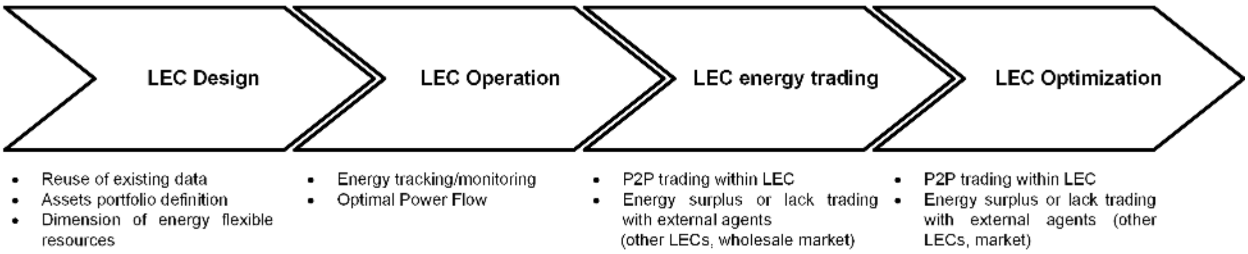

Figure 1: LEC data use cases pipeline

These use cases will become increasingly important as the value addition of energy communities shifts from accumulating capital to overcome investment barriers (i.e. community energy projects) to unlocking diverse streams of revenue with the installed assets by operating them in a flexible manner. More concretely, a number of researchers and practitioners have attempted to formalize such use cases into a number of business models depending on the entity under consideration. For instance, [20] identify a number of different energy community archetypes and the major stakeholders in them. These archetypes include cooperative investment, energy sharing platform, aggregator and microgrid. The availability of data and existing models which can generate this data for the energy community's geographical location can benefit all of these archetypes, as we explain next.

For the cooperative investment archetype, having access to the approximate energy demand requirements, generation potential and climate conditions in the LEC's operational area can help inform the cooperative about its investment decisions. These insights can be used to optimize the dimensions of storage and generation systems in the LEC. The local context (e.g. renewable generation potential and electrification rates) considerably alters the optimal resource mix of a LEC. For instance, LECs in North-Western European countries frequently need to over-dimension storage to bridge the diurnal supply-demand gap, and also consider long-term storage (e.g. in aquifers) to bridge the seasonal supply-demand gap.

For regions with a smaller temporal mismatch between supply and demand, this is often not required. Local tariffs and carbon intensity also affect the profitability and desirability of these investments. The importance of tariff structure and local resource mix means that the local utilities and distribution system operators (DSOs) are also often involved with the design of LECs. In practice, using pre-existing datasets (described in section 4) and models (described in section 5) can help all the relevant stakeholders in achieving their objectives at a lower price point by making more informed decisions: the cooperative gets to make cost-optimal investments, while the DSO can incentivize LEC designs which help stable grid operation in the future and avoid the need for costly grid reinforcement investments.

Likewise, for the energy sharing archetype, having access to this data as well as possible future forecast values can facilitate market players (such as balance responsible parties and DSOs) in optimizing energy flows for maximal self-consumption in energy sharing settings. Modern Peer-to-Peer (P2P) energy trading concepts enable the energy needs of a LEC to be met by internal trade among local prosumers and consumers. This can increase the operational efficiency of a LEC, and avoid issues with large-scale grid injection. On top of that, there are other approaches such as Peer-to-platform [23], or community-based [58], where a common agent or intermediary trades the surplus or lack of energy in an external market, either between LECs or in a wholesale market, when possible. The LEC operator, either prosumers in the LEC itself or a contracted third party, can help setup and operate this platform, making use of tools highlighted in section 6 as well as models presented in section 5.

Finally, for both the aggregator and microgrid archetypes, the value open-sourced models and data bring is both in the LEC design and real-time control they can enable. Real-time control of an optimally designed LEC can be used to provide ancillary services to the broader grid, realize cost arbitrage on the electricity markets, and maximize self-sufficiency etc. Doing so can reduce the costs of energy for energy community members, but also improve the community's resiliency to grid outages. For instance, openly available models for one or more of the demand side, price signals and the local supply can be used as inputs to an optimization algorithm controlling the energy flows in a LEC. In addition to automated control, it is also possible to use these types of models to reduce energy demand by user engagement, for instance by providing comparisons to neighbors and other similar households [21]. It also enables DSOs to invest in LECs rather than costly grid reinforcements, necessitated by distributed generation and elevated energy demand. Consequently, depending on the specific context, different stakeholders in these two archetypes will find all three resources highlihgted in this paper (datasets, models and tools) valuable.

3. The distinction between datasets, models and tools

Before proceeding further, it is important to make the distinction between tools, datasets and models explicit. Figure 3 provides a simplified overview of this distinction.

1. A **dataset** refers to observation data, either in raw or cleaned form, that has been made available for further analysis. This is often in the form of a time series, and

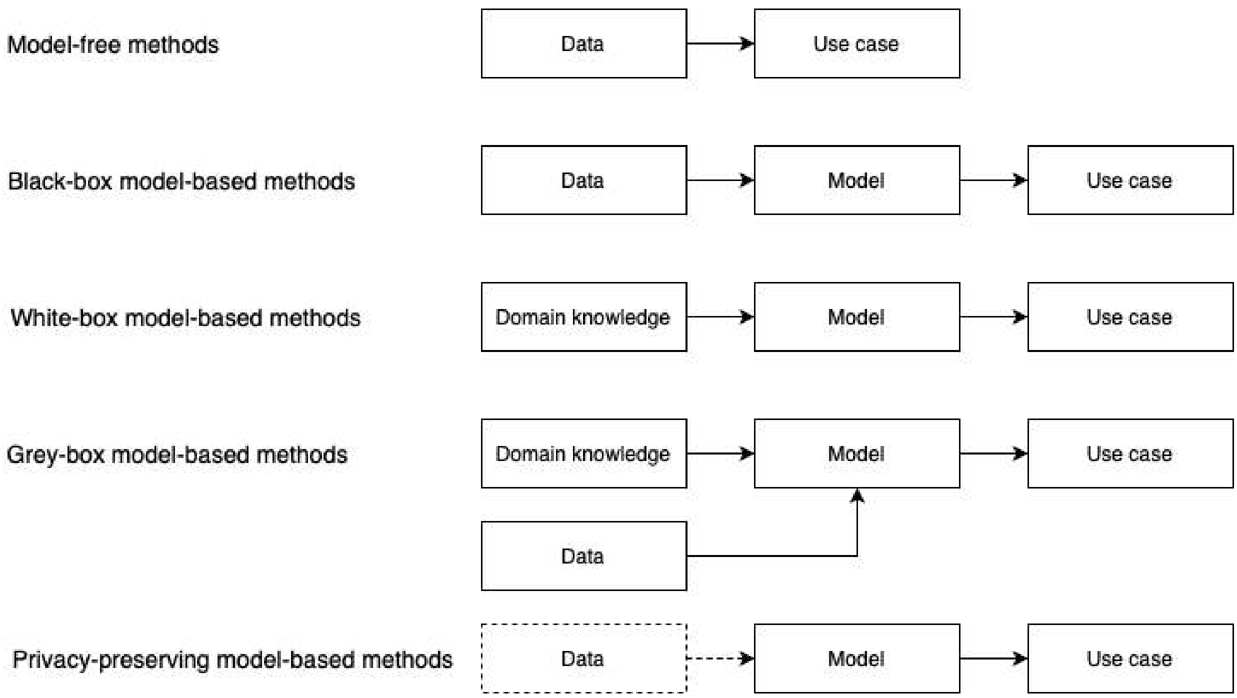

Figure 2: An overview of the different ways data and models can be used in practice; (a) model-free methods directly utilize data to optimize some metric of interest e.g. model-free control of energy flexible resources; (b, c, and d) model-based methods depending on data, domain knowledge or both to construct a model that is used for subsequent analysis; (e) privacy-preserving use cases where only a model trained on user data is utilized. The general purpose tools highlighted in section 6 can be used to build models using gathered data or domain knowledge as well as enabling a host of other services.

can be complemented by additional meta-data. An example of a dataset can be the electricity demand of a building as a time series, while the metadata can provide information about its geographical location and building occupant demographics etc.

2. A **model**, on the other hand, distills the information from a dataset, which may or may not be made available in addition to the model, into a directly usable or interpretable form. The model is therefore more restrictive than the broader dataset, and can, for instance, be used to preserve user privacy. However, it is also possible to develop models from first principles, rather than using observation data. Such models are a valuable source of information in cases where real world data is not available or collecting it can be expensive. These models are also useful for cases where the physics of the process is well understood. Some examples include models for storage systems such as battery-inverter systems and thermal systems, as well as generation models for renewable energy technologies such as solar PV.
3. The **tools** considered in this paper refer to existing, open-source software projects that can be used, either directly or indirectly, to provide services in the energy sector. As this category features some overlap with the models category, we use tools to refer to general purpose software projects that have been, or can be, utilized in LECs as well.

