2020 Building Performance Modeling Conference and
SimBuild co-organized by ASHRAE and IBPSA-USA

URBAN-SCALE ENERGY MODELING: SCALING BEYOND TAX ASSESSOR DATA

Joshua New[1], Mark Adams[1], Eric Garrison[2], Brett Bass[2] and Tianjing Guo[1]
[1]Oak Ridge National Laboratory, Oak Ridge, TN
[2]The University of Tennessee, Knoxville, TN

ABSTRACT

In an attempt to attain building-specific characteristics for urban-scale building energy models, county-specific tax assessors' data is often an initial data source. This data source can contain valuable information such as year built, area, height, HVAC type, and roof/wall descriptions. We will show examples of 2,000 fields from Hamilton County in Tennessee with examples of many fields which are not relevant to urban-scale building energy modeling, are incorrect compared to other data sources, and highlight some lessons learned working with such data.

There are currently 3,142 counties in the United States, each with their own data format, field definitions, and data access policy. As urban-scale involves city-scale analysis potentially covering multiple counties and matures toward state- or nation-scale analysis, county-by-county approaches are not scalable. While there are efforts to unify these datasets, there is an increasing proliferation of data and algorithms that cover wider areas and provide more accurate inputs for urban-scale models. This paper summarizes computer vision of imagery, cartographic layers, building type assessment, and model generation used to achieve scalable detection and analysis of buildings.

INTRODUCTION

Due in part to increasingly available data, ubiquitous computing, and open source software, the nascent field (Reinhart and Davila 2016) of urban-scale building energy modeling has quickly grown to become a dominant topic in related simulation conferences. As an example, IBPSA's Building Simulation 2019 accepted 762 research papers of which 90 (11.8%) fell under "simulation at urban scale,"In addition, ASHRAE has been host to 11 Urban/Multiscale building energy modeling seminars since 2016. Previous urban-scale building energy modeling efforts include insights for CO_2 emissions (Parshall et al. 2010), heating energy demand forecasting (Strzalka et al. 2011), city-scale building retrofit (Chen, Hong, and Piette 2017), and creating a digital twin of a city utility (Copeland and New 2019).

While urban-scale energy modeling has increased in popularity and established increasingly mature applications, it still suffers from a lack of best practices and limited scalability due to the prevalent use of geographically-limited data sources such as county-specific tax assessor's data. The purpose of this publication is to share an approach to urban-scale energy modeling that has been demonstrated for a multi-county utility, delineate data sources and algorithms that facilitate scalability toward nation-scale energy modeling, and might be useful reference toward the establishment of best-practices.

In efforts to reduce an energy burden on a nation's economy, create a more sustainable built environment, foster a resilient critical infrastructure, and develop more flexible electrical distribution networks, our team attempts to facilitate grid-interactive efficient buildings by enabling practical, simulation-informed targeting, prioritization, programmatic, and operational decisions for electrical distributors. Specifically, we have partnered with the Electric Power Board of Chattanooga, TN (EPB). As with most utilities, their geographical extent is beyond an individual county, extending to approximately 600 mile^2 covering parts of eight counties in East Tennessee and Georgia. As such, the authors identified scalable data sources and algorithms that allowed creation of accurate building energy models with less effort than would be required to merge tax data into an irreplicable digital twinning effort. The data sources and algorithms, which we collectively refer to as "Automatic Building detection and Energy Model Creation (AutoBEM)," has been used to create 178,368 distinct OpenStudio and EnergyPlus models for every building in EPB's service territory. The models have since quantified energy, demand, emissions, and cost-reductions under nine monetization scenarios for the utility and is being used to inform programmatic rollout of energy efficiency, demand management, product/service lines, and new business models. In addition, EPB provided 15-minute whole building electricity data from 178,368 premises. Building-specific measured data allowed empirical validation of urban-scale models and tests to establish which combination of data and algorithms tended to provide the most accurate models.

