URBAN-SCALE BUILDING ENERGY MODELING

Market Potential, Data Sources, Computational Methods and Visualization

ABSTRACT

This paper explores urban-scale building energy modeling (UBEM). First, it assesses the market needs of UBEM through interviews with relevant stakeholders to quantify the potential that it holds for large-scale investment in building energy efficiency. The paper also reviews currently-used and potential data sources, computational methods and scalable computing approaching for creating and simulating UBEM of every U.S. building. Lastly, we demonstrate some visual analytics tools that can be employed to visualize UBEM, using sample buildings in Washington D.C. as a case-study.

Amoo, Charles Nii-Baah

Qualifying Exams, Energy Science and Engineering
May 2, 2022

Table of Contents

CHAPTER ONE: BUSINESS MODEL

Introduction

Globally, buildings are responsible for 40% of energy consumption and a third of greenhouse gas emissions [1] and electricity consumption [2]. In the U.S., the situation is no different as building sector energy consumption also stands at 40%, with a similar proportion of carbon footprint (39%) by same [3]. This includes 73% of electricity use and peaks at 80% when demand rises [4]. As developing countries improve access to energy and tropical countries grow in their need for air-conditioning, amid rapid urbanization and increased ownership and use of electrical appliances, global energy demand from buildings is expected to rise significantly [2].

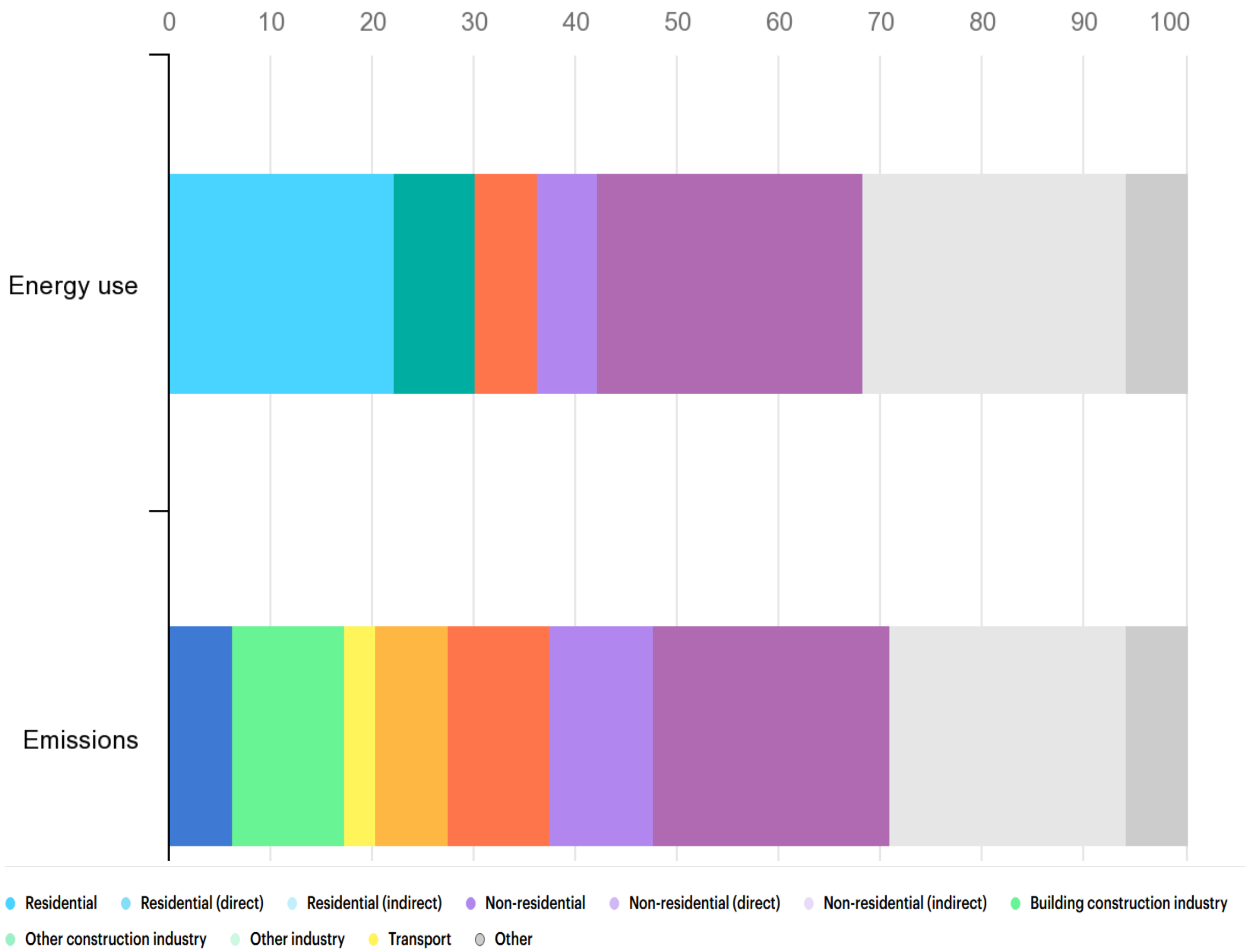

Figure 1: Global energy use and energy-related CO2 emissions by sector, 2020 [2]

Even as governments and government sectors work to enact policies to increase the minimum performance standards of buildings and the deployment of renewable energy in buildings as part of decarbonization efforts, less than 1% of U.S. buildings are improved each year [5]. Also, there is a substantial unserved market for building energy efficiency, especially in the residential and small commercial (less than 50,000 ft^2) building sectors [6]. As a result, buildings are still not on track to achieve carbon-neutral emissions by 2050 [2]. One of the ways to meet this target is for all new buildings and 20% of the existing building stock to be zero-carbon-ready no later than 2030 [2]. Also, building sector energy intensity would need to reduce five times more rapidly over the next decade than in the past five years. Meaning, energy consumption per ten square feet must be 45% less in 2030 than the 2020 baseline.

Among the many recommendations that the International Energy Agency (IEA) proposes as measures to reduce building sector energy consumption and greenhouse gas emissions, technology products that provide consumers with information to help them make informed decisions about energy efficiency stand out. Also, technology awards such as the Global Cooling Prize [7] that reward leading-edge innovation efforts to increase efficiency and reduce costs are highlighted [2].

In the U.S., the Department of Energy's (DOE) Building Technology Office's (BTO) long-term goal is to reduce the energy use intensity (EUI) of U.S. buildings by 50% compared to the 2010 baseline. The short-term goal is to minimize building EUI by 30% by 2030 [8]. The vision of the Office of Electricity is to harness innovation to make the North American energy system more robust, more resilient, and reliable. [9].

Research Goal

Domestic (U.S.) and global building energy efficiency goals underscore the vital role of technology and innovation. One of such innovative and technology-driven solutions is an Urban-Scale Building Energy Modeling (UBEM) software technology called AutoBEM (Automatic Building Energy Modeling), which employs high computational algorithms and has been (or can be) used to generate and simulate almost all 125 million buildings of the U.S. This section investigates the market needs of UBEM, using AutoBEM as a case study, summarizes characteristics of a potential solution via a business model canvas, and seeks to quantify the potential for how building-specific energy analysis might best enable larger-scale investments in building energy efficiency.

Methodology

The data collection methods included both interviews and a literature review. One professional from Ameresco, an Energy Service Company (ESCO), two professionals from the Electricity Power Board (EPB) of Chattanooga, and an Urban Planner were interviewed. A sample of the questions used in the interview can be found in Appendix 1 on page 27.

Developing the Business Model

Assessing Market Needs

The insight gained from the interview with Ameresco suggests a shift of focus from individual buildings to a collection of facilities, including higher education and municipalities. This is because ESCOs appear to be having issues with the business model of providing energy efficiency services to customers and being paid from the cost savings that come from implementing Energy Conservation Measures (ECMs). Some of the reasons given for this include (i) customers are tired of that business model, (ii) customers do not want to incur debt, (iii) customers feel energy savings estimates are not dependable, and (iv) customers think they can do it themselves. Beyond this, traditional physics-based building simulation requires so many inputs and person-hours, which, when scaled to urban levels, even for higher education and municipalities, could become prohibitively laborious using the traditional approach. It is for this reason that AutoBEM has become valuable.

For EPB Chattanooga and other utilities, AutoBEM may help segment customers based on their consumption profile. The actual functionality that could prove more valuable, it seems, is one that allows utilities to dynamically track the consumption profile for customers even as their usage behavior changes or they relocate. Moreover, utilities are looking for ways to prioritize their investments based on demographic data, disaggregate loads and be able to define energy consumption outliers and weed them out. Furthermore, for the application to become useful for non-data-scientists or non-computer-scientists, it needs to have a front end. Indeed, AutoBEM can meet some of the utility's requirements with ongoing work to add more functionalities.

The last interview was with an urban planner with experience in both public sector and private-sector consulting. While this interview did not indicate much of a good fit between AutoBEM and urban planning requirements, it was interesting that the information such as cartographic data for zoning, elevation, and other building information from the Tax Assessor's office may be helpful for land use law firms. Urban planners, however, will find data on tree coverage and building coverage that is separated into roads and sidewalks more helpful in their work.

