

Article

Data Model for Residential and Commercial Buildings. Load Flexibility Assessment in Smart Cities

Simona-Vasilica Oprea , Adela Bâra , Răzvan Cristian Marales and Margareta-Stela Florescu

Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, Romana Square 6, 010374 Bucharest, Romania; bara.adela@ie.ase.ro (A.B.); razvan.marales@gmail.com (R.C.M.); margareta.florescu@ari.ase.ro (M.-S.F.)

* Correspondence: simona.oprea@csie.ase.ro; Tel.: +40-752-29-4422

Abstract: Demand response (DR) programs were usually designed to provide load peak reduction and flatten the load curve, but in the context of rapid adoption of emerging technologies, such as smart metering and sensors, load flexibility will address current trends and challenges (such as grid modernization, demand, and renewables growth) encountered by the evolving power systems. The uncertainty of the renewable energy sources (RES) and electric vehicle (EV) fleet operation has increased the importance of load flexibility that can be managed to provide more support for the stable operation of power systems, including balancing. In this paper, we propose a data model to handle load flexibility and take advantage of its benefits. We also develop a methodology to collect and organize data, combining the consumption profile with several auxiliary datasets such as climate characteristics of the location, independent system operator (ISO) to which the consumer is affiliated, geographical coordinates, assessed flexibility coefficients, tariff rates, weather forecast for day-ahead flexibility forecast, DR-enabling technology costs, and DR programs. These multiple features are stored into a flexibility relational database and NoSQL database for large consumption data collections. Then, we propose a data processing flow to obtain valuable insights from numerous .csv files and an algorithm to assess the load flexibility using large residential and commercial profile datasets from the USA, estimating plausible values of the flexibility provided by two categories of consumers.

Keywords: large datasets management; load flexibility database; DR programs; DR-enabling technology costs



Citation: Oprea, S.-V.; Bâra, A.; Marales, R.C.; Florescu, M.-S. Data Model for Residential and Commercial Buildings. Load Flexibility Assessment in Smart Cities. *Sustainability* **2021**, *13*, 1736. <https://doi.org/10.3390/su13041736>

Academic Editor: Luisa F. Cabeza

Received: 9 December 2020

Accepted: 27 January 2021

Published: 5 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Literature Review

According to [1], it is estimated that by 2030, load flexibility will avoid new generation capacity (57%), lower energy costs by shifting the operation of controllable appliances from peak to off-peak hours (29%), allow new transmission and distribution capacity (12%), and provide frequency regulation of ancillary services regulation (2%). Various demand response (DR) programs have been studied [2,3] as well as methods to assess the load flexibility of buildings [4–6]. In [2], the flexibility was measured for a three-year period using 186 residential consumers' data with a large set of smart appliances from Belgium. The purpose of this study was to identify the impact and performance of demand response (DR) programs. Another residential consumption area was investigated in the Netherlands from the DR point of view. The responsiveness of residential demand to signal tariffs using home energy management systems that shift the flexible load from evening to midday hours and consume energy from local generation was proposed in [3].

A useful classification of demand response services into three main categories (“shed”—reduce load, “shift”—reschedule, and “shimmy”—fast dispatch) is performed in [7], which also provides a comprehensive study on DR-enabling technology costs for control and communication. For residential consumers, flexible appliances, such as heating, ventilation, and air conditioning (HVAC); water heaters; pool pumps; and battery storage, including

electric vehicles (EVs), were proposed to provide DR through direct load control (DLC), automated demand response (ADR—involving communication, measurement, control infrastructure, and a signal that triggers a specific response) programs and programmable communicating thermostats (PCTs). Most of these appliances can provide DR services except for HVAC, which was only investigated to shed and shift with PCTs.

The demand flexibility potential of seven northern European countries with high shares of renewable energy sources was reviewed and estimated in [8]. The study identified and compared the flexibility potential by analyzing the methods used for flexibility potential estimation. The flexibility was assessed to be 15–29% of the peak load. However, the study did not show the value of the flexibility potential or integrate other data sources to provide added value.

The demand potential of HVAC of 12-storey residential buildings in Nordic countries to sustain renewable energy sources (RES) integration was investigated in [9] by taking into account the indoor climate conditions and comfort of the residents. The flexibility potential of HVAC was analyzed to provide ancillary services. The results underline the necessity of aggregating the loads and offering ancillary services, and they lead to a load reduction of 1.57 MW for a Danish region.

Another interesting study investigated and calculated the DR potential in Germany to support RES integration using an electricity market model. Its results showed that DR diminished the curtailments, emphasizing that the flexibility potential varied between DR programs [10]. This study strengthens the idea of implementing various DR programs or combinations of programs to obtain the best results. In addition, the flexibility load was assessed for buildings by considering the potential of heat pumps [11].

An estimation of DR potential using a fitted regression model was performed for residential and commercial buildings with EnergyPlus (a building simulation program to model energy consumption) datasets using two-state models for appliances, with thermostats obtaining an accuracy of 80–90% [12]. The main drawback of this study was that it used only a limited dataset of load. The big data feature and multiple data sources correlated with load data were not considered in the DR prediction. In addition, a quantification of load flexibility potential using cooling and heating systems of office buildings was analyzed in [13]. Several big data [14,15] and business intelligence techniques [16,17] for decision makers were engaged to find patterns in large sets of consumption data, and extensive reviews were performed into smart metering data analytics [18].

The experts in flexible demand from Nordic countries mentioned that the savings are as yet uncertain. They also underline the importance of taking further steps to properly design real-time prices, reconsider revenue regulation of the grid operators to incentivize DR implementation, and envision aggregation services [19].

In this paper, we corroborated the findings of previous studies [1,7] regarding flexibility coefficients estimated at the state level and DR-enabling technology costs by using large load profile datasets. We organized data to obtain useful insights regarding flexibility potential for certain geographical areas. We propose a data model to organize data in order to answer the following two questions: What is the potential size of the load flexibility resources? What are the savings that residential or commercial consumers can obtain? To achieve these goals, the paper is divided into five sections. In Section 1, the most relevant research studies are presented, in the Section 2, the data processing methodology is defined. The data model is built in Section 3. Simulations and results are provided in Section 4, and in Section 5 the conclusions of this study are drawn.

2. Data Processing Methodology

The load profile datasets were downloaded from the OpenEI website that provides a free data source to facilitate research in energy [20]. Load profile datasets for residential consumers and commercial buildings were provided by the Department of Energy in the USA, representing reference houses by location and targeting the energy efficiency for residential and commercial buildings [21].

From a large set of files, 273 load profiles for residential consumers located in various locations of the USA and storing hourly consumption data in kWh with the following attributes were extracted:

- Date/Time
- Electricity: Facility
- Gas: Facility
- Heating: Electricity
- Heating: Gas
- Cooling: Electricity
- HVAC Fans: Electricity
- General, Interior Lights: Electricity
- General, Exterior Lights: Electricity
- Appl., Interior Equipment: Electricity
- Misc., Interior Equipment: Electricity
- Water Heater and Water Systems: Gas

They were classified into gas and electricity consumption. Some of the loads are not controllable whereas some of them can be included into DR programs. The consumption was split into electricity and gas in total and with details for heating, cooling, HVAC, interior and exterior lights, and interior and other appliances. Gas provided heating for the houses and water. Thus, the volume of data for residential consumers was over 2.07 million records. However, the volume of data could expand to hundreds of thousands if the individual consumers were considered. In addition, for commercial buildings, the record count was more than 131 million records.

The datasets were archived as zip files containing large .csv files [20]. For the residential data, 273 .csv files were inside one large archive. For the commercial data, 10 large archives each contained 94 folders (except the last archive that had 90 folders), and each folder contained 16 .csv files and 8760 records, with a total of 14,976 .csv files of commercial data. Considering the size and dimensionality of the data, Mongo DB collections with an aggregation pipeline and Python were used to import, process, and calculate the residential data. Whereas, for the commercial data, the overview looked totally different. Analyzing the folders, files, and data structure, some interesting facts were noticed:

- Only .csv files were analyzed as input datasets, so the data presentation was standard;
- All the .csv files had the same fields/columns structure (Date/Time, Electricity: Facility, Fans: Electricity, Cooling: Electricity, Heating: Electricity, InteriorLights: Electricity, InteriorEquipment: Electricity, Gas: Facility, Heating: Gas, InteriorEquipment: Gas, Water Heater: WaterSystems: Gas);
- Each file name had a pattern which could give us more information about the dataset, as an example for RefBldgLargeHotelNew2004_v1.3_7.1_3A_USA_GA_ATLANTA.csv, the pattern is Ref [LocationType][Year]_v1.3_7.1_3A_[Country]_[State_ISO]_[City].csv. Location type, year, country, state independent system operator (ISO) and city are vital information extracted them for each file;
- The structure of the file path was easy to scan after unzipping each archive. An example is COMMERCIAL_LOAD_DATA_E_PLUS_OUTPUT:part9\USA_TX_Alice.Intl.AP.722517_TMY3\RefBldgStripMallNew2004_v1.3_7.1_2A_USA_TX_HOUSTON.csv having the following pattern [ARCHIVE_NAME]\[ROOT_FOLDER]\[FILE_NAME].csv.