This paper reviews traditional data sources for urban-scale modeling (e.g. tax assessor data) followed by a methodology used to compare and contrast more scalable data sources for overcoming the geographical, format differences, and access limitations of more traditional urban-scale data sources. As such, a description of the full algorithms to turn these data sources into building descriptors and the software workflow of how to turn such descriptors into a simulatable building energy model are beyond the scope of this paper.

TAX ASSESSOR DATA

There are some countries which maintain centralized data on each building in the country. These data sources are often not available for use due to privacy concerns. The United States tends to have data at each county. The primary reason? Tax law. While there are 22 countries in the world which do not have a property tax, the overwhelming majority do. Specifically, the property tax in the United States is an ad valorem tax (Latin for "according to value") based on the fair market value of the property times an assessment ratio times a tax rate. Property tax rates vary from 0.27% to 2.4%. These properties are often tracked as parcels (Figure 1).

Figure 1: Most counties manage parcels. Partial geo-registration of a building is possible since most building are registered to a single parcel. However, this is often non-trivial and a scalability challenge due to non-uniform file formats and field names between counties.

Such tax law necessitates detailed characterization of buildings at a local level since property value is dependent upon the property, building, permanent improvements, and is highly geographically non-linear (i.e. location, location, location). For this reason, the 3,142 counties in the United States collect property information (see Table 1) which overlap the input fields needed to construct a building energy model of each building. Unfortunately, this data is currently collected with little-to-no standardization regarding number of fields collected, definition of building/property descriptors, or data format (e.g. database, CSV, PDF). While most of these are publicly-available due to governmental taxation transparency requirements, the data access policy and approval process is highly variable.

In an effort to be clear about tax assessor data which was considered for this study and contrasted with more scalable energy modeling approaches, we provide summaries with screenshots and statistics to objectively characterize the county-specific data.

Table 1: Typical building-related fields from tax assessor's data for a county in Tennessee. Several fields could improve accuracy of a building energy model.

Column	Field Description	Data Type
1	Parcel ID	Alpha
2	Exterior Type Code	Alpha
3	Exterior Type Description	Alpha
4	Jurist Code	Alpha
5	Jurist Description	Alpha
6	Year Built	Numeric
7	Taxable Building Amount	Numeric
8	Size Adjusted Area	Numeric
9	Story Height	Alpha
10	Roof Structure Code	Alpha
11	Roof Structure Description	Alpha
12	Roof Cover Code	Alpha
13	Roof Cover Description	Alpha
14	Prime Wall Code	Alpha
15	Prime Wall Description	Alpha
16	Second Wall Code	Alpha
17	Second Wall Description	Alpha
18	Heat Type Code	Alpha
19	Heat Type Description	Alpha
20	Account Number	Numeric
21	Card Number	Numeric
22	Street Number	Alpha
23	Street Name	Alpha
24	Land Use Code	Alpha
25	Land Use Description	Alpha
26	City	Alpha

Data Overview

This section provides redacted examples of tax assessor's data from a county in Tennessee in an attempt to highlight uses and challenges of such data for urban-scale energy modeling. Received files included:

- GIS parcel data - shapefiles (*.shp and supporting files) which contain parcels but no building information. Parcel data was available in an online format, but did not allow export for large areas.
- Buildings parcel data - zip file (*.zip) of building specific data. As this data was not available online, it required contact with the GIS administrator.
- Overview letter - listing of data fields (*.pdf) with instructions to run an executable and entry into DOS prompt to extract 300MB of data into a space-delimited text file.
  - Field names/lengths - list of 1,999 property fields with start/end/length records but no definitions/descriptions.

Table 2: This statistical summary of tax assessor data shows the top 10 most frequently-used values from 8 tax assessor data fields related to buildings values with percentage of occurrences out of 139,161 entries. Land Use codes described in text.