Market Potential

Pre-covid annual spending on energy in the U.S. was around US$1.3 trillion, with retail electricity accounting for nearly 32% (US$403 billion). By sector, residential buildings were the second-largest energy spenders (21% of the U.S. total), having surpassed the industrial sector since 2012[10].

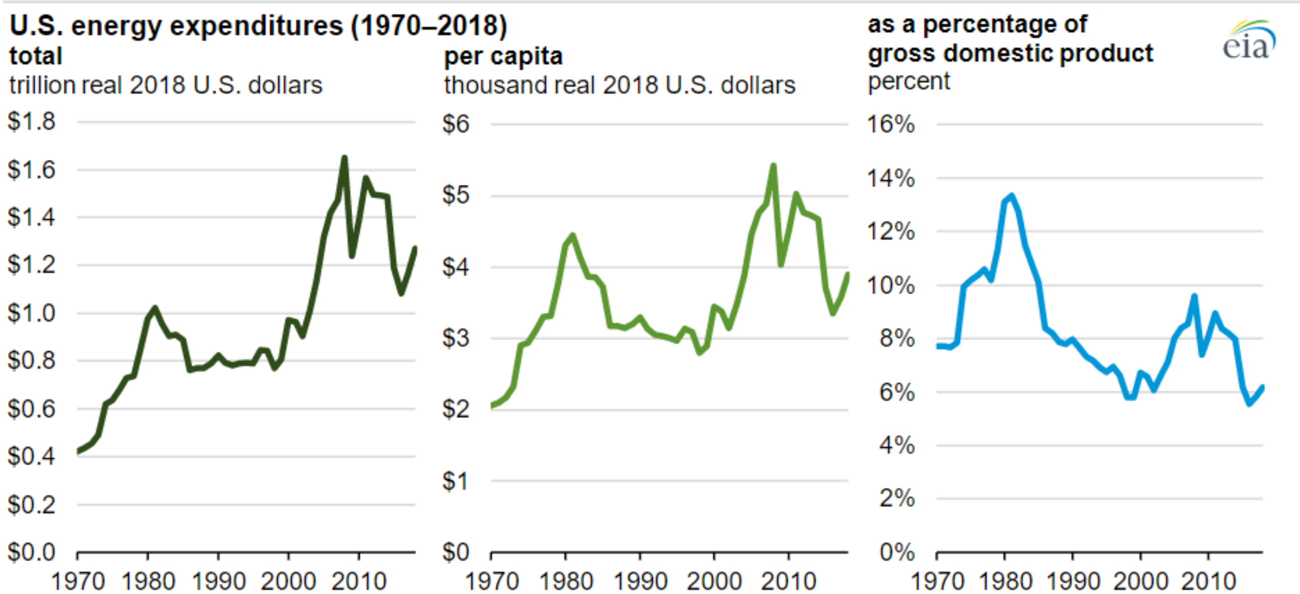

Figure 2: Annual Spending on Energy in the U.S. [10]

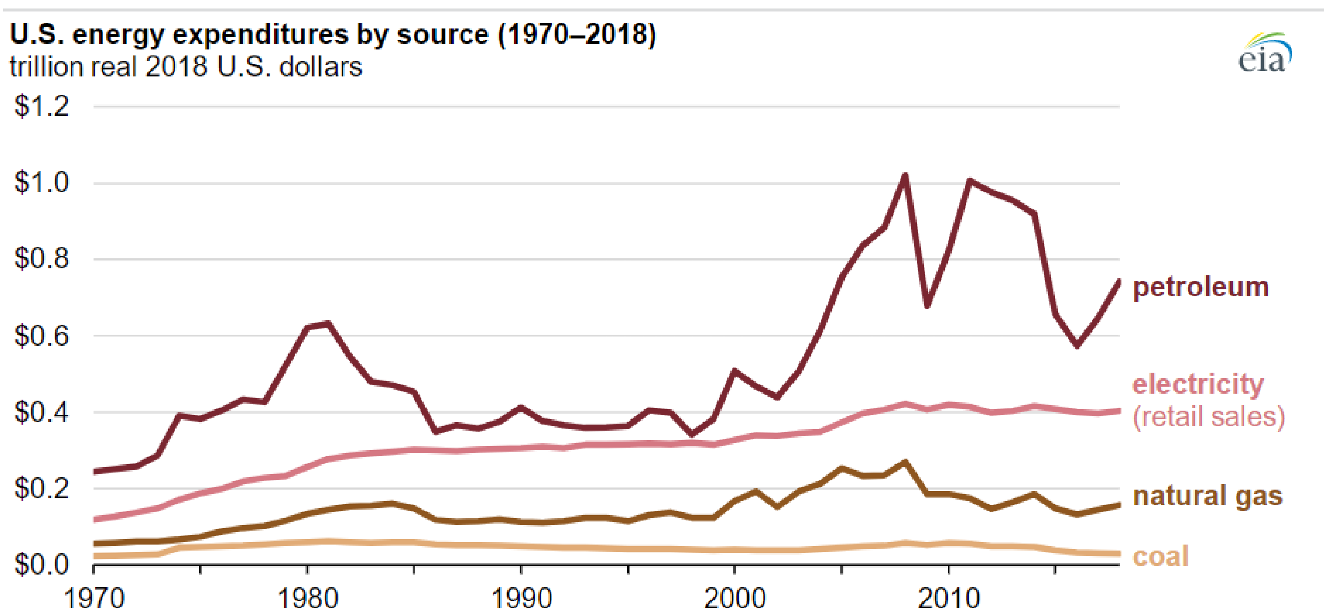

Figure 3: U.S. Annual Spending on Energy by Source[10]

While the global energy as a service (EaaS) market size is estimated to be US$54.5 billion in 2020 and is projected to double by 2030 to US$112.7 billion [11], the global market for energy-efficient devices is forecasted to more than double and reach US$1,771.70 billion by 2028 compared to US$690.52 billion in 2020, with a compound annual growth rate (CAGR) of 12.5% [12]. Of the 2022 market share, North America accounted for 40.2% by region and the residential application segment has the largest share of 48.0%. The residential application segment is projected to increase its market share further as more homeowners adopt smart meters and other home energy management systems, with support from government initiatives toward green buildings guidelines or Energy Conservation Building Code (ECBC). The device type projected to lead the growth is consumer electronics and appliances with a CAGR of 12.8%, followed by smart lighting. Generally, growth in the global energy efficiency market will be fueled by an increasing need to be energy efficient and growing concerns about the effects of climate change. Another factor is the decarbonizing efforts causing firms to invest more in renewable energy resources and technologies. Also, increasing efforts in research and development to make buildings more energy-efficient and produce energy-efficient devices, coupled with rapid urbanization and rising consumer purchasing power, could contribute to the growth of the energy efficiency market [12].

The energy management software is also a big player in the energy efficiency market, with the global energy-management software market size estimated to rise from US$33.2 billion in 2020 to US$69.3 billion by 2027.

National, city and local authorities are the major investors in energy efficiency, especially in buildings and transportation, given their large stock of buildings and fleet [13].

Hindrances to Market Adoption

The major limitations that could hinder growth in this sector are the lack of awareness and information about energy efficiency and energy-efficient devices and the often high initial cost of some energy efficiency measures [12]. In UBEM, the lack of high computational resources and lack of access to data or issues of intellectual property surrounding the data could serve as a major impediment. Another reason businesses give for not investing in building energy efficiency is that energy-efficient technologies are expensive, risky, and hard to operate and maintain. Also, property or facility managers do not have the time to implement energy efficiency measures or do not have the expertise to do so. The difficulty in accessing the information on building energy efficiency or the lack thereof is another roadblock. A significant limitation is the lack of finance or the colossal capital expenditure requirement for some energy conservation measures [14].

Value Proposition for Major Stakeholders

AutoBEM addresses these concerns by providing independently verifiable energy models generated using high-performance computing resources to save both cost and time. Given that lack of trust is a significant concern often raised by customers concerning ESCOs, the models generated by AutoBEM could provide a basis for either an ESCO or a consumer to initiate an energy audit request or service. Generally, the models could provide engineers, architects, and energy managers with the resources needed to set specific energy-efficiency goals for their clients. It can also help building projects meet LEED (Leadership in Energy and Environmental Design) requirements or obtain LEED certification and estimate buildings' future energy usage and cost [15].

For the main customer targets of UBEM – Utilities and ESCOs – AutoBEM offers a variety of values. The ability of AutoBEM to reach beyond the electric meter and independently investigate a building's energy use is one of significant value. Utilities can use the models to map out customers (or geographical areas) based on their energy consumption and use that to formulate peak rate structure. As a result of this mapping or knowledge of peak demand, utilities can be initiative-taking in designing products and services to manage demand through smart appliances such as smart thermostats and water heaters or grid stability measures such as voltage and frequency management. Also, utilities can apply the right emission conversion factor to estimate the emissions contributed by each building within their service area. This will help them put measures in place to meet emissions targets or comply with emissions regulations. Moreover, utilities can use the models to plan energy efficiency programs, especially for high-energy users. Furthermore, these models can be used to educate customers on their energy usage and what measures they could take to reduce their energy consumption [16].