4. Datasets

This section summarizes a number of publicly available datasets relevant to LECs. In addition to datasets for demand and supply, other exogenous factors that may influence the functioning of an energy community, such as ambient conditions as well as electricity prices and carbon intensity are presented as well. Where available, data from developing countries is highlighted as this data is much more sparse, and electrification in these countries represents the bulk of future demand growth.

4.1. Demand side data

The energy demand in a LEC drives most downstream business cases. Therefore, getting greater visibility on the demand side is arguably the most important advantage of openly available datasets for practitioners in the field. The datasets presented next can be distinguished based on a number of different factors, including:

1. The scale of the building constituting the LEC, i.e. whether considering a residential or commercial building.
2. The temporal and spatial resolution of collected data. More specifically, a number of recent studies have focused on disaggregation of electrical loads from the observed load profile - also referred to as non-intrusive load monitoring (NILM). The key factor differentiating these datasets is their high sampling frequency and sub-metering of individual loads.

The remainder of this section is organized as follows. We begin by providing a review of residential energy demand, both on an aggregated and disaggregated level. We follow this by discussing commercial building loads and sub-metered demand in various contexts that could also be considered useful for future LEC projects.

4.1.1. Residential building energy demand

Researchers and practitioners have made available a number of datasets for residential energy demand. These datasets differ not just based on the geography of the buildings, but also in the amount of buildings they provide data for. Furthermore, some datasets only provide electrical consumption on either device or building level, while others also provide metadata or other sensor data such as water and gas consumption. As mentioned before, the temporal resolution of sensing can also be quite different. We begin with datasets that are sampled at a high enough temporal resolution (i.e. 1 Hz or faster) to enable non-intrusive load monitoring:

1. The EMBED dataset [22] includes labeled electricity disaggregation dataset containing plug load consumption data for different appliances in 3 US apartment units in California. The plug load data, collected for 14-27 days, has a sampling frequency of 1-2 Hz (the sampling frequency of current and voltage is much higher at 12 kHz).

2. The REDD dataset (Reference Energy Disaggregation Dataset) [23], likewise, consists of electricity consumption measurements on the building and circuit level for 6 US housing units over 3-19 days. At a high level, REDD provides high frequency information sampled at 0.5 - 1 Hz for up to 20 plug-level monitors, and 24 individual circuits along with a label for the category of appliance connected to it. The power and voltage information are recorded at a much higher sampling frequency of 15kHz.
3. The BLUED energy disaggregation dataset [24] includes current and voltage measurements from one US single-family residence, sampled at 12 kHz for 8 days. Furthermore, the dataset includes time-stamped and labeled state transitions of each appliance during the study period.
4. The PLAID (Plug-Level Appliance Identification Dataset) dataset [25] is another public dataset of electrical appliance measurements for NILM research, and covers 11 different appliance types from 56 households in Pittsburgh, Pennsylvania, USA. It provides voltage and current data, collected during the summer of 2013 and winter of 2014, at a sampling frequency of 30 kHz.
5. The ADRES concept [26] dataset provides data on power consumption and voltage profiles from 30 Austrian households at a sampling frequency of 1 Hz. The measurements are available for a total of two weeks, one in winter 2009 and another in summer 2010. However, the dataset does not include detailed appliance level data.
6. UK-DALE [27] is a domestic appliance level electricity dataset from five households in the UK. The dataset contains both the building and appliance level power demand with a sampling resolution of six seconds. In three of the five houses (houses 1, 2 and 5) the whole-house voltage and current is also recorded at 16 kHz. Data from household 1 is now available for well over 4 years and can facilitate long term analysis of seasonalities and trends.
7. The DRED (Dutch residential energy dataset) [28] dataset provides appliance and house level energy consumption (sampled at 1Hz) data for a Dutch building for 6 months. Furthermore, minute-level ambient conditions measurements (indoor and outdoor temperature, environmental mapping etc.), and metadata about the building (occupancy information, building layout, application mapping etc.) are also included.
8. The ECODS dataset [29] provides measurement data for 6 households over an observation period of 8 months between June 2012 and January 2013 at 1 Hz. This dataset contains both aggregate energy demand, as well as sub-metered measurements for appliances such as refrigerator, freezer, television and coffee machines etc.
9. The ENERTALK dataset [30] provides measurements sampled at 15 Hz for active and reactive power drawn in 22 houses in South Korea, on both the appliance and aggregate building level. The measurements range from 29 days to 122 days, depending on the building, while the monitored appliances included refrigerator, kimchi refrigerator, rice cooker, washing machine, and television.

In addition to the datasets providing high temporal resolution, a number of other datasets also exist. These typically cover a greater number of buildings spanning a longer duration than a few days or weeks, due to the relatively fewer data points.

1. PecanStreet Inc. Dataport [31] is a large scale dataset which contains appliance-level electricity demand from around 1400 houses and apartments located in three areas of the USA (New York, Texas and California) for multiple years, at various granularity levels (1s, 1min, 15min). At the current time, a subset of 25 households of the entire dataset is available for each location and granularity, up to 6 months of data or 1 year of data. This data is however not fully open sourced.
2. The REFIT dataset [32] provides power measurements for 20 households in the UK at both an aggregate and an appliance level. The data have a sampling period of 8 seconds and is available for a period of two years.
3. Smart* data set [33] is another large scale dataset which makes available minute-level electricity usage data from over 400 anonymized houses. Furthermore, this data set also includes electricity consumption and generation, weather conditions and HEMS operational data and meta-data on three households. The same source contains data for 114 single-family apartments in a time granularity of 1 minute and a total of 2 years of records (from 2014 to 2016), both for the aggregated electricity consumption and the weather conditions on these time periods. The closely related NIOM dataset combines electricity consumption with occupancy patterns in the building for 3 weeks of minute level data on consumption and occupancy [34].
4. AMPDs [35] contains minute-level electricity, water, and natural gas measurements for two years for a residential building in Canada. The dataset features 21 power meters, as well as water and natural gas meters. Ambient conditions are also included in the dataset.
5. The PRECON dataset [36] provides minute-level electricity demand data from 42 houses in Lahore, Pakistan for a period of one year. Additional meta-data such as demographics information and device-level sub-metering is also included in the dataset.
6. The CoSSMic (Collaborating Smart Solar-powered Microgrids) dataset[3] [37] contains sub-metered energy demand for 11 households in Konstanz, Germany. The energy demand is sampled at 1 minute intervals, and is available between October 2013 to December 2016.
7. The SustData dataset [38] provides measurements for 50 residential units in Portugal over a period of 1144 days. The sampling rate of the dataset is between 2 and 10 Hz, and sub-metered demand and eco-feedback information from building occupants is also provided in addition to the aggregate energy demand.

The information in this section is summarized in Table 1.

4.1.2. Commercial buildings

Increasingly, commercial - or tertiary - buildings can also form part of an energy community, either on their own or as part of a broader community. This section includes an overview of some public datasets detailing electricity demand in commercial buildings.

[3]https://data.open-power-system-data.org/household$_data/2020-04-15$

Table 1: An overview of open-source residential buildings data

Residential buildings					
Dataset name	Country	Sites	Duration	Resolution	Ref
EMBED	US	3	2-4 weeks	12 kHz (I, V); 1-2 Hz (plug loads)	[22]
REDD	US	6	2-4 weeks	15 kHz (P, V); 0.5-1 Hz (NILM data at plug/circuit level))	[23]
BLUED	US	1	1 week	12 kHz (I, V)	[24]
PLAID	US	56	Summer 2013 and winter 2014	30 kHz (I, V)	[25]
ADRES	Austria	30	2 weeks	1 Hz	[26]
REFIT	UK	20	2 years	0.125 Hz	[32]
UK-DALE	UK	5	4 years	16 kHz (I, V in 3 buildings); 0.17 Hz (appliance-level demand)	[27]
DRED	The Netherlands	1	6 months	1 Hz (energy demand); 1 minute (ambient conditions)	[28]
Dataport	US	1400+ 75 (free-license)	4 years 6 months	1 Hz, 1 minute, 15 minutes	[31]
Smart*	US	3 114-400	3 weeks 1-4 years	1 Hz	[33]
AMPds	Canada	1	2 years	1 minute	[35]
ECODS	Switzerland	6	8 months	1 Hz	[29]
PRECON	Pakistan	42	1 year	1 minute	[36]
CoSSMic	Germany	11	3 years	1 minute	[37]
ENERTALK	South Korea	22	29-122 days	15 Hz	[30]
SustData	Portugal	50	1144 days	2 - 10 Hz	[38]

1. The BLOND dataset [39] provides energy demand data for a typical office environment.

The dataset contains data from 53 appliances belonging to 17 different classes over 213 days. The data is sampled at a very high frequency: 50 kHz for aggregate energy demand and 6.4kHz for individual appliances.