These facts lead to the conclusion that an import of data into another tool for data processing and calculation could be easily automated. For the automation, Python script is used a. Next step was to choose a tool for data processing and calculation. Our criteria for choosing the database were:

- Scalable with a big data load;
- Familiar and easy to use;
- Can run on local machines;
- Easy to setup;

- Open Source.

Therefore, five databases, both SQL and NoSQL, are taken into consideration. In Table 1, the comparison is shown:

Table 1. Databases comparison based on 5 criteria.

	Scalable with a Big Data Load	Familiar and Easy to Use	Can Run on Local Machines	Easy to Setup	Open Source
PostgreSQL	Yes	Yes	Yes	Yes	Yes
MongoDB	Yes	Yes	Yes	Yes	Yes
Cassandra DB	Yes	Partial	Yes	Partial	Yes
Elasticsearch	Yes	Yes	Yes	Partial	Yes
HBase (Hadoop)	Yes	Partial	Yes	Partial	Yes

Based on the table of comparison, we decided to use PostgreSQL, as we are mostly familiar with SQL technology and this kind of databases. However, for handling the two large datasets that come from residential and commercial consumers, the top open source NoSQL and relational databases at the moment were used [22]. Thus, each file from the datasets was scanned and each row from the .csv file was inserted into a row in PostgreSQL. Additionally, location, city and state were extracted from the file name and added to the consumption data. For the beginning, all the rows were inserted into just one table, after that the relations were normalized. To understand the correlation and data source, a synthesis is provided in Table 2.

Table 2. Columns description.

Column Name Database	Data Type Database	Source	Example	Observation
date_time	Timestamp	.csv file	11 July 2004 21:00:00	
Id	Serial			Unique identifier of row, auto generated
electricity_facility_kwhourly	Numeric	.csv file	10.8926491824	
fans_electricity_kwhourly	Numeric	.csv file	32.0021175342	
cooling_electricity_kwhourly	Numeric	.csv file	15.1069992076	
heating_electricity_kwhourly	Numeric	.csv file	5.48122459637	
gas_facility_kwmonthly	Numeric	.csv file	1.8243253836	
interiorlights_electricity_kwhourly	Numeric	.csv file	1.0115996781	
interiorequipment_electricity_kwhourly	Numeric	.csv file	4.04639871239	
gas_facility_kwhourly	Numeric	.csv file	7.29730153440	
heating_gas_kwhourly	Numeric	.csv file	16.4436737891	
electricity_facility_kwmonthly	Numeric	.csv file	24.9203635340	
interiorequipment_gas_kwhourly	Numeric	.csv file	5.48122459637	
water_heater_watersystems_gas_kwhourly	Numeric	.csv file	0.5500053624	
location_type	Text	.csv file name	RefBldgSmallOffice	Extracted from the file name
year	Numeric	.csv file name	2004	Extracted from the file name
state_code	Text	.csv file name	GA	Extracted from the file name

Table 2. Cont.

Column Name Database	Data Type Database	Source	Example	Observation
city	Text	.csv file name	ATLANTA	Extracted from the file name
filename	Text	.csv file name	RefBldgWarehouse-New2004_v1.3_7.1_3A_USA_GA_ATLANTA.csv	Used for investigations
filepath	Text	.csv file path	D:\stats\Data\COM-MERCIAL_LOAD_DATA_E_PLUS_OUTPUT.part1\USA_AL_Troy.Air.Field.722267_TMY3\RefBldg WarehouseNew2004_v1.3_7.1_3A_USA_GA_ATLANTA.csv	Used for investigations, in case we found some irrelevant data we can go directly to file
hour	Numeric	data_time field	23	Used to optimize the queries

When the data was inserted into the database, the insertion time was too long, only inserting 1,007,400 rows took 5270.9 seconds, having in mind that the total target of rows to be inserted is 131,189,412, the insertion time was not acceptable in this state. So, we decided to optimize the insertion by doing two things. Firstly, the insert statement from a simple insert to a prepared statement was changed, so the statement is already processed by the database; and secondly, instead of single row statement, batch statement was used. For the batch statement, some experiments were performed to see which volume per statement is more suitable and performant in our case. Table 3 shows the results of this experiment.

Table 3. Batch statement experiment times.

Number of Inserted Rows per Batch	Total Duration Seconds	Total Rows Inserted
1	5270.980333328247	1,007,400
100	538.0430161952972	1,007,400
200	435.40850925445557	1,007,400
300	382.5205578804016	1,007,400
350	381.44075322151184	1,007,400
400	367.4466257095337	1,007,400
500	371.1627984046936	1,007,400
1000	412.53081250190735	1,007,400

The best insertion time for batch was 367.4466257095337 having the batch size of 400 rows. Keep in mind that this may be different depending on the resources available on the machine (CPU, RAM memory, storage type) and the insertion statement.

We are aware that the insertion procedure can be further optimized, implementing mechanisms such as “in file” import directly into database, parallelism, or resource boosting. In the end, the total insertion of 131,189,412 rows took 47,958.8661942482 s or 13 h and 18 min.

In Figure 1, the algorithm implemented in the Python script for import written in pseudocode is provided. The algorithm steps are the following:

1. The script starts scanning the target path for unarchived data folders (prerequisite step was to download all the archived data and unarchived it to a specific path).
2. The script finds an unarchived folder and scans it inside for a root data folder.
3. The script finds a root data folder and starts to iterate on each file. At this point, the files are in .csv format.

4. Each file is opened, and the columns of .csv format are validated. In case a column is missing, or an unexpected column appears the script will log an error and close the process. In this way, an inconsistency can easily be identified and investigated.
5. If the columns are valid the script will iterate through the rows. The columns of file will be combined with the row values and each row data will be added to the “rows to be inserted” list.
6. If the number of items in list is equals to the number of batch size (to be inserted), the list will be transformed into a big batch statement and ran against the PostgreSQL database.
7. If the end of file is reached and the number of rows to be inserted is not equals to the number of batch size, the insert will be done, so it can be continued with another file.
8. The next file will be processed.
9. If the root folder has no more file, the root folder will be changed.
10. If the unarchived folder has no more root folders, the unarchived folder will be changed.
11. The program will end when there will be no more unarchived folders to be processed.

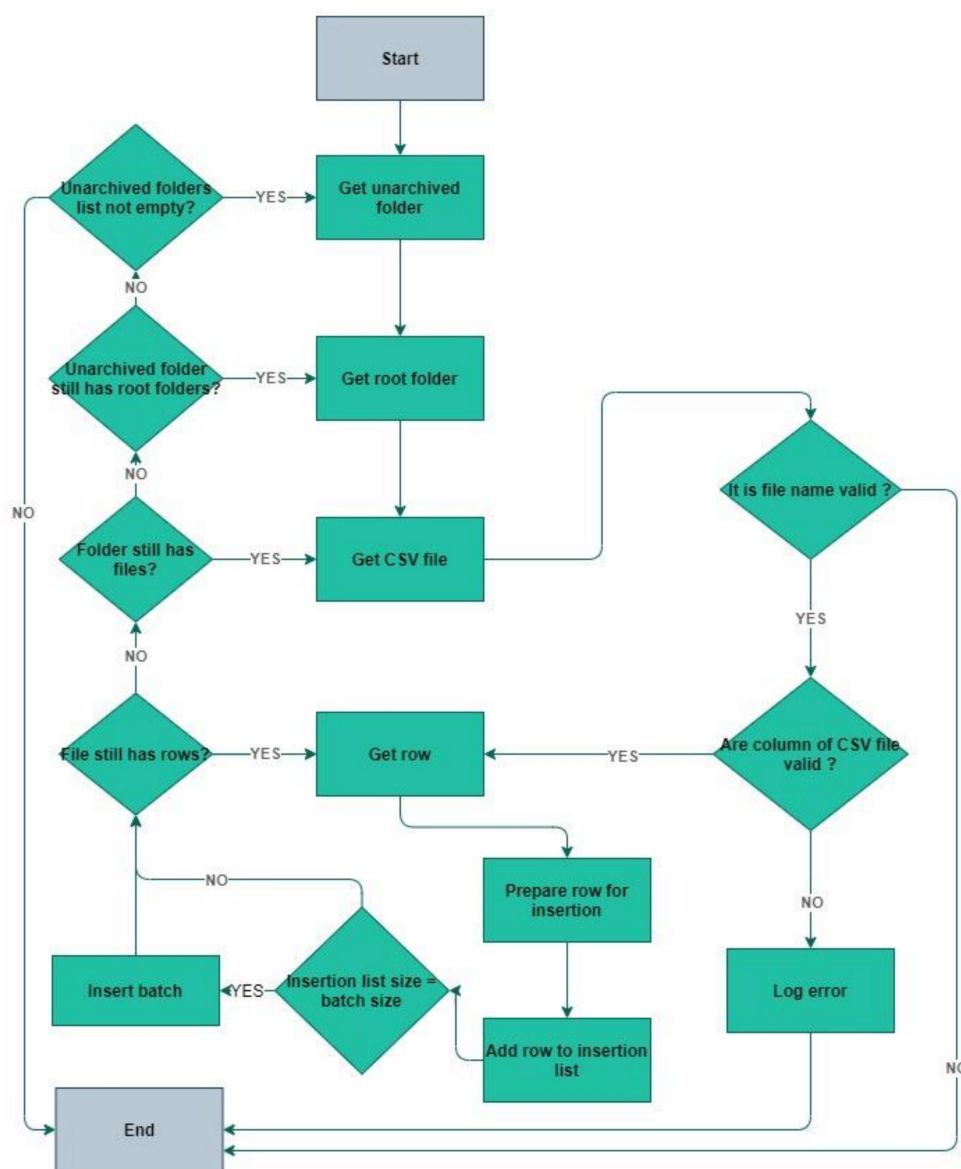


Figure 1. Python script algorithm schema.