Size Adjusted Area	**%**	**Story Height**	**%**	**Land Use code**	**%**	**Heat Type Description**	**%**
1,000-1,499	32.6	1	74.5	RESID	82.3	CENTRL HEAT&	75.4
1,500-1,999	22.1	2	17.4	COMM	7.9	<EMPTY>	12.6
2,000-2,499	12.2	1.5	7.1	MFG	4.7	GRAVITY	7.4
5,000+	9.8	3	0.6	IN	2.4	NO HVAC	3.4
500-999	8.3	>7	0.6	AG	1	FORCED HOT A	0.9
2,500-2,999	6.6	2.5	0.1	EX	0.8	GHA	0.1
3,000-3,499	3.8	4	0.1	DU	0.4	CENTRAL A/C	0.1
3,500-3,999	2.1	5	0.0	EID	0.2	REV CYCLE UN	0.0
4,000-4,499	1.3	6	0.0	RLS	0.1	CENT HEAT &	0.0
4,500-4,999	0.8	7	0.0	BCMT	0.1	NONE	0.0
Roof Structure Description	**%**	**Roof Cover Description**	**%**	**Decade**	**%**	**Prime/Second Wall Description**	**%**
HIP/GABLE	86.0	SHINGLE ASPH	81.9	2000	13.7	<EMPTY>	41.9
WOOD RAFTERS	2.8	SHEET METAL	3.5	1960	13.6	WOOD FR W SH	15.5
BAR JOISTS	2.4	BUILT-UP	3.4	1970	13.5	VINYL	13.5
OPEN STEEL S	2.1	METAL	2.9	1990	11.6	BRICK	13.1
STEEL TRUSS	1.8	<EMPTY>	1.7	1950	11.4	WOOD FR ASBT	2.7
<EMPTY>	1.7	ASPHALT SHIN	1.7	1980	10.8	CONC BLK PLA	1.6
NONE	0.8	CORRUGATED M	1.3	1940	7.6	ALUMINUM	1.4
WOOD TRUSS	0.6	NONE	0.9	2010	6.9	HARDIE BOARD	1.4
FLAT/SHED	0.5	ROLL COMP	0.6	1930	4.8	BRICK VENEER	1.2
GAMBREL	0.3	BUILT UP T &	0.4	1920	3.8	CORRUGATED M	1.2

  - Property code/type - 1-4 word abbreviated description of the property
  - Sales code/type - property transaction type
  - District code/type - name of location
  - Land Use codes - detailed break down of 159 code categorizations for different building uses/types.

- Building data - list (*.csv) of building data. This was attained after significant coordination through the municipal utility and we show the top 10 most common values for these fields in Table 2. We identify some of the most valuable fields for building energy modeling along with some examples of their building-specific values:
  - Size adjusted area - often treated as conditioned square feet
  - Story height - number of floors
  - Land Use - residential, duplex, multi-family. These rarely correspond to canonical reference/protoype buildings most used by energy modelers.
  - Heat type description - type of heating unit. While valuable, this field was most often blank, “Central” (different names were used for this same classification), “no HVAC”, or “gravity.”
  - Roof structure - usually hip/gable, sometimes Gambrel
  - Roof cover -shingle, metal, built-up roof
  - Prime and Second Wall - wood, brick, vinyl

Tax assessor data has the potential to provide many fields relevant to creating a more accurate building energy model. However, this data is not scalable in several ways related to geographical area (in the United States, different for over 3,000 counties), data differences (fields collected and abbreviations vary), format differences (often provided as several files in PDF and/or CSV format), access limitations (special permission from an established stakeholder relationship may be required), and lack of availability for software retrieval (rarely an API for data processing). To overcome these limitations, a methodology is shown for comparing and contrasting more scalable data sources.

SCALABLE DATA CONSIDERATIONS

In an effort to overcome scalability challenges for a digital twin of a utility covering eight counties, we describe several categories of data sources and algorithms the team considered (Yuan et al. 2015; New et al. 2018).