UBEM as an Enabler for Investment in Energy Efficiency

UBEM tools hold many values, but getting the tool to where it is needed to make a real impact is quite another thing. There have been many measures to encourage investment in energy-efficient technologies, including the electrical appliances energy efficiency rating and labeling [17]. The American Council for an Energy-Efficient Economy (ACEEE) suggests ten ways to increase energy efficiency investments [14]. One of the suggestions that aligns more with UBEM is to provide a building performance benchmarking analysis that can help energy consumers with below-average performance easily identify themselves and be stirred toward improvement. The building-specific energy analysis that gets to the consumer must highlight the things they care about, such as their contribution to climate change in terms of carbon emissions, measured on a micro-scale and quantified in simple, relatable metrics such as bags of coal, volume reduction in water supply, or reduced agricultural yield. Also, UBEM developers could partner with utilities to provide energy consumption and efficiency analysis on a monthly or real-time basis using smart energy metering and monitoring devices. This could well serve residential buildings, which happen to be the most underserved segment of the building sector as far as energy efficiency services are concerned. There is already a business case for this. For example, Pacific Gas and Electric (PG&E) power company is experimenting with a pay-for-performance weatherization program whereby the utility will reimburse contractors for the energy savings they achieve [14]. Another way that investment can pour into building energy efficiency is when investors and financing firms have reliable data on potential energy savings to support their decisions. AutoBEM could be an independent energy efficiency data source for such purposes.

Go-to-market Strategy

As UBEM is still a budding field, getting AutoBEM to the market will require more research and publications. Another way is to highlight what AutoBEM can do at industrial and academic conferences and exhibitions. Also, the product will need to be assessed with specific markets to understand its use cases in those markets as well as fill useful functionality gaps that might be missing but required. The data output for AutoBEM would be available through web services and APIs and interactive 3D visualizations such as CesiumJS.

Partnerships for Success

There are no known competitors for this endeavor. The intent is not so much for profitmaking as stimulating a nationwide adoption of and investment into building energy efficiency. The goal of this market assessment investigation could lead to that purpose. Therefore, if, through this report, utilities and ESCOs as partners, rather than competitors, would consider how to leverage AutoBEM or UBEM in general to achieve more significant investment in building energy efficiency, the purpose of this assessment would be achieved. Other partners relevant to this venture include data providers and I.P. attorneys to help deal with I.P. concerns regarding the data used as input for AutoBEM.

Funding Needs

The primary funding needs for AutoBEM will be for continued research and development to refine the models and make them more accessible and meaningful to those who are not experts in the field. Funding may also be required to close context-specific gaps in data, models and simulation results. The key resources to operate and maintain AutoBEM include software developers, data scientists, building technologists, GIS experts, and high-performance computers.

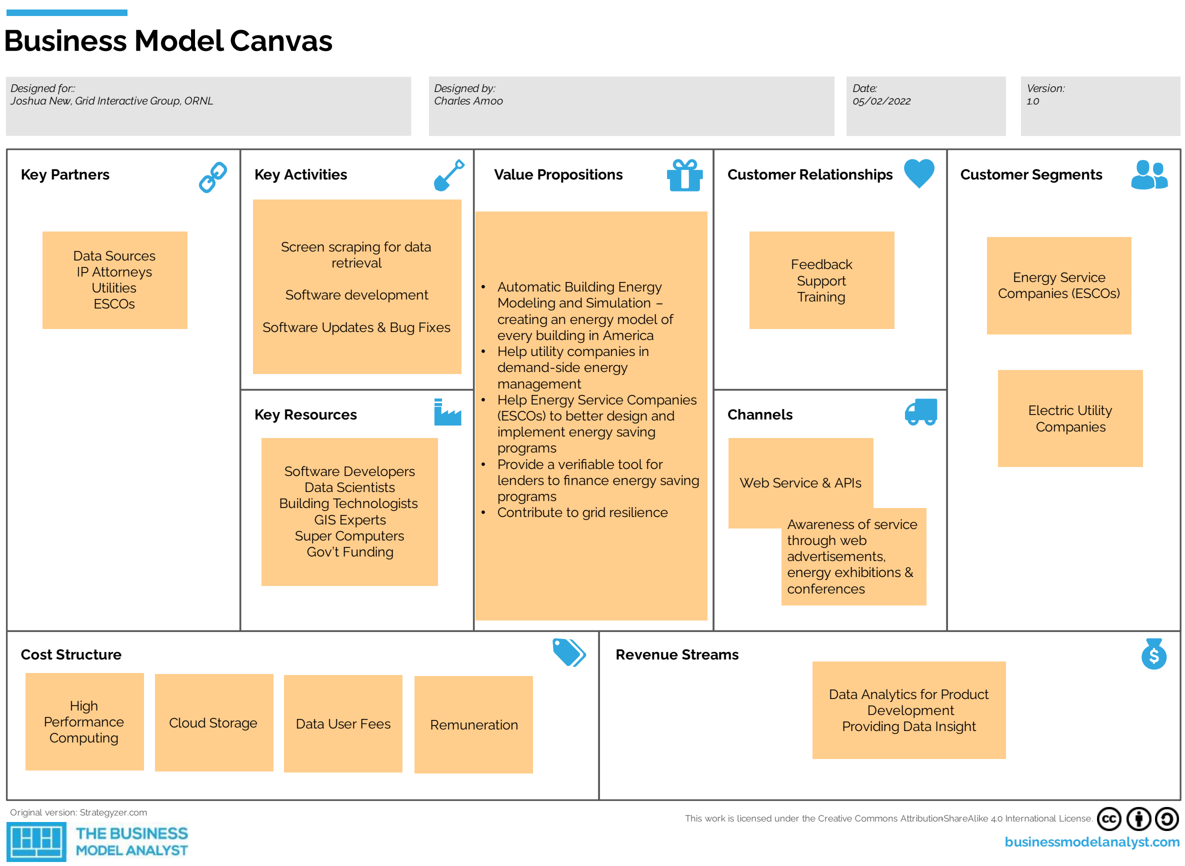

Figure 4: Summary of Business Model Canvas for AutoBEM [18]

CHAPTER TWO: DATA SOURCES, COMPUTATIONAL METHODS AND SCALABLE COMPUTING APPROACH

Data sources

Generating building models and simulating energy use of buildings are heavily data-dependent processes that require lots of building data by the hundreds and well into the thousands, depending on the building type. The table below shows the number of building descriptors for different building types that were modified for an Urban-Scale Energy Modeling (UBEM).

Table 1: Typical Building Descriptors for Building Energy Modeling[19]

	Small Office	Outpatient	Large Office	Medium Office	Hospital	Warehouse	Small Hotel	Large hotel
Inputs	458	3483	1072	760	1955	333	1823	887
	Strip Mall	Retail	Quick Service Restaurant	Full Service Restaurant	Mid Rise Apt	High Rise Apt	Secondary School	Primary School
Inputs	800	438	281	286	1464	4617	1621	1051

Building energy modelers usually rely on the information on climate, building physics, construction materials, equipment components, shading, and occupancy to better characterize a building [20]. While it is possible to obtain almost all the building information input for an individual or a few clusters of buildings, it is highly impractical to get such a high level of details for building energy modeling that involves large stock of buildings at the urban scale. For modeling at the urban scales, such as at the state or national level, where most of the building information is unknown, one of the building information useful to generate energy models of buildings includes building footprint. Other information are height, the number of stories, window-to-wall ratio (WWR), building type, and building vintage [21], [22]. By applying heuristics, suitable input data can be generated for some building properties, such as building type [21]. Also, values from building prototypes such as those provided by the Department of Energy (DOE) [24] can be used to create building models that require minimal input [25]. For building footprints, Microsoft's OpenStreetMap uses deep learning, and image processing algorithms to provide an open, free, downloadable computer-generated footprint of over 129 million U.S. buildings and those of other regions [26]. Representative weather data for the energy simulation can be obtained from OpenStudio or the Office of Energy Efficiency and Renewable Energy website [24]. While the information on building height is more difficult to obtain, Light Detection and Ranging (LiDAR) technology makes it possible to have that information, which certain U.S. counties and regions utilize to make building height data publicly available. But building height information obtained by such means is not without uncertainty and challenges because of the resolution. Height data obtained at low resolution can often be confused with proximate objects with substantial height. It is essential, therefore, that height data is acquired at a higher resolution.