After the data was fully inserted into PostgreSQL database, further investigation to reduce the data is performed as it was still big and the queries were not performant enough, even by adding indexes on some common columns and aggregation query could took over several hours.

Then, the table of consumption is combined with another tables which contains data about state independent system operator (ISO) code together with the ISO code of region, flexibility coefficients, DR enabling technology costs. After this join, the consumption for each hour, region ISO code and location type is analyzed.

The very large volume of consumption datasets especially for commercial consumers falls into big data paradigm. Therefore, the dimensionality of the datasets is reduced as in Figure 2 to process the data and obtain useful insights. First, the peak hours are identified and then built the data model considering other datasets that are correlated with the consumption data. The data model represents the interaction of reduced consumption dataset with the flexibility potential, ISO affiliation, tariff rates, flexibility services, weather forecast, and DR enabling technology costs. The data model is implemented into a database for flexibility assessment and savings computation.

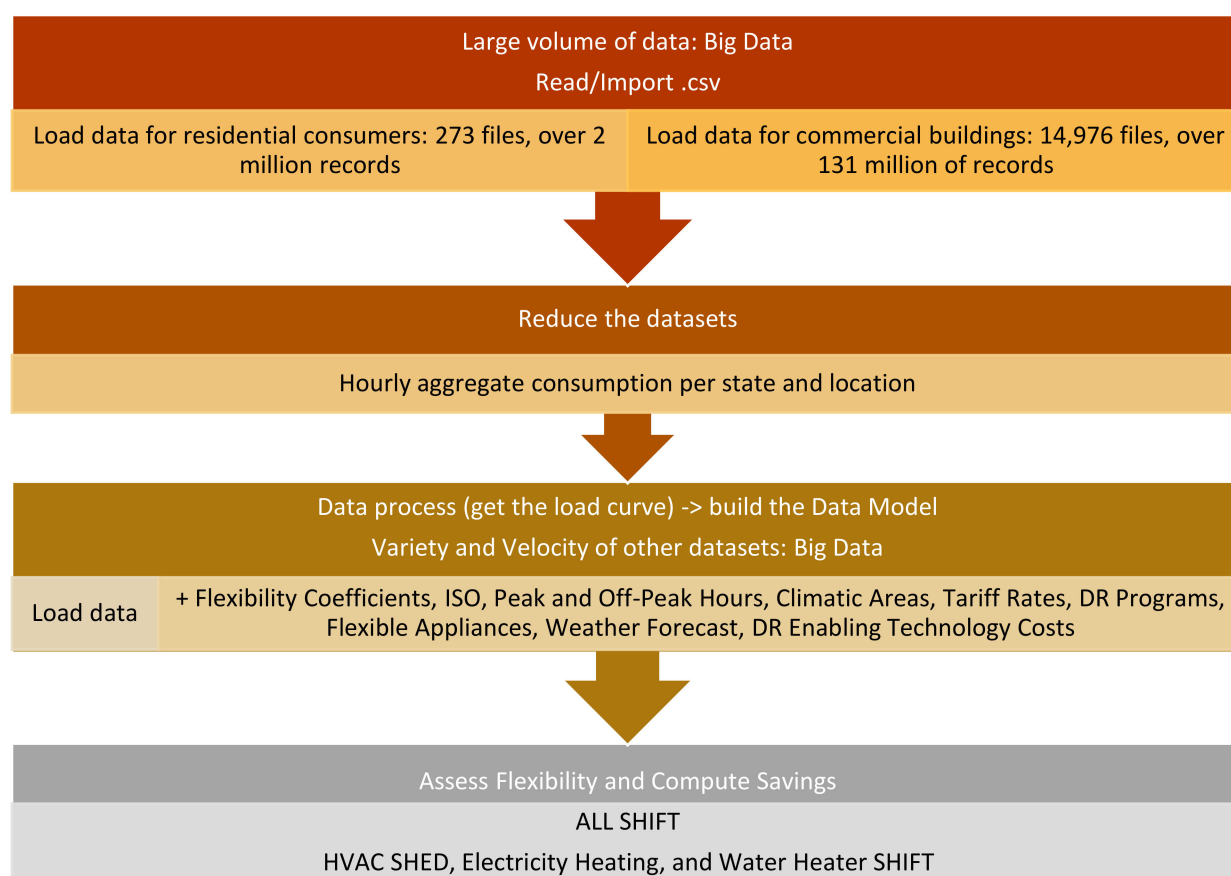


Figure 2. Data processing flow.

3. Building the Data Model and Algorithm for Load Flexibility Assessment

Identification of flexibility potential for residential and commercial buildings starts with the main datasets: the load curve for the two categories of consumers. The load curve represents the mean hourly load. Its shape allows us to identify the peak (h_{PEAK_START} , h_{PEAK_STOP}) and off-peak hours and perform further calculations to assess the flexibility and the savings.

The other components of the data model will be described in the following subsections.

3.1. DR Capability, Climatic Areas, and ISO

The regional differences could be significant in terms of challenges and load flexibility. Thus, the average estimation is not useful, and the consumers are grouped by ISO, states, and locations.

Five flexibility ranges are assessed in [1] and seven climatic areas in the USA are identified in [23] according to Figure 3. Additionally, the states and their affiliation to an ISO is shown in Figure 4.

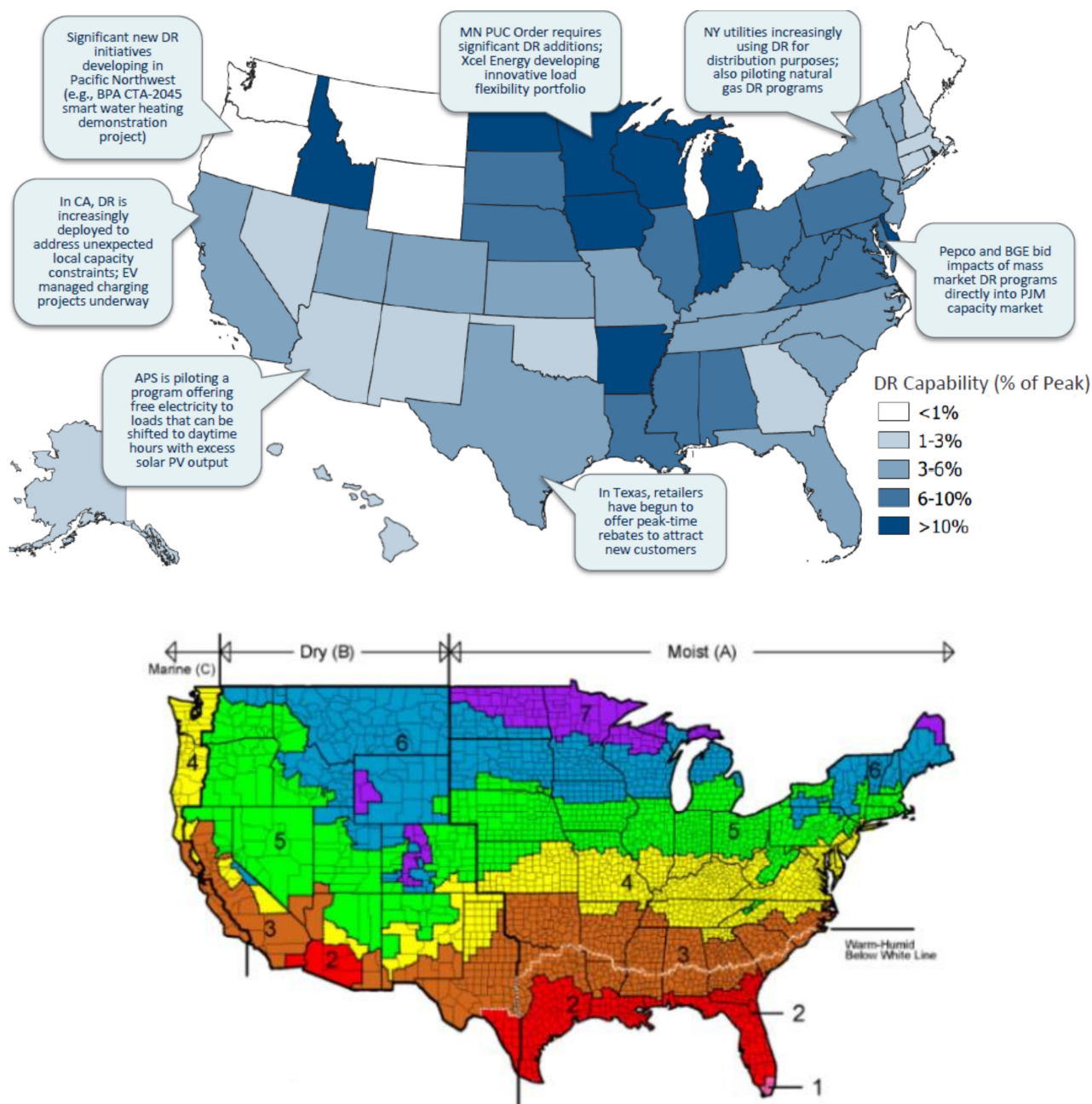


Figure 3. Five flexibility potential ranges [1] and 7 climatic areas in the USA [23].