2. The I-BLEND dataset [40] provides minute-level electricity demand data from a university campus in India. The data is available for 52 months. Additionally, the dataset also includes occupancy information for the campus buildings, sampled at a 10-minute interval.
3. The COMBED (commercial building energy data set) dataset [41] contains energy data for one month from IITD's academic building sampled at more than once a minute.
4. The IEEE Power and Energy Society (PES) has an additional repository of datasets from commercial buildings[4]. These include sub-metered measurements in different offices on the second scale. In addition, a number of competitions hosted on the IEEE Dataport platform also provide energy demand and production data. However, the data collection period is often on the order of a few weeks to months.
5. The ASHRAE great energy predictor III Kaggle challenge [42] provides an openly accessible dataset with hourly energy demand from 1449 buildings for around three years[5]. This competition is a follow-up to earlier competitions also conducted by ASHRAE [43].
6. The Building Data Genome Project [44] provides hourly electrical energy demand data from 507 non-residential buildings for one year. The dataset also contains meta data such as floor area, ambient conditions, and primary use of the building.

The information in this section is summarized in Table 2.

4.1.3. Miscellaneous demand datasets

In addition to energy demand from classical loads in the buildings (e.g. illumination, ventilation etc.), electrification of transport will make charging electric vehicles a significant load as well. This is also in line with European regulations which mandate minimum EV charging pole requirements for commercial buildings. However, openly available data for EV charging remains sparse. An exception is the dataset made available by ElaadNL[6] [45], which contains aggregated data from different types of charging stations. Likewise, the CoSSMic dataset highlighted above contains data on EV charging in some households as well.

Furthermore, while a number of datasets contain sub-metered energy demand for heating, cooling and ventilation in buildings, they seldom contain any information on indoor environment quality indicators such as temperature, humidity, air quality etc. An exception to this is the CU-BEMS dataset [46], which contains electricity demand data in a seven-story 11,700-m2 office building located in Bangkok, Thailand. The data provides measurements

[4]http://sites.ieee.org/pes-iss/data-sets/

[5]https://www.kaggle.com/c/ashrae-energy-prediction

[6]https://www.elaad.nl/news/elaadnl-shares-new-data-sets-demonstrating-the-rise-of-evs-and-usage-of-charging-stations/

Table 2: An overview of open-source commercial buildings data

Commercial buildings					
Dataset name	Country	Sites	Duration	Resolution	Ref
BLOND	Germany	Office/lab	50-230 days	Aggregate: 50-250 kHZ, Individual 6.4 (appliance-level) 50 kHz (aggregated))	[39]
I-BLEND	India	University	52 months	1 minute (load); 10 minutes (occupancy))	[40]
COMBED	India	University	7+ years	0.5 minutes	[41]
Building Data Genome	US, UK, Australia	500 (offices, universities, commercial)	1 year	Hourly	[44]
ASHRAE	Worldwide	1449	3 year	Hourly	[42]
IEEE PES	Multiple	Multiple	Multiple	Multiple	

for indoor environmental sensor data comprise temperature (°C), relative humidity (%), and ambient light (lux) measurements of 33 different zones in the building along with demand data for individual air conditioning units, lighting, and plug loads. The dataset is available on a minute-level resolution for a period of 18 months between July 1, 2018 and December 31, 2019.

Finally, in many cases, energy demand for larger aggregations, such as districts or states are openly available as time series which may be useful to understand the local context for LEC developers. Two examples include the national electricity demand in Belgium made available by Elia through its API, and the regionalized electricity and gas demand data for Germany made available on the OpenEnergy Platform[7].

To fully leverage the potential of demand side management in LECs, detailed data on important energy flexible resources such as heat and transportation is critical. While this is not the case at the moment, addressing this gap should consequently become a central aspect of the next generation of datasets. Moreover, a number of other limitations exist as well, as highlighted in Fig. 3, which plots the datasets based on their geographical locations and whether commercial or residential building data is available. It is obvious from this figure that while some regions are quite well represented (parts of Europe, North America, South Asia), there remain enormous gaps in data that is publicly available, especially from South America, Africa, Middle East, Eastern Europe, Scandinavia and China. The next

[7] https://openenergy-platform.org/dataedit/schemas

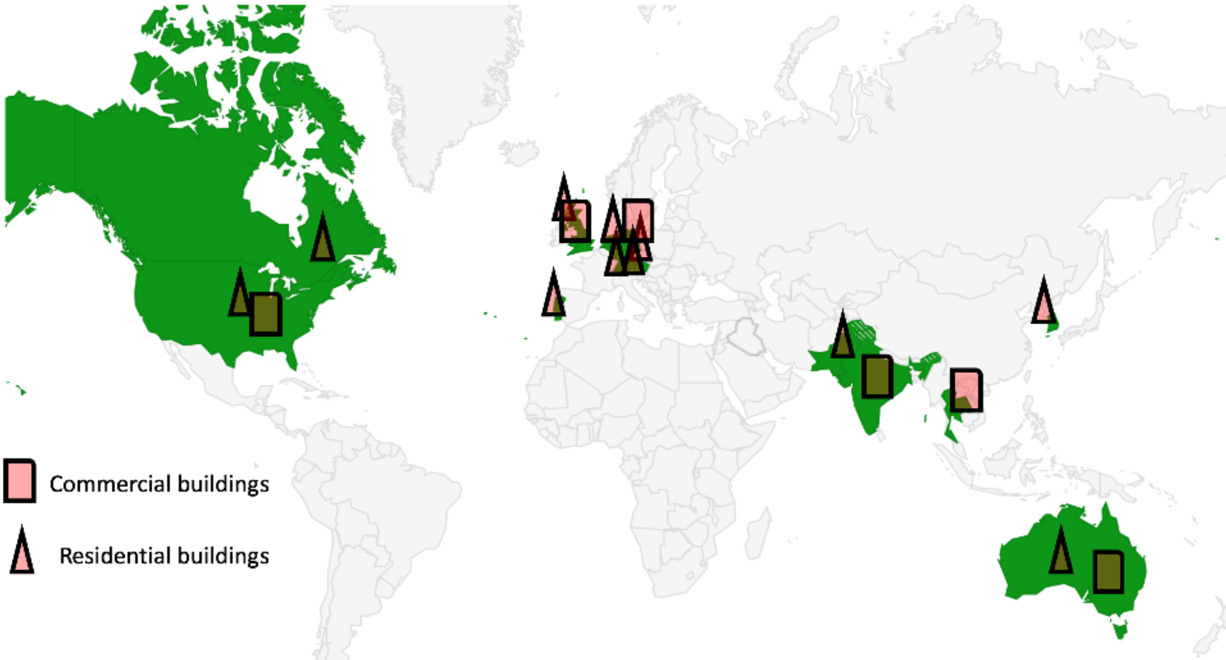

Figure 3: A geographical representation of electricity demand datasets: despite tens of datasets made available, most countries of the world and many climate zones are still not represented.

generation of data gathering should focus on these geographical locations.

4.2. Climate related data

Local weather conditions are among the biggest drivers behind both energy demand and energy flexibility in a LEC at any given time [47]. As such, it is important to know these conditions during both the design and operation phase. In this section, which is divided in three subsections, we briefly explore datasets related to local ambient conditions and climate. First, we present an overview of datasets dealing with real-time observations and forecasts for ambient conditions and solar power generation potential: these are useful for real-time operation. Second, we discuss historical and representative weather conditions in a given location: these are mostly suited for design phase considerations. Finally we discuss datasets that address the long term effect of climate change on ambient conditions. Naturally, there is some overlap, especially between the first two categories, as the same services frequently provide both historic data and short-term forecasts. We address this overlap by positioning these datasets according to their primary use in the field.