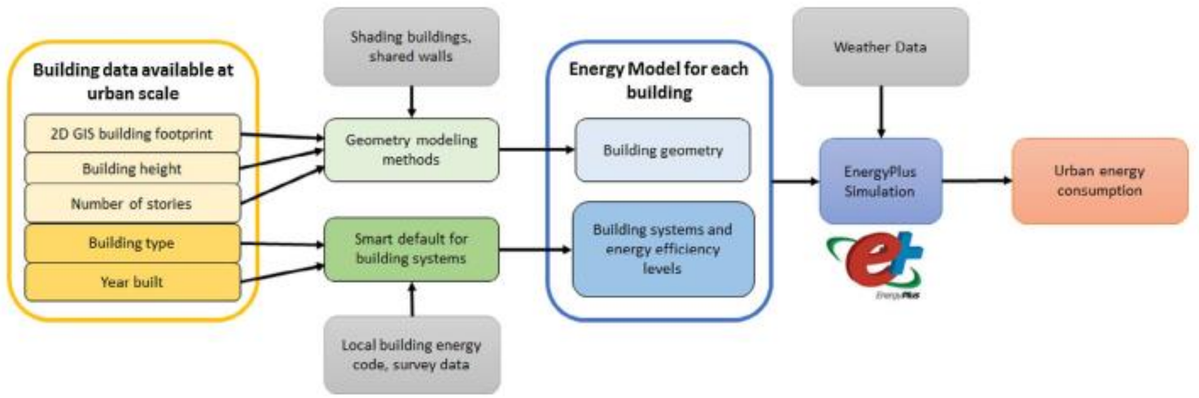

Figure 5: Major Components of Urban Building Energy Modeling[22]

Government agencies that hold valuable information on buildings are important data sources for building modeling. In the U.S., the General Services Administration (GSA) is a source for building footprint, the Department of Planning for land use data, the Tax Assessor's Offices for assessor records, including building vintage, and the Departments of Environment for energy disclosure [23]. In the U.K., a UBEM platform called SimStock, built on EnergyPlus, was developed for the automatic generation of detailed simulation models to predict the energy and environmental performance; another software called 3DStock was created to automatically generate a detailed built form from publicly available national datasets. The publicly available building data that fed the 3DStock software included floor space use type from the Valuation Office Agency (VOA), building addresses from the Ordinance Survey, measured energy data from the Department for Business, Energy & Industrial Strategy, as well as land parcels and sites from the Land Registry [20].

The table below summarizes six general data types that are employed in UBEM:

Table 2: Data types and description for Urban-Scale Building Energy Modeling[16]

Data Type	Data Description
Imagery (satellite, airborne)	These are images taken by satellites and airborne cameras, and they cover extensive urban areas from which building footprints can be detected and extracted.
Street-level Imagery	These are images that contain a collection of photographs that can be used for direction search and navigation
Cartographic data	These are databases that include different geographic data layers used for mapping and GIS analysis. The main components of a cartographic database are the x-y coordinates that define a geographic area [27].
Elevation data	This consists of elevation measurements obtained from LiDAR or computationally-derived models used to detect buildings at high spatial resolution and low resolutions.
Building information databases	This information can be obtained from the U.S. tax assessors' database. Depending on the county, the data from the tax

	assessor’s office may not be digitally available or may be available in an entirely different format from other counties.
3D building model databases	This provides 3D geometry of individual buildings over wide areas.

While building prototypes and government agencies serve as valuable data sources for direct use in UBEM, they do not hold data for all the building descriptors needed for modeling. Therefore, plausible assumptions must be part of the data sources where actual data is unavailable. For example, knowing the building type can help estimate the floor-to-floor height, which can be combined with the actual or estimated height of the building and the number of floors. Again, information on building type also helps determine the WWR, occupancy schedule, and HVAC type. Furthermore, building vintage can inform the kind of insulation and infiltration to use in the model. While the information on building vintage can be obtained from the Tax Assessor’s offices, recent research shows how annual maps of Global Artificial Impervious Areas (GAIA) from 1985 to 2018 could be used to estimate the vintages of several tens of millions of buildings [28]. No combination of these data sources is robust enough to describe the building accurately, but the aim is that aggregating these data sources from actual, estimated and assumed information would provide a representative model that is as close to the actual as possible. Since this data is used to classify buildings, it is imperative to get them right as some building misclassifications could result in energy use of ten times more or less. Where an actual building energy use exists from a utility or measured data, prototypical models of each building category could be compared against actual energy use to assess which ones fit best. This type of benchmarking provides a better representative model for building type, if not for building energy use [29].

Previous work performed to investigate the data sources for UBEM identified 37 data sources. Below is a summary of these data sources and their characteristics [30].

	Data type	Source	Comments
1	Street view image	Google	A free service, downloadable through an API and covers the entire U.S.; can be used to map out the height and window-to-wall (WWR) of buildings as well as building material properties
2	Open map data	OpenStreetMap	This is free, derived data that provides shapefiles of buildings from which building footprints, road network, buildings, and parking lots can be mapped out
3	Satellite image	DigitalGlobe	A paid service that provides satellite imagery from which building footprint, land covers, and their material properties can be mapped out
4	Aerial image	DigitalGlobe	A paid service that provides aerial view images from which building footprint, land covers, and their material properties can be mapped out
5	Image	Pictometry International Corporation	A paid service that provides reports including high-resolution aerial images and 3D diagrams

			from which 3D building models and building exterior materials can be mapped out
6	Ground elevation data	U.S. Geological Survey	A free service that provides main floor ground and road surface elevation that covers the entire U.S.
7	Street view image	Mapillary	A free service that provides crowdsourced, single street view photos from which road pavement materials can be mapped out
8	Remote sensing images	U.S. Geological Survey	A largely free dataset that provides remote sensing imageries from which building footprint, land covers, and their material properties can be mapped out
9	Imagery and terrain	Cesium	Provides an interactive, visualizable 3D platform that is built on JavaScript and HTML/CSS together with the codes from which building footprints and height can be mapped out
10	Land cover image	Multi-Resolution Land Characteristics Consortium	Provides free, land cover information for the whole of the U.S. from which building footprint can be obtained
11	Shapefiles, Geodatabase	U.S. Geological Survey	Provides a free database of geographic information on physical structures that includes name, function and location, among others. From these, the building footprint can be mapped out.
12	Vector and raster images	Volunteers and NACIS	Provides free cultural and physical vector files from which physical properties can be segmented
13	Map images	NASA	Free map data that provides rural-urban mapping, settlement and population
14	Topographic images	ASU[1], SDSC[2], NSF[3]	Provides free topographical information, including raw point cloud data from which building heights could be extracted
15	GIS data	UNEP[4]	A free GIS database that provides visualizations of socio-economic variables, accessible in multiple formats through web services
16	GIS	US Geological Survey	Free digital evaluation maps from which building height could be extracted

[1] Arizona State University
[2] San Diego Supercomputer Center
[3] National Science Foundation
[4] United Nations Environmental Program

17	GIS	City of Knoxville	Free local GIS information for Knox County, Tennessee, suitable for zoning, land use and planning
18	Elevation data	City of Knoxville	Local GIS information for Knox County, Tennessee, suitable for building height extraction
19	Building information	Zillow	A database of building information, including year, age, size and sometimes HVAC systems of more than 100 million buildings (primarily residential buildings), accessible through API
20	Geodatabase, shapefiles	U.S. Geological Survey	Free transportation data consisting of roads, airports, railroads and other features
21	Orthoimagery	US Geological Survey	Free downloadable high-resolution aerial images that combine attributes of an aerial photograph with the spatial accuracy and reliability of a planimetric map. ("Imagery | U.S. Geological Survey") Can provide building footprint and materials.
22	Various data types	ESRI	ArcGIS data is a subscription-based platform that provides high-resolution imagery, base maps, multispectral, historical maps, and boundaries from which building footprint, materials and sizes could be extracted.
23	Various data types	ESRI	Another platform called ArcGIS 3D cities employs Pictometry to produce the most detailed and realistic imagery-texture 3D building models as true representations of the as-built conditions in 3D GIS projects for planning, facility management, and public engagement.
24	3D Model (SketchUp)	Trimble	Provides a wide array of different 3D models from which building footprints, material, sizes, windows, and geometry could be extracted.
25	3D buildings (GeoJSON)	OpenStreetMap	Provides free, wide array of different 3D models from which building footprint, sizes, and geometry could be extracted
26	Imagery, shapefiles, DEMS, orthoimagery	State Clearinghouse	Provides free GIS data of each state from which building footprints, sizes and geometry could be extracted.
27	Geodatabase	Homeland security infrastructure program (HSIP)	Provides various datasets related to energy, water supply, finance, and education. (Not publicly available)

28	LiDAR[5]	NGA	LiDAR data for all the urban areas in the contiguous USA from which building footprint and height could be extracted

Specific algorithms are written to get more data beyond what is often available. One of such algorithms is for building height estimation. This method involves using building footprints on 2D maps and Google Street view images to estimate building height. Another method called factorization-based texture segmentation is used to partition an image into homogeneous regions, distinct from local neighbors and visually meaningful. The goal is to develop an approach to produce meaningful segmentation results efficiently. An application of this method is neighborhood mapping, where aerial images are used to map different socio-economic neighborhoods based on distinct texture appearances with results close to those from human analysts. Another application of texture segmentation is in roadbed extraction from high-resolution aerial images. An algorithm also uses LiDAR data to estimate solar radiation on urban structures for different applications, including solar photovoltaics and road de-icing. One area of need is the ability to evaluate building counts in aerial images. An attempt to solve this problem is through machine learning algorithms to train classifiers that identify line segments corresponding to building edges and observing the relationship (strong correlation) between the number of buildings and the number of lines for similar building types [30].