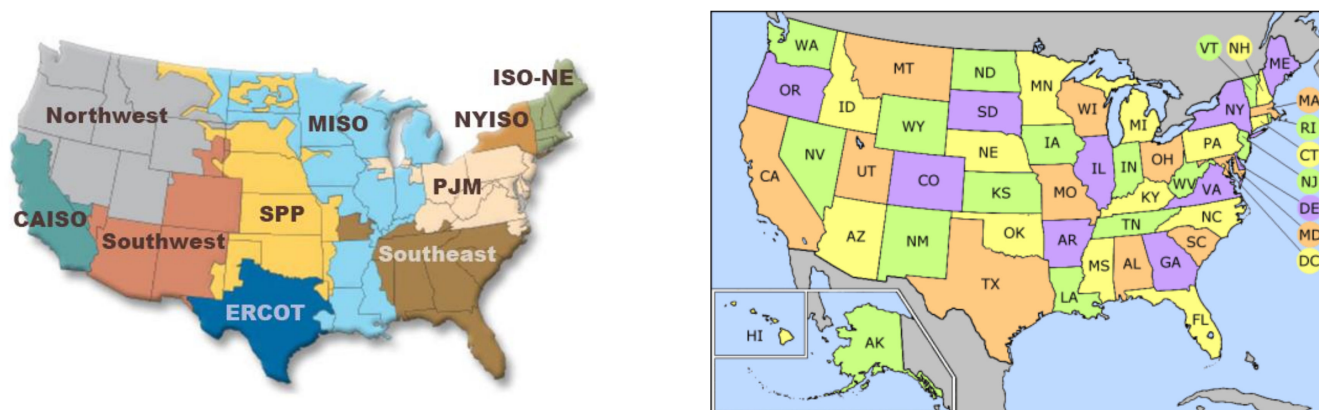


Figure 4. Ten independent system operators (ISOs) and the 52 states in the USA.

The identification of the entities that form the data model contains information that is presented in Figures 3 and 4: Loads, State_flex that embeds the flexibility coefficients; Climate_areas; ISOs; and Locations.

Correlating the information from Figures 3 and 4, the average flexibility coefficients are obtained for residential (a) and commercial consumers (b) at the ISO level as in Figure 5.

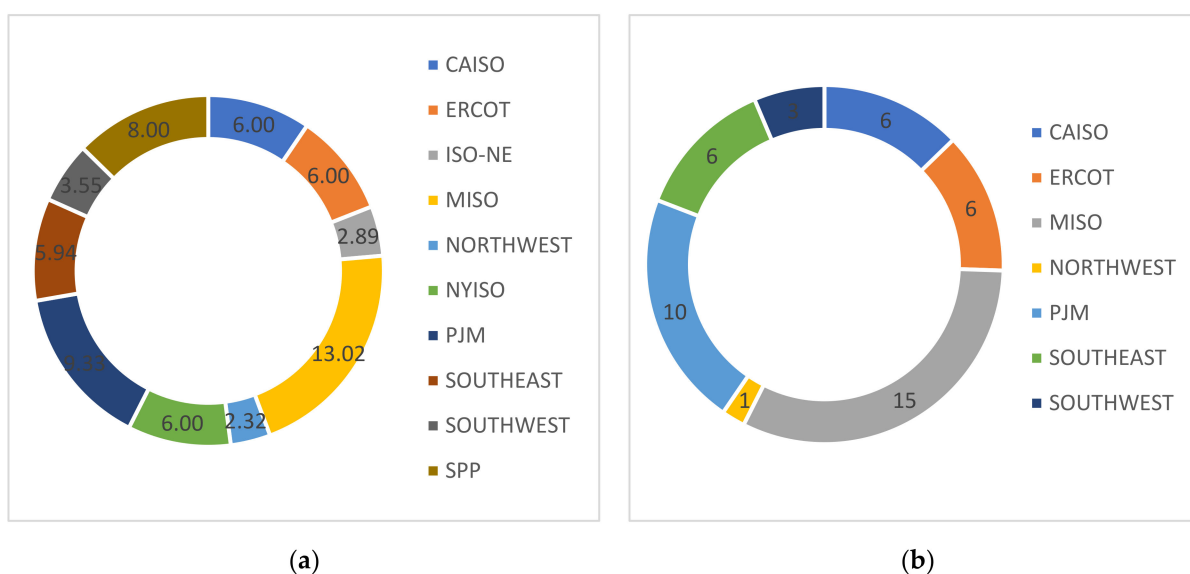


Figure 5. Flexibility coefficients for each ISO. (a) residential consumers; (b) commercial consumers.

3.2. DR Programs

Load flexibility potential is empowered by the emerging programs enabled primarily by smart meters, sensors technologies, smart thermostats, switching modules connecting/disconnecting and controlling the operation of the appliances and local generation and storage. Some of the programs such as DLC should be just revitalized by data analytics showing the benefits of their implementation at large scale. DR programs are briefly described in Table 4.

3.3. DR Enabling Technology Costs

The DR enabling technology costs were assessed for both residential and commercial consumers [7] as in Tables 5 and 6.

Table 4. Demand response (DR) programs.

No.	DR Program	Description
1	ADR or DLC	Control of customers' flexible appliances for DR purposes using an automated signal that triggers a specific response.
2	Smart-thermostats	Temperature is remotely controlled to reduce heating/cooling appliances usage at peak.
3	Discounted rates	Residential consumers reduce consumption to a specific level and get a discounted rate.
4	Bid load	Residential consumers bid the day-ahead curtailment program at 15 min-resolution. If their orders are executed, they must curtail and receive an additional payment. Otherwise, they will encounter a penalty that is usually mentioned in an agreement.
5	Signal tariffs	Static or dynamic signal rates that encourage the consumption at off-peak intervals. These tariffs can be designed as Time-of-Use (ToU) tariffs, critical peak, variable peak, real-time pricing, or dynamic tariff rates set by consumption optimization with game theory.
6	Consumers' awareness	Residential consumers are aware of the load reductions requirements and behave accordingly without a financial incentive. Such programs are tailored to show the advantages and benefits of a certain behavior towards a sustainable consumption.
7	Charging batteries	Residential consumers are stimulated to charge batteries including EV during off-peak interval.
8	Thermal storage	Boilers and other similar appliances operate at off-peak hours. The boiler tanks preserve the water temperature. For commercial buildings, it can be extended to the ice processing. Thus, the water can be frozen at off-peak hours and provide cooling at peak hours.
9	Smart Adaptive Switching Module (SASM)	It automates the control of various appliances, including generation and storage. It is based on a priority of the appliances and control of the load with fuzzy rules.
10	Dispatchable load	The program requires more engagement from consumers.

Table 5. DR enabling technology costs for residential consumers.

No.	Action	DR Program	Flexible App.	Control Tech. and Comm. Cost (\$/End-Use)
1	Shed	DLC	HVAC	166
			Pool pump	147
			Room AC	75
2	Shed and Shift	Smart-thermostat	HVAC	279
3	Shimmy	ADR Control	Water Heater	2136
		DLC	Water Heater	350

Table 6. DR enabling technology costs for commercial consumers.

No.	Action	DR Program	Flexible App.	Control Tech. Cost (\$/kW)	Comm. & Hardware Cost (\$/Site)
1	Shed	DLC	HVAC	62	107
2	Shed and Shift	ADR Control	HVAC	242	-
			Refrigerated Warehouse	289	-
		Smart-thermostat	HVAC	171	-
3	Shimmy	ADR Control	HVAC	310	2066
			Refrigerated Warehouse	289	2066
			Water Heater	166	1000

3.4. Weather Data and Flexibility Forecast

In order to perform the day-ahead flexibility forecast, the meteorological datasets could be extracted from the weather websites. The data is stored in .csv files and then imported in the database. Even if it is an important component of the load flexibility management, it is not an objective of this paper.

The steps required to extract weather data from websites, as in Figure 6, are presented in Table 7.