4.2.1. Observed and forecast weather conditions

These datasets can be divided into two categories. On the one hand they include observations and forecasts for general weather conditions which are required for reliable operation of LECs. On the other hand, they also include on-ground data for electricity production with renewable energy sources such as solar photovoltaics. In subsequent sections, we discuss models that can be used to estimate theoretical generation potential of solar PV systems in a region under ideal conditions, but here the focus is on actual observations and short term forecasts. Some of the resources making this data available include:

1. Online weather services, including OpenWeatherMap, Darksky (now acquired by Apple Inc.), Accuweather, Weatherbit and ClimaCell etc. Most such services have a free tier with limited access that provides historic, observational and future (short-term) forecasts for ambient conditions such as temperature, humidity, precipitation and cloud cover. The forecast horizons vary between a few days to weeks. These services are useful for micro-level analysis, as well as during the operation phase of a LEC. In many countries, national meteorological offices also open-source this data.
2. ECMWF CAMS real time provides ambient conditions data and forecasts for a number of meteorological variables including temperature, precipitation, snowfall and air quality metrics. The forecasts from the service are available on a multiple day horizon, and they are updated every six hours during a day (at 00:00, 06:00, 12:00 and 18:00)[8].
3. ENTSO-E, the European Network of Transmission System Operators for Electricity, compiles and provides day-ahead forecasts of power generation by renewable energy sources e.g. wind, solar PV etc. per bidding zone for the following day through its transparency platform [48]. This information is provided for all bidding zones in European Member States with greater than 1% feed-in of wind or solar power generation. More specifically, ENTSO-E provides (1) the current forecast, (2) the day ahead forecast at 18.00 (on the previous day) and (3) the intraday forecast at 08.00 (on the same day).
4. A number of transmission system operators, such as Elia in Belgium, also provide measurements as well as multi-scale (hour-ahead, day-ahead, week-ahead) forecasts for solar PV electricity production, disaggregated regionally which can provide more granular information to operators of LECs[9]. Likewise, EIA, the US Energy Information Administration, provides hourly electric grid data by eight different generation sources (including solar PV)[10]. Unlike with weather services and ECMWF CAMS real time, this data is specifically focused on actual electricity generation from installed renewable generation.
5. Some of the datasets highlighted in the section on demand (for instance, COSSMIC) also provide in-situ measurements of solar PV production. These resources can be invaluable as they allow researchers and practitioners to understand the deviation between theoretical and actual solar PV production. A number of other resources for solar production data exist as well: for instance, [49] provides data for 100 solar sites in North America from 1 Jan, 2015 to 1 Jan, 2016.

4.2.2. Historic weather conditions

While the datasets presented above are suitable for the operational phase of LECs, they may not be applicable during the design phase when practitioners work with representative

[8]https://apps.ecmwf.int/datasets/data/cams-nrealtime/levtype=sfc/

[9]https://www.elia.be/en/grid-data

[10]https://www.eia.gov/beta/electricity/gridmonitor/dashboard/electric$_{o}verview/US48/US48$

years or actual long-term observations, rather than data from a single year. In this specific case, alternatives based on global climate simulations or long term observations from weather stations or satellites might be preferable. Some of these datasets include:

1. ECMWF ERA5 provides a number of meteorological data fields through a separate reanalysis product, which spans back decades to also optimally combine historical data with future forecasts. The grid in this case is 31x31 km, and the data is only available on an hourly scale. While the service also provides forecasts as a result of its reanalysis, these can be especially inaccurate in coastal, mountainous or cloudy regions.[11].
2. SARAH 2[12] is another free-to-use dataset for ambient conditions with a higher spatial resolution, which is also available on 30 minute intervals.
3. The National Solar Radiation Database (NSRDB) provides hourly and half-hourly values of meteorological data as well as common measures of solar radiation for the United States and a growing subset of international locations. The dataset makes it possible to estimate the amount of solar energy at a given (historical) time and place, and to make forecasts for it for the future.
4. A number of services, including the NSRDB and others, provide Typical Meterological Year (TMY) data to explain the climate of a specified region, often in the EPW format which can be used with tools such as EnergyPlus [50]. These tools provide extensive coverage across the globe, and can be useful for a preliminary analysis, although they do not necessarily account for the effects of climate change or extreme weather events.

4.2.3. Long term forecasts for climate change

In addition to providing historic data or short-term forecasts for the future, a number of openly available data sources also provide long-term projections in light of climate change. These different climate models [51] usually rely on different pathways for future energy systems and the global economy [52]. For instance, a number of energy pathways exist to achieve the Paris climate goal of limiting average temperature increase below 2°C, however they invariably disagree on the actual measures required to make this happen (as well as the timing of these steps). This means that there is a lot of uncertainty about the emissions reductions at any given point in the future. This is exacerbated due to uncertain inertial and tipping point effects in the global climate. Consequently, climate models are inherently noisy as they rely on changes not just to global climate dynamics but also on factors such as the global economy and energy systems. Nevertheless, these scenarios can prove to be useful in future-proofing LECs. Some resources that provide this information include:

1. The WorldClim 2 dataset provides monthly climate data and projections for global land areas with a spatial resolution of 1 km^2. It contains data abaout temperature ranges, precipitation, solar radiation, vapour pressure and wind speed [53]. The monthly values that are made available as forecasts are calculated by averaging over 20 year

[11]https://confluence.ecmwf.int/plugins/servlet/mobile?contentId=129135000content/view/129135000
[12]https://wui.cmsaf.eu/safira/action/viewICDRDetails?acronym=SARAH$_V$002$_ICDR$

Table 3: An overview of openly available sources for climate and weather data

Climate data	
Domain	**Sources**
Historic weather conditions	ECMWF ERA 5, SARAH 2
Observations and forecasts	Macro-level sources: ECMWF CAMS, ENTSO-e (Europe), EIA (US), country specific sources such as Elia in Belgium, REE in Spain etc.; Micro-level sources: OpenWeatherMap, ClimaCell,, COSSMIC, SunDance etc.
Climate change	WorldClim 2, World Bank dashboard

periods (2021-2040, 2 041-2060, 2061-2080, 2081-2100). Both historical and future data are available online, as downscaled CMIP6 future climate projections[13].

2. The World Bank provides similar data fields in a more accessible fashion through their climate change knowledge portal[14]. With this tool, it is straightforward to export historical and future projections for temperature (minimum, maximum and average), precipitation, heating and cooling degree days etc. It is possible to access this information by longitude/latitude or country name. These projections are based on CMIP5 simulations for the global climate, the precursor to CMIP6 simulations used by WorldClim above.
3. A number of service providers, such as White Box Technologies, are also developing EPW files for typical year sets for future ambient conditions, taking into account the effects of climate change. These datasets are however not always open-sourced, and are simply mentioned for the sake of completeness.

The information in this section is summarized in Table 3.

5. Models

Models allow practitioners to generate data for specific scenarios and assets when it is not possible to obtain real data, due to technical or economic constraints. Especially for LECs, this can be quite useful for cases where no similar projects have been executed before (or the data generated in them is not publicly available). In this section we describe a number of demand and generation models following the same classification scheme as before. Additionally, in this section, we also consider resources that model storage and the electricity grid.

[13] https://www.worldclim.org
[14] https://climateknowledgeportal.worldbank.org/download-data

5.1. Models for electricity demand

In this section, we focus on generative models that can simulate energy-related behaviour of buildings and building occupants. These models usually rely either on top-down models, which are typically based on behavioural models of user demand, or on bottom-up approaches, which emphasize individual appliance usage and are built using either diary data or, increasingly, sub-metered or disaggregated appliance loads. The following open-source models can be used to create electricity demand i.e. load profiles for residential and commercial building users:

1. Load Profile Generator (LPG) generates artificial load profiles for residential energy consumption by simulating the people in one household, thereby obtaining their load curves [54]. The model, based on German households, uses insights from psychology and can generate data for large populations of up to 1000 households. As a limitation, it can be time-consuming to define a household from scratch, in the case that the user wants to simulate a household with no flexible assets or DERs associated.
2. The Artificial Load Profile Generator (ALPG) [55] is a model that uses high level demographic statistics as an input to build a bottom-up model and generate electricity load profiles. This is particularly interesting for LECs as the model characterizes devices as uncontrollable, curtailale, time-shiftable, buffer time-shiftable and buffer. Some of its limitations are the low scalability, with a maximum of 100 households to simulate, and the fact that individual flexible assets are not observable, since they have pre-defined labels such as buffered, time-shiftable, etc.
3. The House Load Electricity [56] is an application programmed in MATLAB that generates synthetic electricity load profiles based on consumer loads. The application comes with a Graphical User Interface (GUI) where the end-user can chose the model parameters, as well as change the time resolution and time periods generated.
4. Office Load MATLAB Application generates synthetic load profiles, but in this case for office buildings [57]. As in the previous application, the user can specify the parameters in a MATLAB based GUI, representative for Northen Europe. The model is based on a bottom-up approach, collecting both the behavior of the office workers as well as the appliances installed in the specified office. Energy use of heating and air conditioning systems is also taken into consideration.
5. Demandlib [58] is an open-source Python package that allows users to create power and heat profiles for various sectors by following the BDEW guidelines. Besides specifying building type, it is also possible to generate heat and load profiles in a way that preserves calendar aspects such as national holidays.

5.2. Models for electricity generation

In addition to the data sources mentioned in the previous section, there are a number of models available to simulate the behavior of DERs in new locations. In this section, we focus on solar PV panels. The next section completes the DER picture with battery models.