A widespread tool for generating building models is OpenStudio which can provide model output as an OpenStudio Model in .OSM format or EnergyPlus model in .IDF (Input Data File) format or both. These models can then be simulated (or run) on EnergyPlus to obtain energy consumption information. While traditional building energy modeling and simulation utilizing OpenStudio and EnergyPlus have often been performed on single buildings to a few clusters of buildings on personal laptops and desktop computers, UBEM requires that the modeling and simulation processes be integrated and streamlined in an automated workflow and run on High-Performance Computing (HPC) resources in a scalable manner. The following two sections highlight the computational methods available for UBEM and the scalable computing approach employed in a real-world UBEM project.

Computation Methods

The computational methods employed in generating building energy models and simulating building energy use at the urban scale are varied and many. Some of the computational methods applied include specific Artificial Intelligence (A.I.) algorithms, metaparameter settings, data preparation methods, supervisory signal, training methods, validation methods, and best-in-class metric-based rates of performance. For example, depending on the resolution, satellite or aerial street-view imagery, A.I. algorithms can be used to extract the WWR information and detect

[5] LiDAR data are remotely sensed high-resolution elevation data collected by an airborne collection platform using a combination of laser range-finding, GPS positioning, and inertial measurement technologies to make highly detailed DEMs (digital elevation models) of the earth's terrain, man-made structures, and vegetation.

HVAC systems. Light Detection and Ranging (LiDAR) technology can be used to estimate the footprint of buildings [20] [19].

With all the input parameters in place, generating models or simulating energy use can take a statistical (or spreadsheet) approach or an engineering approach [31], albeit the latter is more common. As developed for use at MIT by Atelier Ten, an environmental design consulting firm, the spreadsheet approach employs statistical techniques to estimate energy use intensity (EUI) for each of the areas of use with internal loads and operating parameters as primary predictors of EUI. With this approach, building envelope, vintage and other design parameters do not play significant roles in predicting EUI. Therefore, the total annual facility energy use is a linear function of the floor area and the EUI of each energy distribution within the facility. In this approach, known values of each building within an assigned energy use classification of the facility are regressed. The main goal of this model is to produce a means of predicting changes in future energy use from future retrofits or other energy use scenarios. Based on modeler experience and knowledge of the facility, the models can be iteratively simulated and compared to actual energy use until there is a good fit and can be used to simulate the future impact of energy efficiency retrofits [31].

The engineering approach, on the other hand, is built on the assumption that building energy use is more complex and dynamic than the mere floor areas of the energy use space and involves envelope variations, electrical and mechanical systems as well as occupant behaviors, all of which can significantly impact on individual building's energy consumption even if their designated uses are similar [31]. The broad categories of input parameters for this modeling approach include envelope characteristics such as building geometry, glazing ratios, wall and roof construction; internal load ranges such as the per unit area of occupants, equipment, and lighting; and mechanical systems such as the cooling and heating setpoints and sources as well as the ventilation flowrate [31].

Scalable compute approaches

Unlike modeling individual buildings or a few clusters of buildings within an organizational environment, generating models and simulating the energy performance of buildings at urban scales could be arduous for personal computers or servers to manage. For UBEM at the city, state or national level, which involves hundreds of thousands to several millions of buildings, computational resources with scalable high-performance capabilities are needed.

Scalable computing, more broadly defined, is the ability to increase computational performance or resources to handle more work [32]. Scaling can happen at the hardware or software level [33], but for UBEM, the scalable computing approach that is more applicable is with the software in what is known as parallelization. Parallelization involves getting the most efficient speedup time from the computing resources (cores or nodes) that are allocated to a work. It is achieved by deploying many processors to work simultaneously to achieve very high computational power that reduces computational time [32]. In parallelization, the main challenge is coming up with the optimal way of breaking up the problem into chunks that can be executed at a reasonably high speed, given the computing resources allocated to the task.

One of the scalable computing approaches for a nation-scale BEM is to categorize buildings into different archetypes – a simplified, theoretical building model composed of several characteristics that are typically found within a category of buildings with similar attributes [34] – and simulate the archetype models for all important climate zones, and multiply the results with the number of buildings for each archetype [21] [35].

In a novel building energy simulation tool built on EnergyPlus and OpenStudio called AutoBEM – Automatic Building Energy Modeling, parallelization techniques were used in the energy simulation workflow to simulate 125 million in the U.S. on Argonne Leadership Computing Facility's (ALCF) Theta supercomputer. Theta has 4,392 compute nodes, with each node having sixty-four cores. Each core is a 1.3GHz Intel Xeon Phi 7230 processor [36]. This makes it possible to parallelize the simulation of sixty-four buildings per node. The challenge to overcome in performing this task was to manage the big data output from the simulation and meet the high computational demands of the process. To achieve this, there was the need to (1) minimize the input-output (I/O) on the main lustre file system (a Linux, high-performance file system used for large-scale cluster computing) by running the application entirely in node memory and synchronously matching data between nodes and the lustre file system at the start and finish of each node's running, (2) ensure data management optimization by splitting the input data and arranging it by core and nodes to parallelize the task, and (3) reduce the total storage requirement by restricting data extraction from the compressed output, and examine its content to retrieve only the needed files [21].

The typical workflow for BEM involves generating OpenStudio and EnergyPlus models from several input building descriptors, then simulating each model using EnergyPlus, based on the representative city's Typical Meteorological Year (TMY) weather file for the building's climate zone. However, simulating 125.7 million U.S. buildings using HPC required that the workload be better managed. Consequently, the buildings were divided into the 9 U.S. census regions as well as California, as shown in the table below.

Table 3: Number of Buildings by U.S. Census Regions

Region	States	#Buildings
1	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont	5,435,392
2	New Jersey, New York, Pennsylvania	12,637,184
3	Illinois, Indiana, Michigan, Ohio, Wisconsin	22,528,155
4	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota	12,463,109
5	Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia	23,558,752

6	Alabama, Kentucky, Mississippi, Tennessee	9,977,403
7	Arkansas, Louisiana, Oklahoma, Texas	15,480,692
8	Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming	7,551,785
9	Alaska, Hawaii, Oregon, Washington	5,149,591
10	California	10,933,546

Each region in the table above represents a Comma Separated Values (CSV) file containing building I.D.s, building type, standard and height, among other descriptors in the data preparation process. Each CSV by region is used to generate input for the HPC jobs, which are submitted to the queue on a shared login node. Depending on the number of buildings in each region and the HPC configuration, the maximum size of the job to be submitted is determined. The next thing to determine is the number of buildings to be simulated by each core during the job, also called *churns*. Each region is treated separately to keep track of the processes so that issues can be easily identified and resolved. A batch (each job within a region) script calls the run generation code for each batch of jobs. The run generation code takes the number of nodes and the churns, and the input CSV file for the region, now split up so that each core gets a CSV file with just the correct number of buildings for the requested number of churns. Each node receives a tarball consisting of the split CSV files compressed together and saved to disk. The batch script also generates a job submission script which automatically adds the job to the queue [21].

Job configuration is an essential aspect of HPC. In generating and simulating building energy on Theta, the critical considerations included the determination of how much data (i.e. the number of buildings) to process. For 125 million buildings at 80% computing capacity on Theta (i.e. $3{,}514 nodes \times 64\ cores$), it is possible to run about 224,896 buildings in parallel. As experimental results showed that it takes a quarter of an hour (15 minutes) to simulate a building, based on an 80% compute capacity, it would take 139 hours ($\frac{125M}{224{,}896} \times 0.25hr$) to process all buildings. Another consideration is how long jobs wait in the queue, which is a factor of the number of nodes being utilized by a job and the maximum time the job can run, also called *walltime*. Node-intensive jobs would take time to complete before there is space for larger jobs to be run. For this consideration, it was determined that a *walltime* of 2 hours per job (i.e. eight churns or eight buildings per core within 2 hours of running a job) could be ideal in reducing the risk of having a lot of idle cores while waiting for some cores to finish working on complex buildings. The last consideration in the job configuration is resource optimization, to ensure that the greater number of cores are running when they should be running instead of being idle (waste of resources). Since processing times can vary widely between buildings, balancing out running cores and completing a building run is essential. One of the observations made in this project was that jobs that utilized 20-25% of Theta’s nodes waited in the queue for 4-5 days, whereas those that used 80% of the nodes queued for 6-7 days. Also, since there is a limitation on the number of jobs a single user can have in a queue, it was best to reduce the number of jobs that could be in the queue. Jobs must be split into chunks such that no one job is too small to run as a production job, that is, the theoretical minimum job size that can be run on not less than 128 nodes [21].

Parallelization starts on the compute node immediately after leaving the queue, but first, the simulation code and the required input data have to be initialized by copying them from the lustre file system and untarring (unzip) in the on-node memory. After the initialization, which happens on one core (single-threading), the job is then spread out on all cores on the node, with each core receiving a CSV file containing building information for the number of churns required. Next, the OpenStudio/EnergyPlus workflow is started to generate the building models and simulate each building's energy use and record any error logs that come up to the memory. Close to the end of the job's *walltime*, each core compresses its output into a tarball and copies it onto the lustre file system.