Website format

	Conditions			Comfort			
Time		Temp	Weather	Wind		Humidity	Barometer
00:50 Tue, 1 Jan		9 °C	Passing clouds.	19 km/h	↗	87%	999 mbar
01:50		8 °C	Passing clouds.	15 km/h	→	93%	1000 mbar
02:50		7 °C	Passing clouds.	15 km/h	↘	81%	1000 mbar
03:50		7 °C	Passing clouds.	17 km/h	→	81%	1001 mbar
04:50		6 °C	Clear.	17 km/h	↘	81%	1002 mbar
05:50		5 °C	Clear.	13 km/h	→	81%	1003 mbar
06:50		4 °C	Passing clouds.	17 km/h	→	87%	1003 mbar
07:50		4 °C	Clear.	17 km/h	→	81%	1004 mbar
08:50		3 °C	Passing clouds.	11 km/h	↗	93%	1006 mbar
09:50		5 °C	Passing clouds.	15 km/h	→	81%	1007 mbar
10:50		6 °C	Passing clouds.	20 km/h	→	76%	1007 mbar
11:50		7 °C	Passing clouds.	20 km/h	→	71%	1008 mbar
12:50		7 °C	Passing clouds.	24 km/h	↘	61%	1008 mbar

Csv format

Figure 6. Cont.

	A	B	C	D	E	F
1	Time	Temp	Weather	Wind	Humidity	Barometer
2	00:50	9°C	Overcast.	6 km/h	82%	1024 mbar
3	01:50	7°C	Passing clc	2 km/h	87%	1024 mbar
4	02:50	6°C	Clear.	No wind	93%	1024 mbar
5	03:50	4°C	Passing clc	2 km/h	93%	1024 mbar
6	04:50	3°C	Passing clc	2 km/h	93%	1024 mbar
7	05:50	3°C	Fog.	9 km/h	93%	1024 mbar
8	06:50	5°C	Fog.	4 km/h	93%	1025 mbar
9	07:50	4°C	Fog.	6 km/h	100%	1026 mbar
10	08:50	5°C	Fog.	4 km/h	93%	1026 mbar
11	09:50	6°C	Fog.	6 km/h	93%	1026 mbar
12	10:50	7°C	Fog.	6 km/h	93%	1027 mbar
13	11:50	10°C	Partly sunr	4 km/h	82%	1026 mbar
14	12:50	9°C	Partly sunr	4 km/h	82%	N/A
15	13:50	14°C	Scattered	2 km/h	63%	N/A
16	14:50	14°C	Sunny.	7 km/h	63%	N/A
17	15:50	15°C	Scattered	11 km/h	63%	1026 mbar
18	16:50	13°C	Sunny.	9 km/h	63%	1026 mbar
19	17:50	10°C	Clear.	11 km/h	71%	1026 mbar
20	18:50	9°C	Clear.	6 km/h	71%	1027 mbar
21	19:50	7°C	Clear.	4 km/h	87%	1027 mbar
22	20:50	7°C	Clear.	No wind	81%	1027 mbar
23	21:50	5°C	Clear.	4 km/h	87%	1027 mbar
24	22:50	4°C	Passing clc	4 km/h	93%	1028 mbar
25	23:50	3°C	Passing clc	6 km/h	93%	1028 mbar

Figure 6. Scraping the date from website to .csv file.

Table 7. Steps to extract weather data from the websites written in PHP.

No.	Step
1	//Using CURL PHP library to download data via HTTP //curl session initiation \$ch = curl_init();
2	//setting option with curl_setopt //CURLOPT_URL option to specify the URL curl_setopt (\$ch, CURLOPT_URL, "https://www.timeanddate.com/scripts/cityajax.php?n=fl/ states&mode=historic&hd=\{\$data\}&month=\{\$dataMonth\}&year=\{\$dataYear\}" (accessed on 2 February 2021)) //post request, 1-true curl_setopt (\$ch, CURLOPT_POST, 1); //CURLOPT_RETURNTRANSFER returns the content of the page curl_setopt (\$ch, CURLOPT_RETURNTRANSFER, true);
3	//execute session \$ch and display the result in browser \$server_output = curl_exec(\$ch);
4	//HtmlDomParser returns the content of the HTML page \$dom = HtmlDomParser::str_get_html (\$server_output);
5	//create and save the extracted data into a .csv file \$fp = fopen ("Weather data {\\$dataA}isare.csv", 'w');

3.5. Data Model

The relational data model reflects the above-mentioned components that can be correlated with load data to calculate flexibility potential and savings. Moreover, the flexibility can be daily estimated using weather forecast that is based on weather readings.

The most linked entity is *LOCATIONS* connecting 9 entities: *WEATHER_READINGS*, *TARIFF_RATES* (through *TARIFF_LOCATIONS*), *WEATHER_FORECAST*, *CLIMATE_AREAS*, *ISOS*, *STATE_FLEX*, *LOADS*, *LOAD_FORECAST*. The load forecast is stored at the level of an appliance that belongs to a specific location. *DR_PROGRAMS* and *DR_APP_COSTS* store data about programs described in Table 4 and DR enabling technology costs presented in Tables 5 and 6. In *WEATHER_READINGS* data is collected as in Table 7. Based on the data model, a procedure is written to implement algorithm presented in Table 8 to calculate the flexibility potential and savings of the residential or commercial consumers.

Table 8. Algorithm for flexibility potential and savings.

Notations: $h \in [h_{PEAK_START}, h_{PEAK_STOP}]$—hour; h_{PEAK_START}—peak start hour; h_{PEAK_STOP}—peak stop hour; W_f—flexible capacity; α_f—flexibility coefficient;	C_{EH}—consumption of electricity heating; C_{WH}—consumption of water heater; C_{HVAC}—consumption of HVAC; S—savings from flexibility usage; t_{PEAK} tariff rate at peak hours; $t_{OFFPEAK}$ tariff rate at off-peak hours.
Case 1: ALL SHIFT	Case 2: HVAC SHED, EH, WH SHIFT
Assess flexibility	Assess flexibility
IF $h \geq h_{PEAK_START}$ AND $h \leq h_{PEAK_STOP}$ THEN : $W_f = \alpha_f \times \sum_h (C_{EH} + C_{WH} + C_{HVAC})$; END IF;	IF $h \geq h_{PEAK_START}$ AND $h \leq h_{PEAK_STOP}$ THEN : $W_{f_SHIFT} = \alpha_f \times \sum_h (C_{EH} + C_{WH})$ $W_{f_SHED} = \alpha_f \times \sum_h C_{HVAC}$ END IF;
Compute savings	Compute savings
$S = \sum_h W_f \times (t_{PEAK} - t_{OFF-PEAK})$	$S_{SHIFT} = \sum_h W_{f_SHIFT} \times (t_{PEAK} - t_{OFFPEAK})$ $S_{SHED} = \sum_h W_{f_SHED} \times t_{PEAK}$ $S = S_{SHIFT} + S_{SHED}$

Starting from this model, useful insights are extracted using performant queries that release synthetic information related to load flexibility. The results achieved in this paper are based on the data model presented in Figure 7.

3.6. Algorithm for Load Flexibility Assessment

To assess the flexibility potential and compute the savings, we propose the algorithm described in Table 9. For calculation, two cases are simulated: CASE1 ALL SHIFT algorithm in which a small percent of the electric heating, water heater and HVAC consumptions are shifted from peak to off-peak hours; and CASE2 HVAC SHED, EH, WH SHIFT in which a small percentage of the HVAC consumption is shed, whereas also a small percentage of electric heating and water heater consumption is shifted. The percentages assessed as in [1] are converted to the flexibility coefficients. To estimate the savings, different rates for peak t_{PEAK} and off-peak hours $t_{OFFPEAK}$ are considered.

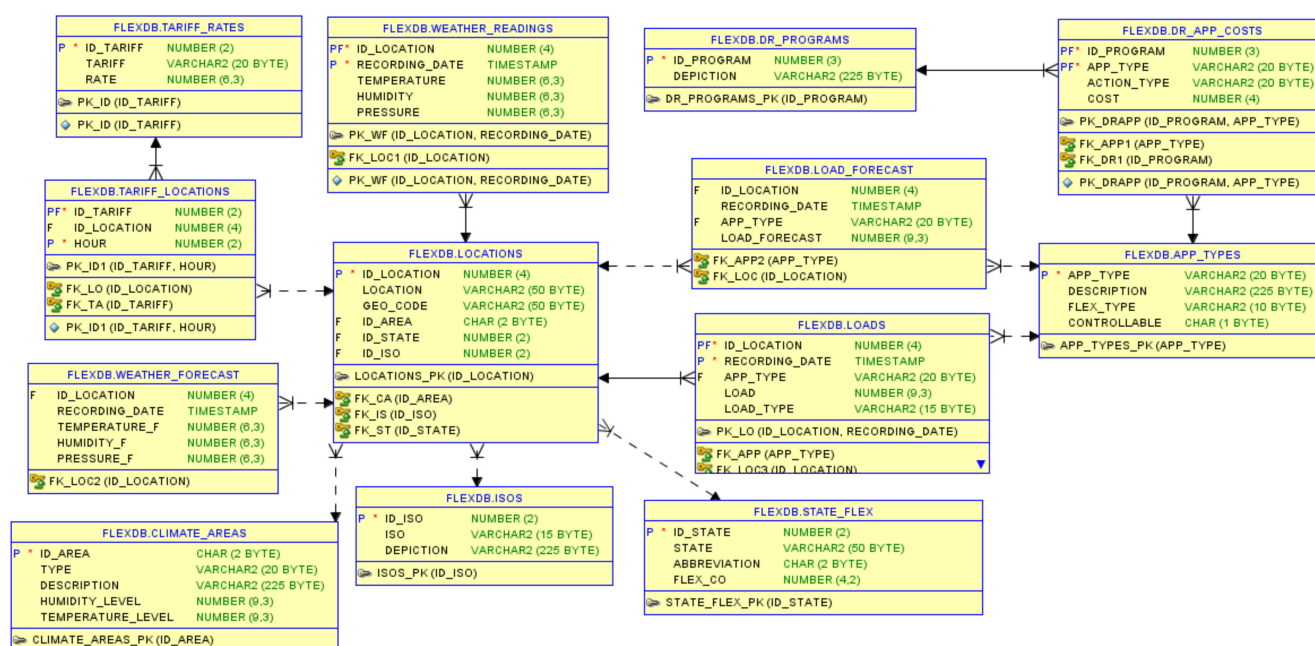


Figure 7. Data relational model.