1. The PV Performance Modeling Collaborative (PVMPC) is an initiative from the Sandia National Laboratories which is working towards improving the accuracy and technical rigor of PV performance models and analyses. The first models coming out of this initiative were provided as MATLAB libraries [59], which were later ported to Python as the pvlib-python library [60]. Like its MATLAB counterpart, this open-source library can also be used to develop PV power forecasting tools, as well as evaluating different configurations of PV systems [61]. The models include a number of parameters such as the sun position, irradiance and insolation, array orientation, shading, soiling and reflection losses, as well as inverter technical characteristics that can be specified by the end-user.
2. PVWatts [62], developed by the National Renewable Energy Laboratory, provides a browser-based PV system model that estimates the electricity output and economic costs of grid-connected PV energy systems, based on geographic location, technical characteristics of the PV system and local market conditions.
3. Renewables.ninja is a web-based system that contains a PV model to estimate its electricity power output for any location (input as either country or a latitude-longitude combination) [63]. The user can modify the technical specifications of the system, such as tilt, azimuth or capacity, or work with ready-made datasets available by country in the tool. Renewables.ninja is a GUI of the python model developed in [64].
4. OEMOF's feedinlib is a Python package[15] that allows users to estimate power production of solar PV and wind power plants as time series, based on defined system parameters and weather conditions. In essence, for solar PV systems, this library acts as a high level wrapper to the pvlib library described above.
5. Atlite[16] is an open-sourced Python package developed using the Aarhus University RE Atlas [65] which allows users to convert weather data (temperature, irradiation, wind speed etc.) into energy demand and generation time series. The package can create time series for wind power, solar power, hydro power and heating demand at hourly resolution on a 30 km x 30 km grid using the recommended ERA5 weather reanalysis dataset. The library is designed to work with big weather and climate datasets at a low computational cost.

5.3. Storage models

Energy storage systems (ESS) are among the most effective flexibility sources and can enable greater proliferation of renewables by enabling services such as self-consumption. In the case of batteries, a general lack of real world observation data and the increasing number of projects integrating batteries has pushed the community to create and share their models, playing a key role in helping the integration of ESS in LECs. Some of these models include:

1. QuESt - Optimizing Energy Storage is a python-based application that contains energy storage models for simulation and analysis purposes [66, 67]. Furthermore, the

[15] https://feedinlib.readthedocs.io/en/latest/
[16] https://atlite.readthedocs.io/en/

application includes a data acquisition option that allows import of market data, transmission system data, load profiles and PV power profiles to calculate the profitability of the ESS.

2. OSESMO provides a battery model (for Lithium-ion and flow batteries) and calculates the optimal charge-discharge strategy in 15-minute time periods using linear programming, in order to minimize the end-user monthly bill [68].
3. EnergyBoost is a learning-based control tool for home batteries that models a Lithium-ion battery, connected with a PV energy system [69]. This physical model is then used to control the battery charge and discharge operation by considering it an optimal control problem.
4. While EV battery models are comparatively more difficult to find than their fixed battery counterparts, Geotab has recently open-sourced a detailed EV battery degradation model over time[17]. This model allows users to visualize battery degradation over time for a number of different makes and models of EVs. Another recently open-sourced model for the temporal modeling of EV sessions is presented in [70]. This synthetic data generator uses real world charging sessions to jointly model the arrival and departure times of EVs. emobpy[18] is another recent Python package that uses empirical data to model and simulate EVs in some detail [71]. More concretely, it can generate time series for electricity charging demand of EVs as well as their mobility state (i.e. where they are located etc.) and grid availability in terms of available infrastructure. vencopy is another recently open-sourced tool that can be used to estimate the energy flexibility potential of an EV fleet [72]. However, both of these tools, emobpy and vencopy, are currently in the early stages of their development. Finally, the Simulation of Urban MObility (SUMO) open source package [73], while not EV specific, allows users to simulate traffic in the transportation system as a whole. This can be useful in better understanding the dynamics of EV energy flexibility in the broader context.

5.4. Other models of interest

In addition to electricity demand, storage and generation (via renewables), a number of other models can also be utilized to optimize LECs. An example is the System Advisory Model (SAM) created by NREL [74], which provides end-to-end decision making support for micro-grids and LECs. SAM incorporates different models e.g. of renewable energy systems such as solar PV systems, energy storage systems etc. and can therefore be used to obtain data for the entire community considering different renewable and flexible assets installed [75].

Additionally, models for power systems can also be critical to better understanding the practical feasibility of LECs, especially with increasing proliferation of DERs, and heat and transport electrification. There are a number of such models available, even though

[17] https://storage.googleapis.com/geotab-sandbox/ev-battery-degradation/index.html
[18] https://gitlab.com/diw-evu/emobpy/emobpy

actual data remains sparse. These include the PowerGenome project [76], the Open Energy Modeling Framework (OEMOF) [77], the multi-vector simulator [78], and the renpassGIS [79]. A number of low-level alternatives exist as well. For instance, Pandapower [80], a general-purpose python-based power system analysis tool, allows the user to run DC optimal power flow calculations of electrical grids. These can be used to maximize the utility of the LECs as described earlier. Pandapower is based on the now deprecated Pypower and a well-known MATLAB Tool for power system analysis, MATPOWER [81]. Another (non-Pythonic) open-source alternative for simulating electric power systems is OpenDSS [82]. OpenDSS models and simulates distribution networks as stand-alone executable programs, and can be used for planning a LEC and operating the flexibility resources contained in it.

Finally, the possibility to simulate the performance and operation of building systems can assist in the optimal definition of resources to locate in a local energy community or to change their operation once they have been installed. While the focus so far was on electric storage, a number of modelling tools also exist for thermal systems. These include Python libraries such as RC BuildingSimulator[19] and TEASER[20], which can be used to create detailed thermal models for the built environment. More general-purpose tools such as EnergyPlus can also be useful for this purpose. In the absence of detailed indoor temperature measurements, these can be a valuable resource for LEC practitioners in understanding and leveraging flexibility in practice. Other alternatives that take a more holistic approach exist as well such as City Learn, an OpenAI Gym environment[21] [83]. This tool allows the implementation of different (reinforcement learning-based controllers to reshape electricity demand in a building by controlling the operation of flexible resources such as domestic water boilers, electric heaters, PV panels, and space cooling.

The information in this section is summarized in table 4.

6. General purpose tools

As mentioned previously, a number of openly available, general-purpose tools exist to enable practitioners to analyze and draw insights from the datasets and models discussed above. These tools, in general, require some knowledge of data science and decision support systems. Consequently, we expect them to only be useful for practitioners in LECs interested in taking a deeper dive into their data. As there is a large number of such tools, we include only a broad overview of some of the most promising and widely utilized tools in Python, arguably the most popular data science language at the moment. Similar packages typically exist in R, MATLAB and other programming languages. The applications in this section broadly mirror the same use cases as the ones highlighted earlier in this review.

[19]https://github.com/architecture-building-systems/RC$_{B}$*uildingSimulator*

[20]https://github.com/RWTH-EBC/TEASER

[21]OpenAI Gym is a general purpose toolkit for developing and comparing reinforcement learning algorithms

Table 4: An overview of open-source models that can be used by LECs, and their typical applications

Models	
Domain	**Sources**
Synthetic demand data	Load Profile Generator (LPG) [54, 84], Artificial Load Profile Generator (ALPG) [85, 55], House Load Electricity [86, 56] , Office Load MATLAB application [87, 57], demandlib [58]
Generation potential of renewables (PV)	Sandia Labs PV Performance model Program (PVPMC) [59], PVLIB- Python [60], NREL PVWatts [62], Renewables.ninja [63], feedinlib, atlite
Storage models	QuESt [66, 67], OSESMO [68], EnergyBoost [69], Geotab-Sandbox, emobpy [71], vencopy [72], SUMO [73]
Power grids and systems	System Advisory Model (SAM) PowerGenome Project [76], Open Energy Modeling Framework (OEMOF) [77, 88], Multi-Vector Simulator [78], renpassGIS [79]
Building systems simulation	TEASER, RC BuildingSimulation, City Learn
Power systems simulation	Pandapower, MATPOWER, OpenDSS

6.1. *Data wrangling and visualization*

We have highlighted numerous data sources as well as models that can be used to generate data in the previous sections. Once the data have been acquired, the next step is exploratory data analysis (EDA) and create dashboards to provide actionable feedback to users. This includes both summarization of key metrics (which can be achieved through libraries such as Pandas in Python), and visualizaiton of key trends. While Matplotlib is arguably the most well-known python library for visualizing data, other libraries such as Seaborn [89] allow for the creation of much more visually appealing and intuitive plots. Recently, a number of tools such as Dash Plotly [90] and Streamlit [91] have emerged that allow users to create engaging dashboards, which can be shared with other people easily over the internet. Grafana is another popular choice for creating custom dashboards quickly. It must be emphasized here that the services enabled by these lower-level Python libraries are quite distinct from the ones highlighted with energy-specific datasets and models in the previous sections as they allow far greater customization.