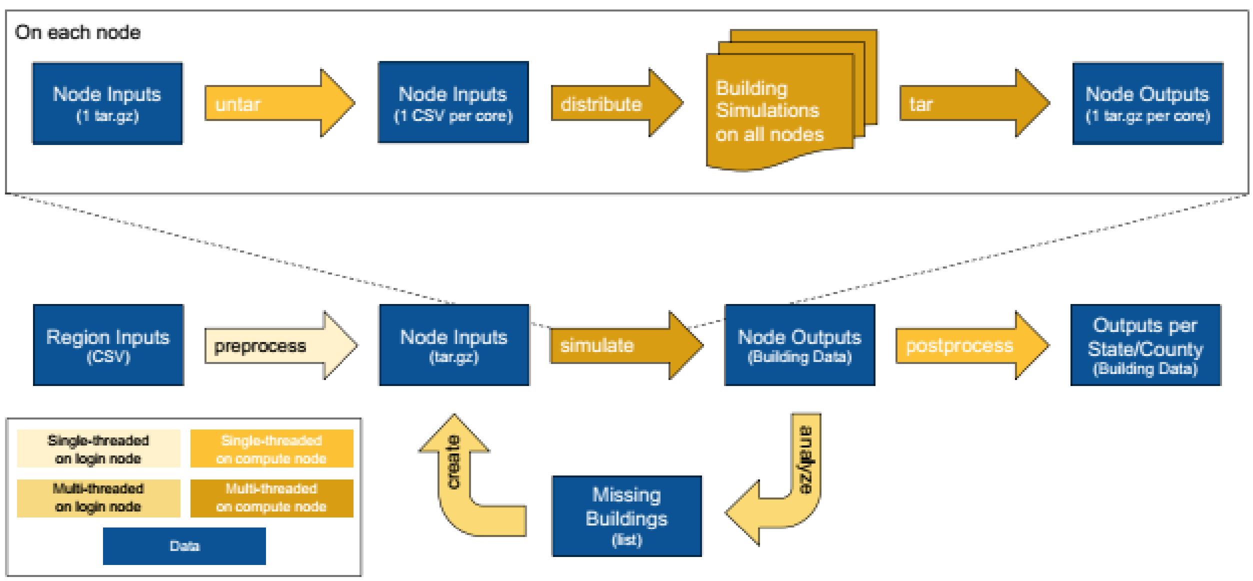

Figure 6: High-level Workflow for Nation-Scale Building Energy Modeling and Simulation with AutoBEM[21]

Due to constraints that compute resource optimization imposes on the ability of the system to complete its runs within an assigned *walltime*, a job process can often be completed with some buildings not simulated. Some factors that are responsible for that, as explained earlier, are the variability in buildings' processing times owing to building and geometry complexity and the number of building systems. While there are diverse ways of managing such faults, a redundancy-based approach was taken with this project. The redundancy strategy employed is that the initial job run starts with eight churns per core and then half the churn in the next job resubmission for buildings that could not complete their run in the previous job. Suppose during a particular iteration at a prevailing run, there are not enough buildings for a full job (3,514 nodes). In that case, those missing buildings are added to a pool of buildings (labelled "Remainder" in Figure 7) to reduce the job size in a queue. If there are enough unprocessed buildings from a run to constitute a production run, they will be added to the next job iteration.

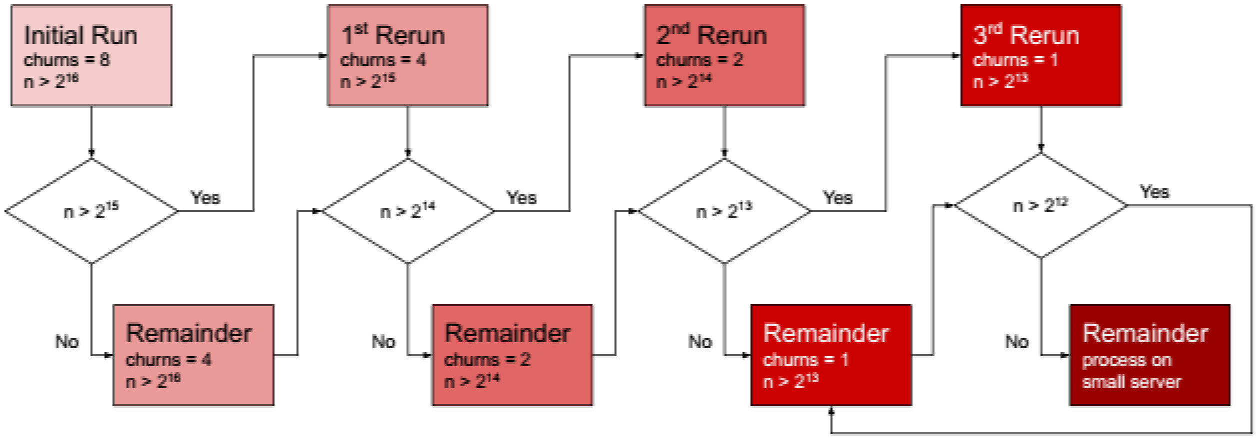

Figure 7: iterative Process to Handle Incomplete Jobs [21]

The iterations for each run at the regional level, indicated by the top-row rectangles in Figure 7, do not increase in size as unprocessed buildings are not added to the next run. However, the remainder pool, indicated by the bottom-row rectangles, gets larger because it keeps receiving more input whenever a run is not completed fully. In both cases, the churn-reducing, iterative processes remain the same. Since the walltime of 2 hours remains the same for each iteration of run, jobs which were not completed in a previous run will have more time to run in the next. Through this process, buildings get twice as much time to run in subsequent iterations compared to the previous ones.

The initial run, involving 92 jobs and 8 churns, yielded an 88.4% completion or processing rate of all buildings submitted to it. The 1st rerun consisted of 21 jobs and 4 churns, and completed 34.3% of the buildings that were submitted in that run. The 2nd rerun of 21 jobs with only 2 churns had an 86.2% success rate. The last run or 3rd rerun uses a single churn. Of the 4 jobs that were submitted, the completion rate ranged between 20.1% and 92.1%.

Table 4: Result of Simulated Buildings in each Iteration by Region

#	Buildings	Initial	Rerun1	Rerun2	Rerun3
1	5,435,392	4,362,202	518,576	291,327	207,103
2	12,637,184	9,019,214	1,358,741	1,821,282	
3	22,528,155	22,467,370			
4	12,463,109	10,748,054	791,385	820,444	
5	23,558,752	23,207,301			
6	9,977,403	4,517,672	1,442,752	3,046,398	97,638

7	15,480,692	15,452,216			
8	7,551,785	7,531,113			
9	5,149,591	5,051,863			
10	10,933,546	10,781,507			
R			742,174		
%	125,716,992	113,138,512	4,853,628	5,979,451	305,741
		89.99%	3.86%	4.76%	0.16%

The output from this process was worth over 80 T.B. of data on the lustre filesystem. The postprocessing mechanism adopted to manage the data involved extracting only the content that was certainly necessary to analyze. All this is done without untarring the entire content of each node’s tarball output because untarring all outputs at once would result in over 628.6 million files on disk and cause the system to run into issues. This postprocessing analysis made use of *tarfile* and *multiprocessing* libraries in Python. Completed buildings would contain an EnergyPlus file in IDF format named after the building I.D., while incompletely processed or unprocessed buildings would not contain such files.

CHAPTER THREE: DATA VISUALIZATION

As shown in Chapter 2 under Data sources, building energy modeling involves huge numbers of building descriptors. Moreover, running the simulation generates time series data of building energy performance in terms of electricity and natural gas consumption. Depending on the time intervals (or timesteps) and duration within which the performance is measured, thousands or tens of thousands of energy performance data could be generated. These could be aggregated to provide valuable insight on energy performance or demand by month, floor area, zone, load, or end-use. Also, by applying the right conversion factors, these data could be utilized to estimate the energy cost and carbon footprint of a building's energy consumption, as well as savings estimates from identified Energy Conservation Measures (ECMs).

The amount of information to digest from a simulation's output could be overwhelming for the untrained mind to handle and appreciate. To meet the goal of making building energy efficiency data more accessible and adoptable by users for their education and implementation of recommended ECMs, data must get to them in processed, interactive, exciting and easy-to-understand formats. Several works have utilized the energy consumption data in comma-separated value (csv) format to give a visual representation of building energy performance in the form of tables and various statistical and engineering charts such as flowcharts, pie charts, scatter plots, bar graphs, line graphs, histograms, time plots, box and whisker [37][38]. While these kinds of visualization could be enough for single buildings, solely using such visualization approaches for energy performances of clusters of buildings at the urban scale could inundate a user's view.