Table 9. Savings and flexibility capacity of residential consumers.

	Shifted Load (kWh)	Shed Load (kWh)	Savings (Euro)
ALL SHIFT	12,434.42	0	3605.98
Total CASE1	12,434.42	0	3605.98
EH SHIFT	327.46	0	94.96
WH SHIFT	4025.12	0	1167.29
HVAC SHED	0.00	8081.84	3071.10
Total CASE2	4352.58	8081.84	4333.35

4. Simulations and Results

Apart from consumption data, in the simulations the datasets presented in the data model above were partially considered. As it is complex, the consumption forecast, DR enabling technology costs and climate area are eluded in the simulations that could be the subject of another study. Therefore, the flexibility forecast is not in the scope of this paper.

The hourly average load is aggregated and grouped by ISO to obtain the daily load curve (as in Figure 8a for residential and Figure 8b for commercial consumers). The commercial consumers are located in the control area of only 7 ISO. In both cases, we can notice that consumers that belong to SOUTHEAST and ERCOT ISO have a higher consumption than the other consumers. However, the average profile consumption for the two types of consumers varies very much in shape and amplitude.

As the datasets contain gas and electricity consumption, the total hourly consumption is compared, noticing that for residential consumers the gas consumption is always higher than electricity, especially at the morning hours when electricity consumption is lower (as in Figure 9a). However, at evening the gas and electricity are more balanced, therefore, it is assumed that gas is replaced by electricity. In case of commercial consumers (as in Figure 9b), the electricity always exceeds the gas consumption. Further, gas consumption peak records in the morning, whereas electricity consumption peak starts around 9 and decreases at 19. Therefore, different time intervals are taken into calculating the flexibility value.

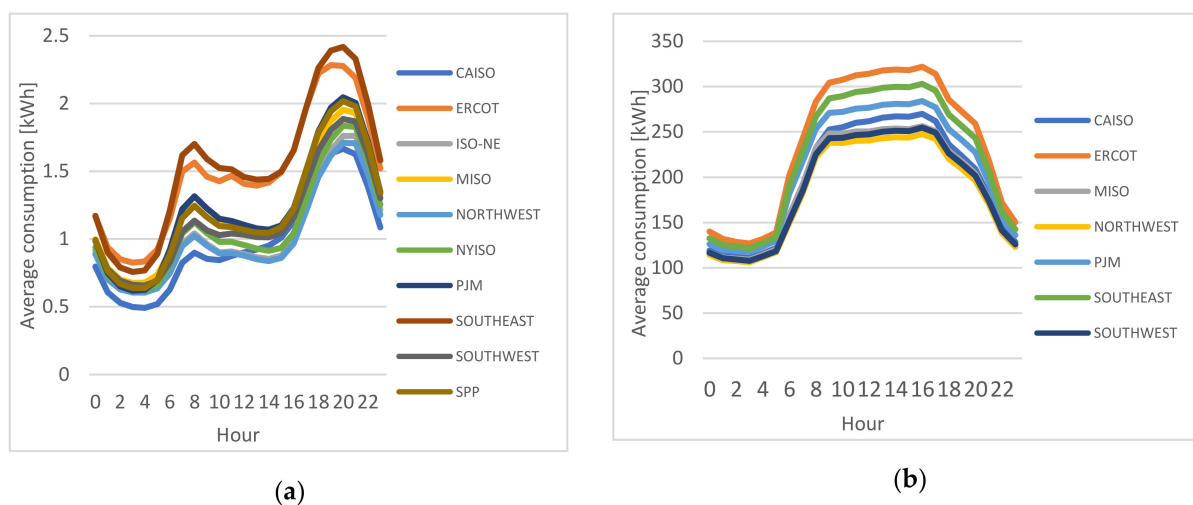


Figure 8. Daily load curves of the residential (a) commercial (b) consumers grouped by ISO.

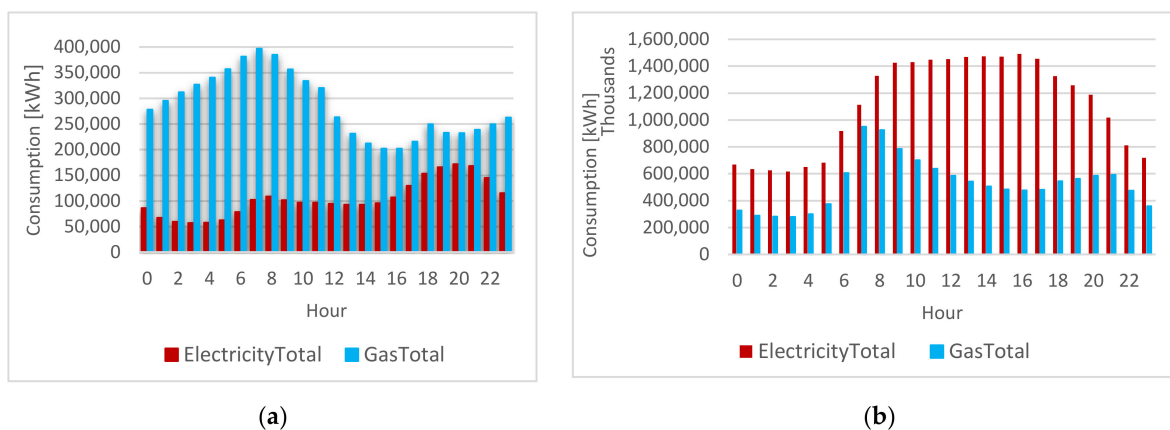


Figure 9. Electricity and gas curves of the residential consumers. (a) residential consumers; (b) commercial consumers.

The electricity daily load curve breakdown is presented in Figure 10a for residential and Figure 10b commercial consumers. Both consumers have three flexible appliances that are measured separately and may bring savings.

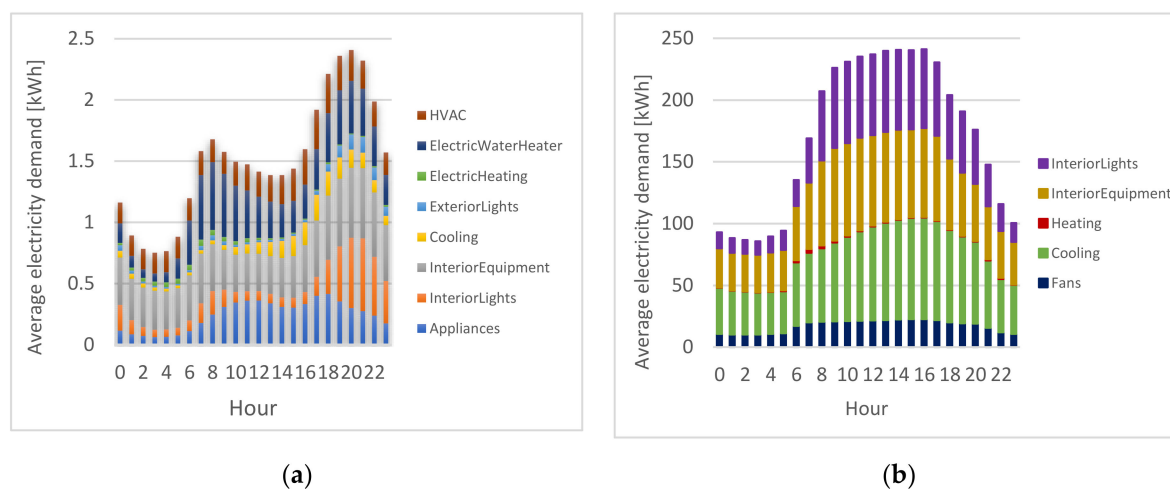


Figure 10. Daily average electricity load curve of the residential (a) commercial consumers (b) by components.

The sum of controllable and non-controllable appliances consumption is shown in Figure 11a for residential and Figure 11b for commercial consumers. The controllable appliances count for 26% of the total consumption of the residential consumers. The share of the controllable appliances is much bigger for commercial consumers showing an increased flexibility potential.