6.2. *Modelling and forecasting*

Forecasting demand and generation using data-driven (or physical) models is arguably one of the core activities in a LEC. While we have highlighted services that provide forecasts for some relevant variables including solar generation and ambient conditions, these usually do not cover all the forecasts that need to be made in a LEC, including for local demand as well as grid and market conditions. These services can either be out-sourced to a third party, or they can be done using statistical and machine learning algorithms. Modern low-code libraries such as Darts, sktime [92] and Pycaret [93] allow users to quickly build forecasting models in a few lines of code. While many of these libraries are wrappers around other low-level Python libraries, they also offer other important tools such as data imputation and backtesting to understand how well a model will perform in real world conditions. Increasingly, no-code alternatives such as Dataiku, which include a free-tier for non-commercial users, are gaining traction as well.

6.3. *Design and operational optimization*

As opposed to time series analysis and machine learning methods for modelling and forecasting energy loads, decision support tools are necessary to optimally design and operate the LEC. Decision support systems can take on numerous form, some of these were already presented in the modelling section. However, a large body of general-purpose optimizers exist as well. These include tools that are primarily used for convex problems such as Pyomo [94], PuLP [95] and CVXPY. Many such tools allow users to formulate the problem in a higher level language, while using well-known solvers such as GLPK, CPLEX and Gurobi as the backend. Some of these frameworks also support solving stochastic programs, which can be important for risk averse (or risk aware) planning. Likewise, a number of libraries such as DEAP [96], Nevergrad [97] and Optuna allow the solution of nonconvex problems, which can often be required when the underlying models are built using machine learning algorithms. Finally, in recent years, many variants of reinforcement learning have emerged as feasible optimisation strategies for the energy domain, especially in settings where a model

Table 5: An overview of open-source tools that can be used by LECs and their typical applications

Tools	
Domain	**Sources**
Data visualization and dashboarding	*Visualization*: Matplotlib, Seaborn, Plotly Dash; *Interactive dashboarding*: Panel and Streamlit
Modelling and forecasting	*Time series modelling*: Darts, Sktime, Facebook Prophet; *Machine learning modelling and clustering*: Scikit-learn, Tensorflow, PyTorch, Pycaret etc.
Design and operational optimization	*Convex optimization*: Pyomo, Gurobi, PuLP, CVXPY; *Non-convex optimization*: Nevergrad; DEAP; *Reinforcement learning*: City Learn project, ChainerRL, KerasRL, TensorForce

for the environment is not available. While this is a rapidly evolving space, libraries such as ChainerRL [98], KerasRL [99] and TensorForce [100] provide off-the-shelf implementations of a number of standard reinforcement learning algorithms.

6.4. Tracking tools

A relative newcomer that can automate the tracking of energy flows in a LEC is blockchain or distributed ledger technology. Blockchain technology has found accelerating adoption in LECs and peer-to-peer energy trading concepts. Such blockchains can be implemented in the definition of smart contracts, enhancing the deployment of peer-to-peer markets and therefore ensuring the traceability of the energy that is being produced and consumed within the community. Hashlib is a popular tool for Blockchain definition in Python [101], however this is a rapidly evolving field as well.

The information in this section is summarized in table 5.

7. Conclusions

In this paper, we have focused on local energy communities, and provided a detailed overview of publicly available datasets, tools and models that can be used to optimize their design and operation. While the energy domain has, in general, lagged behind other sectors when it comes to digitization, we have gathered a multitude of resources in this review. These include openly available datasets for energy demand and generation (via renewable energy sources). Since there are lots of regions of the world which still do not have any open datasets, we complement this information with models that can generate such data in more general settings. Where one or the other should be used, depends on the use case. A general rule of thumb is to use actual datasets if they are similar enough to the current use case. Models excel, however, when the design must use projections for the future. This will become increasingly important while designing LECs keeping the effects of climate change and a changing energy supply mix in mind. Finally, the tools we highlighted in this paper include general purpose data science related frameworks which can allow data visualization and modelling as well as optimization and tracking of energy flows. This is especially important for players in the platform archetype highlighted in section 2, but can also be useful in the aggregator and microgrid archetypes.

More concretely, the resources presented in this paper can be utilized by different roles in real-world LECs in numerous ways. These are explained in section 2 on data use case in LECs. Besides these archetypes, it is also possible to distil these guidelines based on whether the LEC is currently operational or not:

1. Different stakeholders in **aspiring LECs** (e.g. prosumers, DSOs, balance responsible parties, third party service providers etc.) often need to make assumptions about the energy demand and generation potential in the LEC to optimally dimension the energy flexible resources (e.g. battery-inverter etc.) and distributed or community generation. This analysis can be considerably improved by making use of real-world data from similar projects (when it is available) or models that can simulate this data. Likewise, such aspiring LECs can use the frameworks and tools presented in this paper with pre-existing data or models to simulate future performance in different scenarios and estimate profitability while keeping local regulations in mind.
2. **Existing LECs** will likely find the tools presented in this paper more useful than existing datasets. For instance, the data visualization and modelling tools can help third party LEC operators develop tools that can be used to engage users and prosumers in the LEC. On the other hand, the optimization algorithms and frameworks presented in this paper can be used to both optimally dimension the energy flexible resources (e.g. DERs, heat pumps etc.) and also to operate them in a way that it minimizes grid energy offtake and maximizes self-consumption. Finally, the grid models and frameworks (e.g. OpenDSS, Pandapower etc.) can be used by advanced practitioners to ensure that the energy flexible resources in the LEC can also help keep distribution grid operation stable (e.g. minimizing voltage and congestion issues etc.).

In gathering these resources, we have also unearthed some obvious shortcomings. These include a general lack of openly available data for electric vehicles. This extends beyond the energy demand, and encompasses mobility patterns and the long-term longevity of EV batteries, especially if they are used with fast chargers or in vehicle-to-grid mechanisms. Likewise, detailed data from space conditioning (heating, cooling, ventilation) and hot water production is quite limited, when compared with electricity demand data. Even in cases where the energy demand for heating and cooling is sub-metered, it is not often that detailed temperature values are recorded inside the medium of interest (e.g. building and/or hot water vessel). This lack of temperature data limits the usefulness of the data with regards to estimating the amount of available flexibility. Finally, even though demand data from thousands of buildings is openly available, it is often limited either in space or time. Temporal limitations mean it is difficult to generalize the observed trends and detect seasonality in the data. Likewise, the fact that most datasets are concentrated in a few geographical regions does not lend itself to broader generalization. As LECs are seen as a potential solution for sustainable community-driven electrification in developing countries, this is a major shortcoming at the moment which needs to be addressed in future work.

The situation with open-source tools is rather different. Here, numerous options exist to help LECs analyze and better understand data. However, even thought this is slowly changing, many of these tools require coding proficiency which can limit their general applicability. More applications with intuitive user interfaces will hopefully alleviate these issues in the years to come. Future work to address these issues will therefore greatly assist in not just the design and operation of LECs, but the broader field of smart buildings and grids as a whole. Using the resources highlighted in this review as a basis for an easy to use, holistic framework, which integrates the many different elements of a LEC, can also further lower the barriers to widespread operationalization of LECs.

8. Acknowledgments

Hussain Kazmi acknowledges support from EIT InnoEnergy and Research Foundation – Flanders (FWO), Belgium (research fellowship 1262921N) in the preparation of this manuscript. The authors would also like to acknowledge the editor and anonymous reviewers for their feedback and contribution.

References

[1] IEA, Digitalisation and energy, Technical Report (2017).

[2] G. Walker, What are the barriers and incentives for community-owned means of energy production and use?, Energy policy 36 (12) (2008) 4401–4405.

[3] G. Walker, P. Devine-Wright, Community renewable energy: What should it mean?, Energy policy 36 (2) (2008) 497–500.

[4] B. P. Koirala, E. Koliou, J. Friege, R. A. Hakvoort, P. M. Herder, Energetic communities for community energy: A review of key issues and trends shaping integrated community energy systems, Renewable and Sustainable Energy Reviews 56 (2016) 722–744.

[5] C. Romero-Rubio, J. R. de Andrés Díaz, Sustainable energy communities: a study contrasting spain and germany, Energy Policy 85 (2015) 397–409.

[6] B. Wiersma, P. Devine-Wright, Decentralising energy: comparing the drivers and influencers of projects led by public, private, community and third sector actors, Contemporary Social Science 9 (4) (2014) 456–470.

[7] B. J. Kalkbrenner, J. Roosen, Citizens' willingness to participate in local renewable energy projects: The role of community and trust in germany, Energy Research & Social Science 13 (2016) 60–70.

[8] P. Devine-Wright, Community versus local energy in a context of climate emergency, Nature Energy 4 (11) (2019) 894–896.

[9] I. F. Reis, I. Gonçalves, M. A. Lopes, C. H. Antunes, Business models for energy communities: A review of key issues and trends, Renewable and Sustainable Energy Reviews 144 (2021) 111013.

[10] E. Union, Directive (eu) 2018/2001 of the european parliament and of the council of 11 december 2018 on the promotion of the use of energy from renewable sources, Official Journal of the European Union 5 (2018) 82–209.