A more tolerable approach toward urban-scale building energy data visualization is to provide a means to allow users to choose buildings or areas of interest whose data they would like to view, with data filtering functions for easy identification of relevant buildings. One such platform is CesiumJS, an open-source JavaScript library for creating interactive 3D globes and maps for sharing dynamic geospatial data [39]. Cesium may be installed using the node package manager (npm) in Node JS or downloaded and saved in a local directory [40]. The feature of CesiumJS that is most relevant to demonstrating the visualization of energy data is 3D Tiles, which allows the streaming, styling and interaction with 3D buildings and photogrammetry using CesiumJS 3D Tiles. CesiumJS has a web-based live-coding app that can be used to view CesiumJS examples called Sandcastle. On CesiumJS Sandcastle [41], one can find a gallery of examples of CesiumJS features together with their JavaScript, HTML and CSS codes. A few of the CesiumJS features which may be utilized in the demonstration of urban-scale building energy data include 3D Tiles Feature Picking [42], 3D Tiles Feature Styling [43], and Camera Options [44], among others. These features may be applied to select buildings based on specific building properties such as building materials, building type, or age, among others. The camera options feature also allows users to navigate buildings based on geolocation data. Moreover, there are features that allow the display of building metadata by hovering around or clicking on a building. Figure 8 and Figure 9 below show screenshots of the features of CesiumJS on Sandcastle.

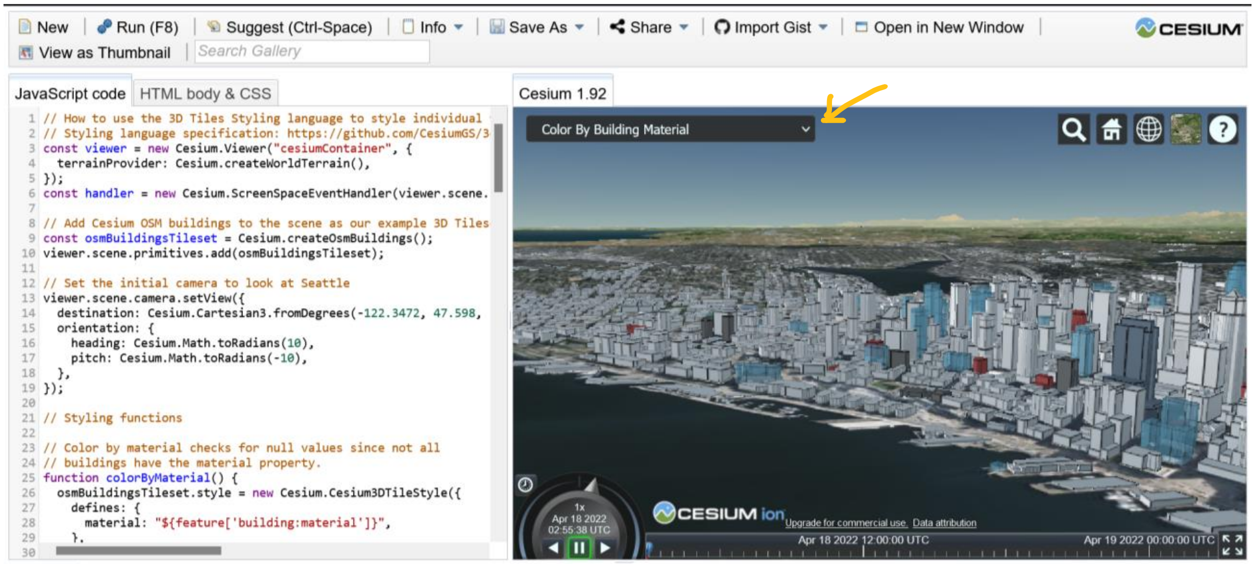

Figure 8: 3D Tiles Feature Styling on Cesium Sandcastle[43]. The yellow arrow at the top of the image on the right points to the dropdown button which displays the feature styling options to choose from.

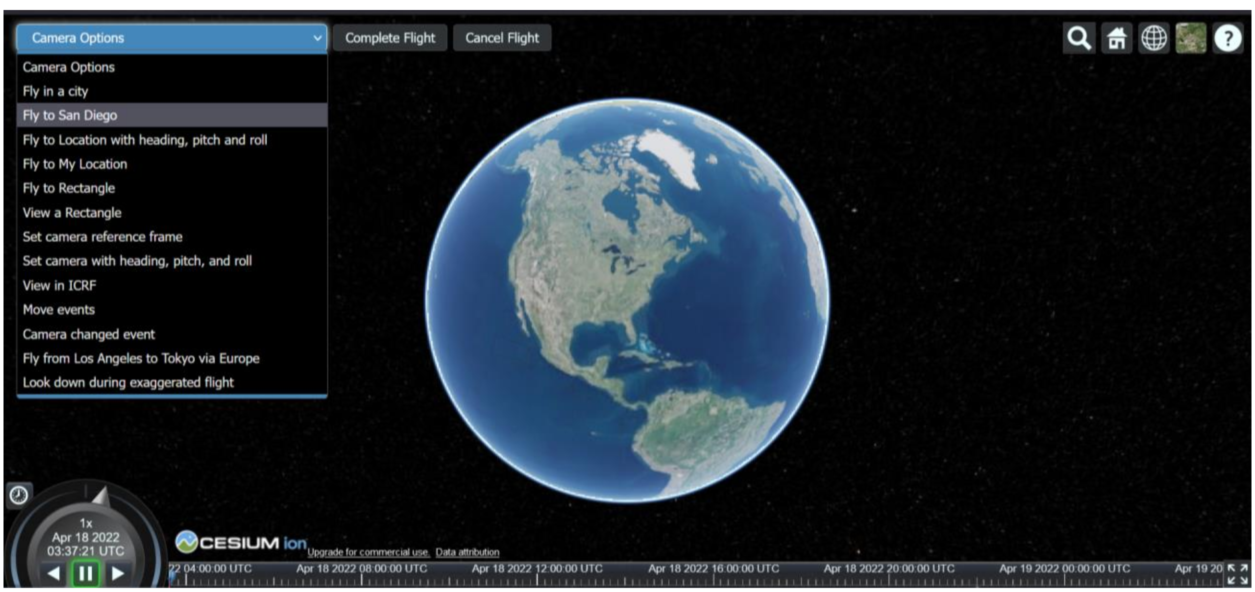

Figure 9: Camera option Features on Cesium Sandcastle[44]. The dropdown button at the top left corner displays options that could be added to the camera views.

To demonstrate Cesium visualization with some building energy data, we make use of two sample neighborhoods (Soapstone and Waterfront) of buildings in Washington, D.C. The initial camera view (the area which is loaded onto the screen when the page starts) was set to Waterfront. The Cesium container has all the standard viewer widgets options in the upper right corner of the page such as address search bar, Home, base layer selector with imagery and terrain options and instructions on how to navigate the globe. There is a mouseover function that displays building IDs upon hovering over a building and also pops out a table of building metadata when a building is left-clicked. The demonstration also includes two toolbars at the top left corner with color filtering and camera functions. A screenshot of the demonstration is shown in Figure 10. The interactive features can be explored by following the link [here](#).

Figure 10: CesiumJS Visualization Demonstration of Virtual DC [45].

References

[1] J.-P. Tricoire, "Buildings are the foundation of our energy-efficient future," *World Economic Forum*. https://www.weforum.org/agenda/2021/02/why-the-buildings-of-the-future-are-key-to-an-efficient-energy-ecosystem/ (accessed Mar. 24, 2022).

[2] "Buildings – Topics," *IEA*. https://www.iea.org/topics/buildings (accessed Mar. 24, 2022).

[3] "Buildings," *Alliance to Save Energy*. https://www.ase.org/categories/buildings (accessed Mar. 24, 2022).

[4] "About the Building Technologies Office," *Energy.gov*. https://www.energy.gov/eere/buildings/about-building-technologies-office (accessed Mar. 25, 2022).

[5] "Buildings – Sustainable Recovery – Analysis," *IEA*. https://www.iea.org/reports/sustainable-recovery/buildings (accessed Apr. 23, 2022).

[6] C. J. Bell, S. Sienkowski, and S. Kwatra, "Financing for Multi-Tenant Building Efficiency: Why This Market Is Underserved and What Can Be Done to Reach It," p. 55.

[7] "Global Cooling Prize," *Global Cooling Prize*. https://globalcoolingprize.org/ (accessed Mar. 25, 2022).

[8] "Building Technologies Office Multi-Year Program Plan." Department of Energy. [Online]. Available: https://www.energy.gov/sites/prod/files/2016/02/f29/BTO%20Multi-Year%20Program%20Plan%20-%20Final.pdf

[9] "About the Office of Electricity," *Energy.gov*. https://www.energy.gov/oe/about-office-electricity (accessed Mar. 25, 2022).

[10] "In 2018, U.S. energy expenditures increased for the second consecutive year," Sep. 10, 2020. https://www.eia.gov/todayinenergy/detail.php?id=45076 (accessed Mar. 26, 2022).

[11] "Energy as a Service Market Size to Reach USD 112.7 Billion by 2030 at CAGR 7.6% - Valuates Reports," Jan. 05, 2022. https://www.yahoo.com/now/energy-market-size-reach-usd-163000368.html (accessed Mar. 26, 2022).