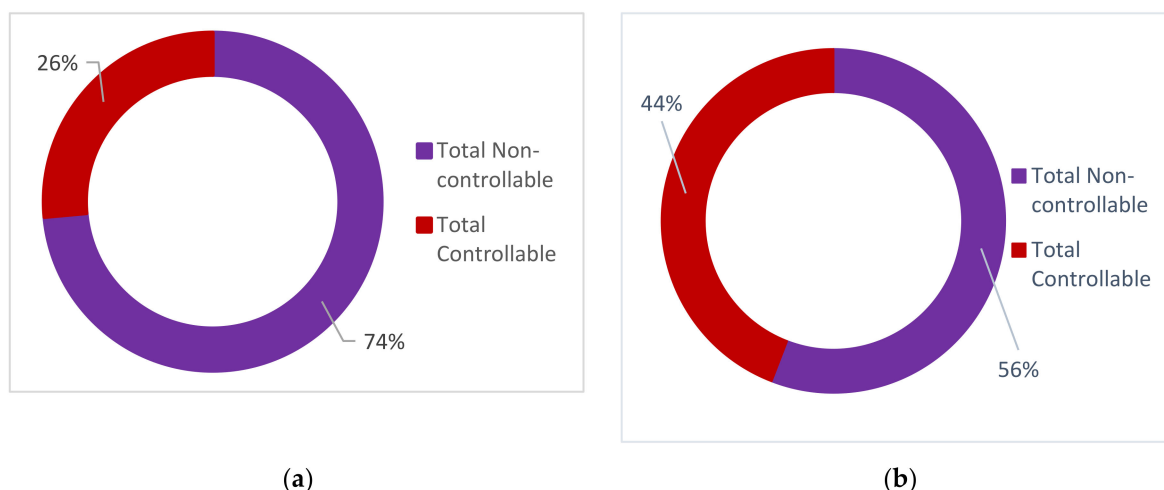


Figure 11. Controllable and non-controllable appliances share. (a) residential consumers; (b) commercial consumers.

The breakdown of the controllable appliances is presented in Figure 12a,b. Heating has the smallest share in both cases, whereas the HVAC and Cooling have the largest shares.

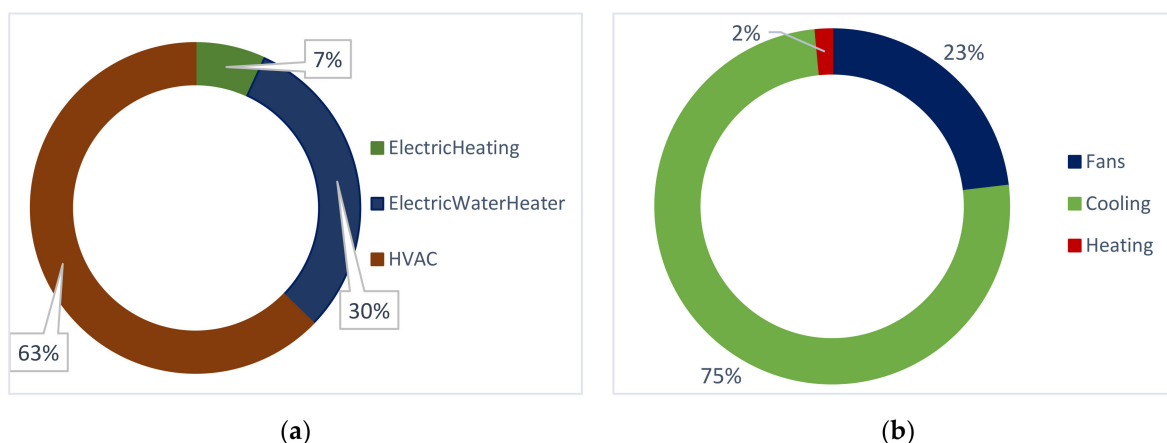


Figure 12. Electricity consumption breakdown by controllable appliances for residential (a) and commercial consumers (b).

The structure of the controllable appliances by ISO is shown in Figure 13. Most of the flexible consumers belong to SOUTHEAST, MISO, and ERCOT ISO for residential, respectively, MISO, SOUTHEAST, and PJM for commercial consumers.

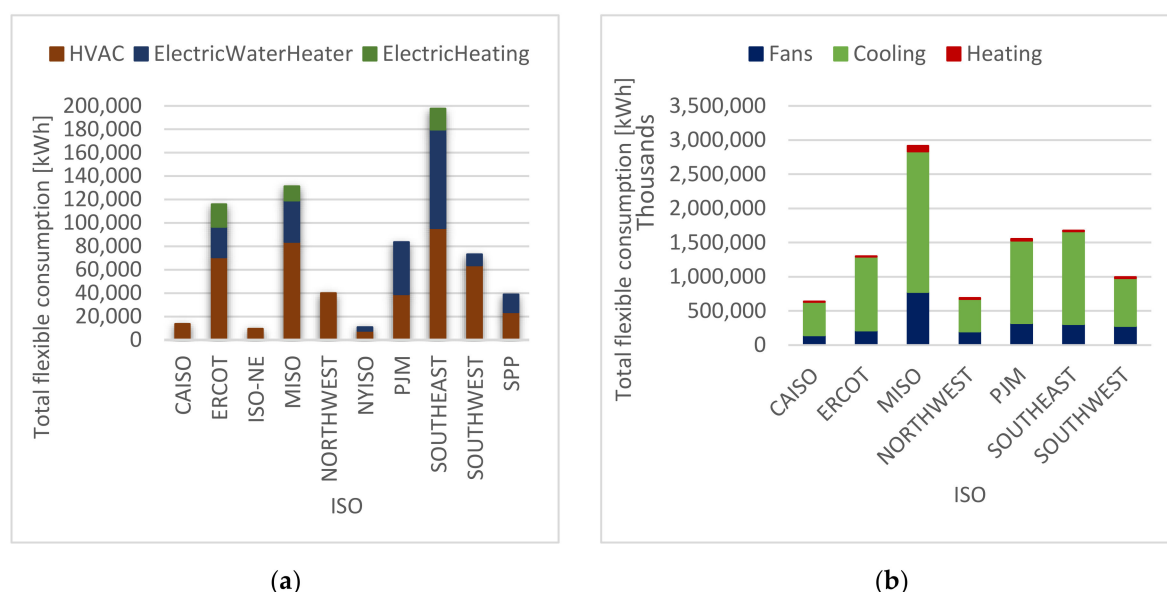


Figure 13. Flexible electricity consumption breakdown by ISO. (a) residential consumers; (b) commercial consumers.

Based on the proposed methodology and algorithm presented in Figure 2 and Table 5, the flexible capacity and savings are calculated. The results are shown in Table 8.

The annually savings are quite modest even in the second case considering the total number of residential consumers (273), the average value being almost 16 Euro. Therefore, the DR enabling technology costs were not taken into account from evident reasons. They could be supported by the aggregator (retailer) or grid operator. To be more efficient, the reward that result from the tariff rates difference ($t_{PEAK} - t_{OFFPEAK}$) should be accompanied by additional incentives. As peak rate and off-peak rates for residential consumers, $t_{PEAK} = 0.38$ Euro and $t_{OFFPEAK} = 0.09$ Euro, respectively $t_{PEAK} = 0.27$ Euro and $t_{OFFPEAK} = 0.09$ Euro are considered in calculation.

The flexibility coefficients are taken from a previous study and used as input data. They vary from less than 0.1 to 0.15 of the peak load. In case the flexibility coefficients are increased up to 0.15 of the peak, the results of this alternative scenario are shown in Figure 14.

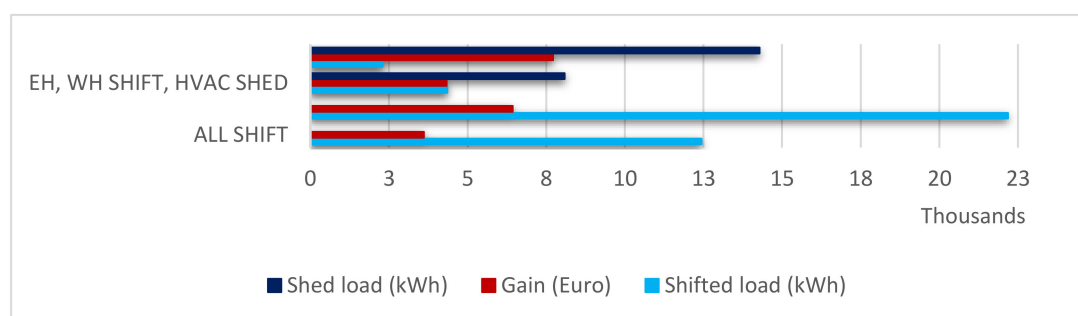


Figure 14. Flexible potential and savings for the assessed flexibility coefficients and 15-scenario.

For commercial consumers, the shifted power totalized 9.3 million MWh with savings of 88.1 million Euro. In this case, the savings are significant considering the total number of commercial consumers (14,976), the average value being almost 5900 Euro per consumer. However, the DR enabling costs are much higher for commercial consumers and their supportability can be shared or they can be on aggregator.

5. Conclusions

In this paper, we started from the findings of previous studies, regarding flexibility coefficients estimated at the state level and DR enabling technology costs and combined them with large load profile datasets to obtain useful insights regarding flexibility potential for certain geographical areas.