[11] I. Munné-Collado, E. Bullich-Massagué, M. Aragüés-Peñalba, P. Olivella-Rosell, Local and Micro Power Markets, John Wiley Sons, Ltd, 2019, Ch. 2, pp. 37–96.

[12] B. Koirala, D. Vaghela, M. Hiremath, R. Kulenthran, Opportunities and challenges of community energy systems: analysis of community micro-hydro systems in south and south-east asia (ssea), Micro Energy System, TU Berlin and Berkeley Rural Energy Group, Univeristy of California (2014) 63.
[13] L. Lyons, Digitalisation: Opportunities for heating and cooling, Publications Office of the European Union: Luxembourg.
[14] D. Spichtinger, Open access in horizon 2020 and the european research area (2012).
[15] G. Von Krogh, S. Haefliger, S. Spaeth, M. W. Wallin, Carrots and rainbows: Motivation and social practice in open source software development, MIS quarterly (2012) 649–676.
[16] P. S. Kochhar, E. Kalliamvakou, N. Nagappan, T. Zimmermann, C. Bird, Moving from closed to open source: Observations from six transitioned projects to github, IEEE Transactions on Software Engineering.
[17] S. Pfenninger, J. DeCarolis, L. Hirth, S. Quoilin, I. Staffell, The importance of open data and software: Is energy research lagging behind?, Energy Policy 101 (2017) 211–215.
[18] Z. El Mrabet, N. Kaabouch, H. El Ghazi, H. El Ghazi, Cyber-security in smart grid: Survey and challenges, Computers & Electrical Engineering 67 (2018) 469–482.
[19] L. A. Greening, D. L. Greene, C. Difiglio, Energy efficiency and consumption—the rebound effect—a survey, Energy policy 28 (6-7) (2000) 389–401.
[20] F. Tounquet, Energy communities in the european union.
[21] O. Iweka, S. Liu, A. Shukla, D. Yan, Energy and behaviour at home: a review of intervention methods and practices, Energy Research & Social Science 57 (2019) 101238.
[22] F. Jazizadeh, M. Afzalan, B. Becerik-Gerber, L. Soibelman, Embed: a dataset for energy monitoring through building electricity disaggregation, in: Proceedings of the Ninth International Conference on Future Energy Systems, 2018, pp. 230–235.
[23] J. Z. Kolter, M. J. Johnson, Redd: A public data set for energy disaggregation research, in: Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA, Vol. 25, 2011, pp. 59–62.
[24] A. Filip, Blued: A fully labeled public dataset for event-based nonintrusive load monitoring research, in: 2nd Workshop on Data Mining Applications in Sustainability (SustKDD), 2011, p. 2012.
[25] J. Gao, S. Giri, E. C. Kara, M. Bergés, Plaid: a public dataset of high-resoultion electrical appliance measurements for load identification research: demo abstract, in: proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, 2014, pp. 198–199.
[26] A. Einfalt, C. LEITINGER, D. TIEFGRABER, S. GHAEMI, Adres concept–micro grids in österreich, Internationale Energiewirtschaftstagung an der TU Wien. Wien: IEWT.
[27] J. Kelly, W. Knottenbelt, The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes, Scientific data 2 (1) (2015) 1–14.
[28] A. S. Uttama Nambi, A. Reyes Lua, V. R. Prasad, Loced: Location-aware energy disaggregation framework, in: Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, 2015, pp. 45–54.
[29] 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 conference on embedded systems for energy-efficient buildings, 2014, pp. 80–89.
[30] 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 6 (1) (2019) 1–13.
[31] O. Parson, G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, W. Knottenbelt, A. Rogers, Dataport and nilmtk: A building data set designed for non-intrusive load monitoring, in: 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2015, pp. 210–214.
[32] D. Murray, L. Stankovic, V. Stankovic, An electrical load measurements dataset of united kingdom households from a two-year longitudinal study, Scientific data 4 (1) (2017) 1–12.
[33] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, J. Albrecht, et al., Smart*: An open data set and tools for enabling research in sustainable homes, SustKDD, August 111 (112) (2012) 108.
[34] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, P. Shenoy, Non-intrusive occupancy monitoring using

smart meters, in: Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings, 2013, pp. 1–8.

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

[36] A. Nadeem, N. Arshad, Precon: Pakistan residential electricity consumption dataset, in: Proceedings of the Tenth ACM International Conference on Future Energy Systems, 2019, pp. 52–57.

[37] A. Amato, R. Aversa, B. Di Martino, M. Scialdone, S. Venticinque, S. Hallsteinsen, G. Horn, Software agents for collaborating smart solar-powered micro-grids, in: Smart Organizations and Smart Artifacts, Springer, 2014, pp. 125–133.

[38] L. Pereira, F. Quintal, R. Gonçalves, N. J. Nunes, Sustdata: A public dataset for ict4s electric energy research, in: ICT for Sustainability 2014 (ICT4S-14), Atlantis Press, 2014, pp. 359–368.

[39] T. Kriechbaumer, H.-A. Jacobsen, Blond, a building-level office environment dataset of typical electrical appliances, Scientific data 5 (2018) 180048.

[40] H. Rashid, P. Singh, A. Singh, I-blend, a campus-scale commercial and residential buildings electrical energy dataset, Scientific data 6 (2019) 190015.

[41] N. Batra, O. Parson, M. Berges, A. Singh, A. Rogers, A comparison of non-intrusive load monitoring methods for commercial and residential buildings, arXiv preprint arXiv:1408.6595.

[42] C. Miller, P. Arjunan, A. Kathirgamanathan, C. Fu, J. Roth, J. Y. Park, C. Balbach, K. Gowri, Z. Nagy, A. D. Fontanini, et al., The ashrae great energy predictor iii competition: Overview and results, Science and Technology for the Built Environment 26 (10) (2020) 1427–1447.

[43] J. Haberl, S. Thamilseran, The great energy predictor shootout ii, ASHRAE journal 40 (1) (1998) 49.

[44] C. Miller, F. Meggers, The building data genome project: An open, public data set from non-residential building electrical meters, Energy Procedia 122 (2017) 439 – 444, {CISBAT} 2017 International ConferenceFuture Buildings amp; Districts – Energy Efficiency from Nano to Urban Scale. doi:https://doi.org/10.1016/j.egypro.2017.07.400.
URL `http://www.sciencedirect.com/science/article/pii/S1876610217330047`

[45] M. Zweistra, S. Janssen, F. Geerts, Large scale smart charging of electric vehicles in practice, Energies 13 (2) (2020) 298. doi:10.3390/en13020298.
URL `http://dx.doi.org/10.3390/en13020298`

[46] M. Pipattanasomporn, G. Chitalia, J. Songsiri, C. Aswakul, W. Pora, S. Suwankawin, K. Audomvongseree, N. Hoonchareon, Cu-bems, smart building electricity consumption and indoor environmental sensor datasets, Scientific Data 7 (1) (2020) 1–14.

[47] A. Balint, H. Kazmi, Determinants of energy flexibility in residential hot water systems, Energy and Buildings 188 (2019) 286–296.

[48] L. Hirth, J. Mühlenpfordt, M. Bulkeley, The entso-e transparency platform–a review of europe's most ambitious electricity data platform, Applied energy 225 (2018) 1054–1067.

[49] D. Chen, D. Irwin, Sundance: Black-box behind-the-meter solar disaggregation, in: Proceedings of the Eighth International Conference on Future Energy Systems, 2017, pp. 45–55.

[50] N. Fumo, P. Mago, R. Luck, Methodology to estimate building energy consumption using energyplus benchmark models, Energy and Buildings 42 (12) (2010) 2331–2337.

[51] G. A. Meehl, C. A. Senior, R. J. Stouffer, K. E. Taylor, Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization, Geoscientific Model Development 9 (5) (2016) 1937.

[52] A. G. Dagnachew, A. F. Hof, P. L. Lucas, D. P. van Vuuren, Insight into energy scenarios: A comparison of key transition indicators of 2° c scenarios (2019).

[53] S. E. Fick, R. J. Hijmans, Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas, International journal of climatology 37 (12) (2017) 4302–4315.

[54] N. Pflugradt, Modellierung von Wasser-und Energieverbräuchen in Haushalten, Technischen Universität Chemnitz (April) (2016) 373. doi:10.3945/jn.113.174375.