[12] "Energy Efficient Devices Market Size USD 1,771.70 Billion by 2028 | BEMS Industry Growth 12.5% CAGR." https://www.emergenresearch.com/industry-report/energy-efficient-devices-market (accessed Mar. 26, 2022).

[13] D. Thorpe, "Market for Energy Efficiency Is $360 billion/yr and Bigger Than Renewables | Smart Cities Dive." https://www.smartcitiesdive.com/ex/sustainablecitiescollective/market-energy-efficiency-360-billionyr-and-bigger-renewables/1001556/ (accessed Mar. 26, 2022).

[14] S. Nadel, "How can we increase energy efficiency investments? Here are 10 suggest," Aug. 15, 2017. https://www.aceee.org/blog/2017/08/how-can-we-increase-energy-efficiency (accessed Apr. 02, 2022).

[15] A. Patil, "The Benefits of Early-Stage Energy Modeling," Jan. 07, 2022. https://www.ny-engineers.com/blog/the-benefits-of-early-stage-energy-modeling (accessed Mar. 31, 2022).

[16] J. R. New *et al.*, "Automatic Building Energy Model Creation (AutoBEM) for Urban-Scale Energy Modeling and Assessment of Value Propositions for Electric Utilities," Jun. 2018, [Online]. Available: https://www.osti.gov/biblio/1474682

[17] J. E. McMahon and S. Wiel, "Energy-efficiency labels and standards: A guidebook for appliances, equipment and lighting," LBNL--45387, 836221, Feb. 2001. doi: 10.2172/836221.

[18] "Business Model Analyst - Business Model Canvas Examples and Analysis." https://businessmodelanalyst.com/ (accessed Apr. 24, 2022).

[19] J. R. New, “AutoBEM Capabilities for Creating a Digital Twin of a Utility,” in *Seminar 14 - Multiscale Building Energy Modeling, Part 10*, Atlanta, Georgia, vol. Part 10. [Online]. Available: http://web.eecs.utk.edu/~jnew1/presentations/2019_ASHRAE_MBEM10_ORNL.pdf

[20] P. Ruyssevelt, “Modelling London’s Building Stock and Its Associated Energy Use,” in *Multiscale Building Energy Modeling*, Jan. 2019, p. 30. [Online]. Available: http://web.eecs.utk.edu/~jnew1/presentations/2019_ASHRAE_MBEM9_UK.pdf

[21] A. S. Berres, B. C. Bass, M. B. Adams, E. Garrison, and J. R. New, “A Data-Driven Approach to Nation-Scale Building Energy Modeling,” in *2021 IEEE International Conference on Big Data (Big Data)*, Orlando, FL, USA, Dec. 2021, pp. 1558–1565. doi: 10.1109/BigData52589.2021.9671786.

[22] Y. Chen and T. Hong, “Impacts of building geometry modeling methods on the simulation results of urban building energy models,” *Applied Energy*, vol. 215, pp. 717–735, Apr. 2018, doi: 10.1016/j.apenergy.2018.02.073.

[23] Y. Chen and T. Hong, “Impacts of Building Geometry Simplification on Energy Simulation Results of Urban Building Energy Models,” presented at the 2018 ASHRAE Annual Conference, Houston, Texas, Jun. 2018.

[24] Office of Energy Efficiency and Renewable Energy, “Prototype Building Models.” https://www.energycodes.gov/prototype-building-models

[25] X. Luo, D. Macumber, J. R. New, and R. Judkoff, “Seminar - Multiscale Building Energy Modeling,” in *Multiscale Building Energy Modeling*, Atlanta, Georgia, Jan. 2019, vol. 10. [Online]. Available: http://web.eecs.utk.edu/~jnew1/presentations/2019_ASHRAE_MBEM10_NREL.pdf

[26] Microsoft®, *Microsoft Building Footprint*. [Online]. Available: https://www.microsoft.com/en-us/maps/building-footprints

[27] “Cartographic Database - PrintWiki.” http://printwiki.org/Cartographic_Database (accessed Apr. 01, 2022).

[28] P. Gong *et al.*, “Annual maps of global artificial impervious area (GAIA) between 1985 and 2018,” *Remote Sensing of Environment*, vol. 236, p. 111510, Jan. 2020, doi: 10.1016/j.rse.2019.111510.

[29] B. C. Bass, “What data sources can you use to do modeling at the urban scale and beyond?” *Ask A Modeler*, Dec. 21, 2021. https://www.ibpsa.us/ask-a-modeler-december-2021/ (accessed Mar. 12, 2022).

[30] J. Yuan, J. R. New, J. Sanyal, and O. Omitaomu, “Urban Search Data Sources,” ORNL, Oak Ridge, TN, ORNL Internal Report ORNL/TM-2015/397, Jul. 2015.

[31] S. Nagpal and C. F. Reinhart, “A comparison of two modeling approaches for establishing and implementing energy use reduction targets for a university campus,” *Energy and Buildings*, vol. 173, pp. 103–116, Aug. 2018, doi: 10.1016/j.enbuild.2018.05.035.

[32] X. Li, “Scalability: strong and weak scaling,” *PDC Blog*. https://www.kth.se/blogs/pdc/2018/11/scalability-strong-and-weak-scaling/ (accessed Mar. 15, 2022).

[33] “Scaling tutorial,” *HPC Wiki*. https://hpc-wiki.info/hpc/Scaling_tutorial (accessed Mar. 15, 2022).

[34] C. S. Monteiro, C. Cerezo, A. Pina, and P. Ferrão, “A METHOD FOR THE GENERATION OF MULTI-DETAIL BUILDING ARCHETYPE DEFINITIONS: APPLICATION TO THE CITY OF LISBON,” p. 6, 2015.

[35] C. F. Reinhart and C. Cerezo Davila, “Urban building energy modeling – A review of a nascent field,” *Building and Environment*, vol. 97, pp. 196–202, Feb. 2016, doi: 10.1016/j.buildenv.2015.12.001.
[36] “Theta/ThetaGPU | Argonne Leadership Computing Facility.” https://www.alcf.anl.gov/alcf-resources/theta (accessed Mar. 16, 2022).
[37] M. Buevich, A. Rowe, and R. Rajkumar, “SAGA: Tracking and Visualization of Building Energy,” in *2011 IEEE 17th International Conference on Embedded and Real-Time Computing Systems and Applications*, Toyama, Japan, Aug. 2011, pp. 31–36. doi: 10.1109/RTCSA.2011.83.
[38] Y. Chen, X. Liang, T. Hong, and X. Luo, “Simulation and visualization of energy-related occupant behavior in office buildings,” *Build. Simul.*, vol. 10, no. 6, pp. 785–798, Dec. 2017, doi: 10.1007/s12273-017-0355-2.
[39] “CesiumJS,” *Cesium*. https://cesium.com/platform/cesiumjs/ (accessed Apr. 17, 2022).
[40] “Downloads,” *Cesium*. https://cesium.com/downloads/ (accessed Apr. 17, 2022).
[41] “Cesium Sandcastle,” *Cesium*. https://sandcastle.cesium.com/index.html (accessed Apr. 17, 2022).
[42] “Cesium Feature Picking.” https://sandcastle.cesium.com/gallery/3D%20Tiles%20Feature%20Picking.html (accessed Apr. 17, 2022).
[43] “Cesium 3D Tiles Feature Styling,” *Cesium*. https://sandcastle.cesium.com/?src=3D%20Tiles%20Feature%20Styling.html (accessed Apr. 17, 2022).
[44] “Cesium Camera Options.” https://sandcastle.cesium.com/gallery/Camera.html (accessed Apr. 17, 2022).
[45] “Virtual DC.” http://volweb2.utk.edu/~camoo/ (accessed Apr. 23, 2022).

Appendices

Appendix 1: Customer Discovery Questions for ESCOs and Utilities

Understanding the Customer

- What do you do professionally and what's your role at the company?
- How important is energy efficient buildings to your operations?
- What are some building-specific energy needs you have?
- What specifically about building energy use within your utility service areas is critical to your company?

Understanding the Problem

- Do you find it hard to generate energy models in a short time, at a low cost, and without needing so much data from the building occupants?
- How have you managed to regulate building energy use within your service areas as a means of meeting normal or peak demands?
- Are they any shortfalls in utility program formulation, ensuring grid resiliency, demand-side management or building energy efficiency?
- How important is data, time of simulation, cost, and reliability of the models to the work you do?
- How motivated are you to solve/improve the challenges in your work?
- If you had a solution to this problem, what would it mean to you/how would it affect you?

Suitability of a Plausible Solution

- Beyond individual or a few collection of building energy models, have you considered large-scale building energy models at the city, state or national levels?
    - If no, why not? If yes, how might they be of help?
- What do you think of a solution or software the enables the automatic detection and creation of building energy models, leveraging several data sources and high computational abilities? Would it solve your problem?
- Would you ever use such a solution? Why? Why not?
- What would you like to see in an [ideal] building energy modelling solution or utility-scale energy modeling solution that meets your current/future needs?
    - Would you be willing to start using this right away?
    - Would you be willing to pay for the product? How much are you willing to pay?

Improving the product or the idea

- What will motivate you to continue using this product?
- What are the barriers that would prevent you from using this product?