Usually, large datasets challenge researchers when it comes to achieving valuable insights from numerous data. Thus, 273 load profiles for residential consumers located in various locations of the U.S. storing hourly consumption data with the numerous attributes were extracted. The volume of data for residential consumers was over 2.07 million records. However, the volume of data could expand to hundreds of thousands if the individual consumers were considered. In addition, for commercial buildings, the records counted for more than 131 million records. To handle such volume of data, the files were automatically imported with a Python script, reduced and for data processing, storage, and calculation, two top databases (both SQL and NoSQL) are considered.

Hence, in this paper, we proposed a data model that handles large volumes of data and correlates load data with other datasets to assess or calculate the savings from flexibilities. The algorithm can be written as a stored procedure and applied separately for residential and commercial consumers considering different peak intervals, DR programs, costs, and tariff rates.

By implementing the proposed data flow, model, and algorithm, we obtained total savings of over 7500 Euro per year for residential consumers and 88.1 million Euro for commercial consumers. the annual savings are quite modest even in the second case considering the total number of residential consumers (273), the average value being almost 16 Euro. For commercial consumers, the shifted power totaled 9.3 million MWh with savings of 88.1 million Euro. In this case, the savings are significant considering the total number of commercial consumers (14,976), the average value being almost 5900 Euro per consumer. However, we noticed a significant difference in the population sizes and consumption levels of the two categories. As individual savings could be considered as reasonable for commercial, savings for residential consumers are very modest and may not incentive enough to implement DR programs.

As a further study, we will focus on emergent data management solutions in relation to the local flexibility markets and their implementation by means of direct load control and machine learning algorithms that are able to forecast the generation of local energy sources (Photovoltaic (PV) and wind generating systems) and balance the consumption requirements, storage and provide the surplus to the local markets. Furthermore, the model implementation can be further enhanced with consumption forecast, DR enabling technology costs and climate area characteristics.

Author Contributions: The authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI-UEFISCDI, project number 462PED/28.10.2020, project code PN-III-P2-2.1-PED-2019-1198, within PNCDI III. This paper also an extension and presents the scientific results of the project “Intelligent system for trading on wholesale electricity market” (SMARTRADE), co-financed by the European Regional Development Fund (ERDF), through the Competitiveness Operational Programme (COP) 2014–2020, priority axis 1—Research, technological development and innovation (RD&I) to support economic competitiveness and business development, Action 1.1.4-Attracting high-level personnel from abroad in order to enhance the RD capacity, contract ID P_37_418, no. 62/05.09.2016, beneficiary: The Bucharest University of Economic Studies.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

DR	Demand Response
RES	Renewable Energy Sources
EV	Electric Vehicle
ISO	Independent System Operator
DLC	Direct Load Control
PCT	Programmable Communicating Thermostat
ADR	Automated Demand Response
HVAC	Heating, Ventilation, and Air Conditioning
SASM	Smart Adaptive Switching Module
EH	Electricity Heating
WH	Water Heater
ToU	Time-of-Use
$h \in [h_{PEAK_START}, h_{PEAK_STOP}]$	Hour
h_{PEAK_START}	Peak start hour
h_{PEAK_STOP}	Peak stop hour
W_f	Flexible capacity;
α_f	Flexibility coefficient;
C_{EH}	Consumption of electricity heating
C_{WH}	Consumption of water heater
C_{HVAC}	Consumption of HVAC
S	Savings from flexibility usage
t_{PEAK}	Tariff rate at peak hours
$t_{OFFPEAK}$	Tariff rate at off-peak hours

References

- Hledik, R.; Faruqui, A.; Lee, T.; Higham, J. The National Potential for Load Flexibility: Value and Market Potential through 2030. Available online: https://brattlefiles.blob.core.windows.net/files/16639_national_potential_for_load_flexibility_-_final.pdf (accessed on 2 November 2020).
- D'hulst, R.; Labeeuw, W.; Beusen, B.; Claessens, S.; Deconinck, G.; Vanthournout, K. Demand response flexibility and flexibility potential of residential smart appliances: Experiences from large pilot test in Belgium. *Appl. Energy* **2015**, *155*, 79–90. [\[CrossRef\]](#)
- Klaassen, E.A.M.; Kobus, C.B.A.; Frunt, J.; Slootweg, J.G. Responsiveness of residential electricity demand to dynamic tariffs: Experiences from a large field test in the Netherlands. *Appl. Energy* **2016**, *183*, 1065–1074. [\[CrossRef\]](#)
- Lopes, R.A.; Chambel, A.; Neves, J.; Aelenei, D.; Martins, J. A Literature Review of Methodologies Used to Assess the Energy Flexibility of Buildings. *Energy Procedia* **2016**, *91*, 1053–1058. [\[CrossRef\]](#)
- Reynders, G.; Amaral Lopes, R.; Marszał-Pomianowska, A.; Aelenei, D.; Martins, J.; Saelens, D. Energy flexible buildings: An evaluation of definitions and quantification methodologies applied to thermal storage. *Energy Build.* **2018**, *166*, 372–390. [\[CrossRef\]](#)
- Klaassen, E.A.M.; van Gerwen, R.J.F.; Frunt, J.; Slootweg, J.G. A methodology to assess demand response benefits from a system perspective: A Dutch case study. *Util. Policy* **2017**, *44*, 25–37. [\[CrossRef\]](#)
- Potter, J.; Cappers, P. Demand Response Advanced Controls Framework and Assessment of Enabling Technology Costs. Available online: https://emp.lbl.gov/sites/default/files/demand_response_advanced_controls_framework_and_cost_assessment_final_published.pdf (accessed on 2 November 2020).
- Söder, L.; Lund, P.D.; Koduvere, H.; Bolkesjø, T.F.; Rossebø, G.H.; Rosenlund-Soysal, E.; Skytte, K.; Katz, J.; Blumberga, D. A review of demand side flexibility potential in Northern Europe. *Renew. Sustain. Energy Rev.* **2018**, *91*, 654–664. [\[CrossRef\]](#)
- Rotger-Griful, S.; Jacobsen, R.H.; Nguyen, D.; Sørensen, G. Demand response potential of ventilation systems in residential buildings. *Energy Build.* **2016**, *121*, 1–10. [\[CrossRef\]](#)
- Müller, T.; Möst, D. Demand Response Potential: Available when Needed? *Energy Policy* **2018**, *115*, 181–198. [\[CrossRef\]](#)
- Arteconi, A.; Polonara, F. Assessing the demand side management potential and the energy flexibility of heat pumps in buildings. *Energies* **2018**, *11*, 1846. [\[CrossRef\]](#)
- Yin, R.; Kara, E.C.; Li, Y.; DeForest, N.; Wang, K.; Yong, T.; Stadler, M. Quantifying flexibility of commercial and residential loads for demand response using setpoint changes. *Appl. Energy* **2016**, *177*, 149–164. [\[CrossRef\]](#)
- Klein, K.; Herkel, S.; Henning, H.M.; Felsmann, C. Load shifting using the heating and cooling system of an office building: Quantitative potential evaluation for different flexibility and storage options. *Appl. Energy* **2017**, *203*, 917–937. [\[CrossRef\]](#)
- Diamantoulakis, P.D.; Kapinas, V.M.; Karagiannidis, G.K. Big Data Analytics for Dynamic Energy Management in Smart Grids. *Big Data Res.* **2015**, *2*, 94–101. [\[CrossRef\]](#)
- Wang, Y.; Chen, Q.; Kang, C.; Xia, Q. Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications. *IEEE Trans. Smart Grid* **2016**, *7*, 2437–2447. [\[CrossRef\]](#)

16. Muntean, M. Business intelligence issues for sustainability projects. *Sustainability* **2018**, *10*, 335. [[CrossRef](#)]
17. Muntean, M.; Cabău, L.G.; Rînciog, V. Social Business Intelligence: A New Perspective for Decision Makers. *Procedia Soc. Behav. Sci.* **2014**, *124*, 562–567. [[CrossRef](#)]
18. Wang, Y.; Chen, Q.; Hong, T.; Kang, C. Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. *IEEE Trans. Smart Grid* **2019**, *10*, 3125–3148. [[CrossRef](#)]
19. Nordic Countries of Ministers. *Flexible Demand for Electricity and Power: Barriers and Opportunities*; Nordic Energy Research: Oslo, Norway, 2017.
20. Office of Energy Efficiency & Renewable Energy (EERE). Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States. Available online: <https://openei.org/datasets/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3-locations-in-the-united-states> (accessed on 30 October 2020).
21. National Renewable Energy Laboratory Building America House Simulation Protocols. Available online: <https://www.nrel.gov/docs/fy11osti/49246.pdf> (accessed on 30 October 2020).
22. mongoDB. Comparing MongoDB vs PostgreSQL. Available online: <https://www.mongodb.com/compare/mongodb-postgresql> (accessed on 30 October 2020).
23. National Renewable Energy Laboratory, U.S. Department of Energy Commercial Reference Building Models of the National Building Stock. Available online: <https://www.nrel.gov/docs/fy11osti/46861.pdf> (accessed on 30 October 2020).