[55] G. Hoogsteen, A cyber-physical systems perspective on decentralized energy management, 2017.

doi:10.3990/1.9789036544320.
URL http://purl.org/utwente/doi/10.3990/1.9789036544320
[56] C. Sandels, J. Widén, L. Nordstrom, House Load Electricity Model (2018).
URL https://www.mathworks.com/matlabcentral/fileexchange/63375-house-load-electricity
[57] C. Sandels, D. Brodén, J. Widén, L. Nordström, E. Andersson, Modeling office building consumer load with a combined physical and behavioral approach: Simulation and validation, Applied Energy 162 (2016) 472–485. doi:10.1016/j.apenergy.2015.10.141.
URL http://dx.doi.org/10.1016/j.apenergy.2015.10.141
[58] M. Hellwig, Entwicklung und anwendung parametrisierter standard-lastprofile, Ph.D. thesis, Technische Universität München (2003).
[59] Sandia National Laboratories, PVPerformance Modeling Collaborative - PVLIB MATLAB.
URL https://pvpmc.sandia.gov/
[60] W. Holmgren, C. Hansen, M. Mikofski, pvlib python: a python package for modeling solar energy systems, Journal of Open Source Software 3 (29) (2018) 884. doi:10.21105/joss.00884.
URL https://doi.org/10.21105/joss.00884
[61] D. J. Gagne, A. McGovern, S. E. Haupt, J. K. Williams, Evaluation of statistical learning configurations for gridded solar irradiance forecasting, Solar Energy 150 (2017) 383–393. doi:https://doi.org/10.1016/j.solener.2017.04.031.
URL http://www.sciencedirect.com/science/article/pii/S0038092X17303158
[62] National Renewable Energy Laboratory, PVWatts v6.1.3.
URL http://pvwatts.nrel.gov/
[63] ETH Zürich, Imperial College London, Renewables.ninja - Electricity power output calculation.
URL https://www.renewables.ninja/
[64] S. Pfenninger, I. Staffell, Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data, Energy 114 (2016) 1251–1265. doi:https://doi.org/10.1016/j.energy.2016.08.060.
URL http://www.sciencedirect.com/science/article/pii/S0360544216311744
[65] G. B. Andresen, A. A. Søndergaard, M. Greiner, Validation of danish wind time series from a new global renewable energy atlas for energy system analysis, Energy 93 (2015) 1074–1088.
[66] R. H. Byrne, R. J. Concepcion, C. A. Silva Monroy, Estimating potential revenue from electrical energy storage in pjm.doi:10.1109/PESGM.2016.7741915.
[67] T. A. Nguyen, R. H. Byrne, D. Copp, R. J. Concepcion, B. Chalamala, Opportunities for energy storage in caiso: Day-ahead and real-time market arbitrage.
[68] R. Mann, J. Grygier, Simulation of Proposed GHG-Reduction Solution with Open-Source Energy Storage Model (2018).
[69] B. Qi, M. Rashedi, O. Ardakanian, Energyboost: Learning-based control of home batteries, 2019, pp. 239–250. doi:10.1145/3307772.3328279.
[70] M. Lahariya, D. F. Benoit, C. Develder, Synthetic data generator for electric vehicle charging sessions: Modeling and evaluation using real-world data, Energies 13 (16) (2020) 4211.
[71] C. Gaete-Morales, H. Kramer, W.-P. Schill, A. Zerrahn, An open tool for creating battery-electric vehicle time series from empirical data: emobpy, arXiv preprint arXiv:2005.02765.
[72] N. Wulff, Emobility in python (emobpy) & vehicle energy consumption in python (vencopy). demonstration of two open source tools describing electric vehicle energy demand.
[73] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, E. Wießner, Microscopic traffic simulation using sumo, in: The 21st IEEE International Conference on Intelligent Transportation Systems, IEEE, 2018.
URL https://elib.dlr.de/124092/
[74] U.S. Department of Energy, WINDExchange: Wind Energy Models and Tools.
URL https://sam.nrel.gov/
[75] N. Blair, N. Diorio, J. Freeman, P. Gilman, S. Janzou, T. W. Neises, M. J. Wagner, System Advisor Model (SAM) General Description (NREL/TP-6A20-70414).

URL https://www.nrel.gov/docs/fy18osti/70414.pdf
[76] G. Schively, PowerGenome.
URL https://github.com/gschivley/PowerGenome
[77] S. Hilpert, C. Kaldemeyer, K. Uwe, S. Günther, c. wingenbach, G. Pleßmann, The open energy modelling framework (oemof) - a new approach to facilitate open science in energy system modelling, Energy Strategy Reviews 22 (2018) 16–25. doi:10.1016/j.esr.2018.07.001.
[78] Reiner Lemoine Institut, MVS - Multi-Vector Simulator of the E-land toolbox (2019).
URL https://github.com/rl-institut/mvs_eland
[79] C. Kaldemeyer, M. Söthe, S. Hilpert, C. Wingenbach, renpassG!S - A Free and Open Python Tool for Simulating Energy Supply Systems (Feb 2019). doi:10.17605/OSF.IO/KP4MH.
URL osf.io/kp4mh
[80] L. Thurner, A. Scheidler, F. Sch, J.-h. Menke, J. Dollichon, F. Meier, pandapower - an Open Source Python Tool for Convenient Modeling , Analysis and Optimization of Electric Power Systems (2018) 1–12arXiv:arXiv:1709.06743v3.
[81] R. D. Zimmerman, C. E. Murillo-Sánchez, R. J. Thomas, Matpower: Steady-state operations, planning, and analysis tools for power systems research and education, IEEE Transactions on Power Systems 26 (2011) 12–19.
[82] R. C. Dugan, The Open Distribution System Simulator (OpenDSS) - Reference Manual, Electric Power Research Institute (EPRI) (April) (2018) 1–199.
URL http://download2.nust.na/pub4/sourceforge/e/el/electricdss/OpenDSS/OpenDSSManual.pdf
[83] J. R. Vázquez-Canteli, J. Kämpf, G. Henze, Z. Nagy, Citylearn v1.0: An openai gym environment for demand response with deep reinforcement learning, in: Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, BuildSys '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 356–357. doi:10.1145/3360322.3360998.
URL https://doi.org/10.1145/3360322.3360998
[84] N. Pflugradt, Load Profile Generator (2016).
URL https://www.loadprofilegenerator.de/
[85] G. Hoogsteen, A. Molderink, J. L. Hurink, G. J. Smit, Generation of flexible domestic load profiles to evaluate Demand Side Management approaches, 2016 IEEE International Energy Conference, ENERGYCON 2016doi:10.1109/ENERGYCON.2016.7513873.
[86] C. Sandels, J. Widén, L. Nordström, Forecasting household consumer electricity load profiles with a combined physical and behavioral approach, Applied Energy 131 (2014) 267–278. doi:10.1016/j.apenergy.2014.06.048.
URL http://dx.doi.org/10.1016/j.apenergy.2014.06.048
[87] C. Sandels, D. Brodén, L. Nordström, E. Andersson, Office Load Model: A MATLAB Application.
URL https://www.mathworks.com/matlabcentral/fileexchange/63376-office-load-electricity
[88] Reiner Lemoine Institut, Offgridders (2019).
URL https://github.com/rl-institut/offgridders
[89] M. Waskom, Seaborn - Python library for Data Visualization.
URL https://seaborn.pydata.org/
[90] Dash Plotly - Interactive Dashboard Python library.
URL https://plotly.com/dash/
[91] Streamlit - Python library for web-app creation.
URL https://docs.streamlit.io/
[92] M. Löning, A. Bagnall, S. Ganesh, V. Kazakov, J. Lines, F. J. Király, sktime: A unified interface for machine learning with time series, arXiv preprint arXiv:1909.07872.
[93] M. Ali, PyCaret: An open source, low-code machine learning library in Python, pyCaret version 2.2 (April 2020).
URL https://www.pycaret.org
[94] W. Hart, J.-P. Watson, D. Woodruff, J.-P. Watson, Pyomo: Modeling and solving mathematical programs in python, Mathematical Programming Computation 3 (2011) 219–260. doi:10.1007/s12532-

011-0026-8.
[95] S. Mitchell, M. O'Sullivan, I. Dunning, PuLP: A Linear Programming Toolkit for Python, Department of Engineering Science, The University of Auckland.
URL `http://www.optimization-online.org/DB_FILE/2011/09/3178.pdf`
[96] F.-A. Fortin, F.-M. De Rainville, M.-A. Gardner, M. Parizeau, C. Gagné, Deap: Evolutionary algorithms made easy, Journal of Machine Learning Research 13 (70) (2012) 2171–2175.
[97] J. Rapin, O. Teytaud, Nevergrad-a gradient-free optimization platform, version 0.2. 0, https://GitHub.com/FacebookResearch/Nevergrad.
[98] Y. Fujita, T. Kataoka, P. Nagarajan, T. Ishikawa, Chainerrl: A deep reinforcement learning library, arXiv preprint arXiv:1912.03905.
[99] M. Plappert, Keras-rl, GitHub Repository.
[100] M. Schaarschmidt, A. Kuhnle, K. Fricke, Tensorforce: A tensorflow library for applied reinforcement learning, Web page.
[101] Python Standard Libraries, hashlib Python Library.
URL `https://docs.python.org/3/library/hashlib.html`