1. Imagery (satellite, airborne) – computer vision can be applied for extraction of building footprints.
2. Elevation data – LiDAR and computationally-derived Digital Elevation Models (DEM) can be used to determine building height and number of floors using heuristics.
3. Geometry data – computationally-derived 3D tessellation or photogrammetry of buildings.
4. Cartographic data – Geographic Information System (GIS) analysis layers can be used to inform zoning for building types, critical facilities (e.g. hospitals), and other properties.
5. Street-level Imagery – computer vision can potentially extract higher-resolution details of buildings windows and façade type, but currently suffers from a lack of robust techniques.
6. Building information databases – this meta-category can include tax assessor's data or any building-specific information, as is common in Multiple Listing Service (MLS) data used for real estate sales.

For each of these, a "comparison matrix" was created which allows side-by-side comparison of datasets. There was significant iteration and debate over what dimensions should be considered. We provide our final list of data source considerations in hopes it saves others the debate:

1. Title – short label for referring to the dataset
2. Summary – short description of the data
3. Data type – the format in which the information is stored (usually image, database, or computationally derived from multiple data sources)
4. Company – name of the organization that makes the data available
5. Website – hyperlink to the most pertinent information necessary for using this dataset
6. Temporal resolution – how often the datasets are collected (e.g., 25 years)
7. Spatial resolution – the dimensions of the data (e.g., 1 km^2 per pixel)
8. Measure accuracy – information available regarding the accuracy of the database based on input sources or sensor calibration
9. Cost – any initial or recurring costs required to access/retain rights to the data
10. Format – the standard file format in which the datasets are stored
11. Mapping to building input variables – indicates whether these datasets are useful in identifying properties necessary or useful for constructing a software model of a building (e.g., building type, square footage, window-to-wall ratio, façade material type, façade material thickness, façade material density)
12. Mapping to area properties – indicates whether these datasets are useful in segmenting area type (e.g., buildings, roads, open/vegetated spaces)
13. Mapping to material properties – indicates whether these datasets are useful in determining material types (e.g., concrete, brick, soil, gravel, asphalt, granite)
14. Coverage of United States (US) – indicates the extent to which the data provided are local versus national
15. Orientation – where relevant, the general view from which the data were taken (e.g., street view, single side of a building, multiple sides of building, perspective, oblique)
16. Existing internal software – does the current team have software capabilities that leverage this dataset for purposes that could be synergistically leveraged for this project
17. Existing expertise – does the current team have any unique knowledge or skills that would be vital to the successful use of the data for this project
18. Restrictions – what are the limitations on the use of the data (e.g., legal/privacy ratings, number of Application Program Interface [API] calls per day)
19. Comments – any major observations about the data that do not fit in the previous categories

While the original analysis covered 37 data sources, a comparison methodology is shown for only a few examples - satellite imagery, aerial imagery, elevation data, street-level imagery, and geological information (see Figures 2-5).These were used successfully by the team to create a digital twin of a utility with evidence that these might stimulate scalability beyond the urban context.

	DigitalGlobe Standard Imagery	DigitalGlobe Precision Aerial image
Summary	Satellite imagery including panchromatic and multispectral images (4 bands or 8 bands)	Aerial imagery, including panchromatic and multispectral images
Data type	Image	Image
Company	DigitalGlobe	DigitalGlobe
Website	www.digitalglobe.com	www.digitalglobe.com
Temporal resolution	N/A	N/A
Spatial resolution	Pan: 0.5/0.6 m; MS 2.0/2.4 m	0.3 m
Measure accuracy	High	High
Cost	Pan: $24 per sq. km; Pan+MS $27 per sq. km	$11 per sq. km; Pricing URL
Format	GeoTiff	GeoTiff
Building inputs	Building footprint	Building footprints
Area properties	Vegetated areas, road surface, buildings, parking lots	Vegetated areas, road surface, buildings, parking lots
Material properties	Road pavement materials (e.g., concrete, asphalt), parking lots (e.g., gravel, soil)	Road pavement materials (e.g., concrete, asphalt), parking lots (e.g., gravel, soil)
Coverage of US	High	Over 10 million km^2 of coverage of the contiguous US
Orientation	Aerial	Aerial
Existing expertise	Remote sensing data analysis tool	Remote sensing data analysis tool
Restrictions	Contract-specific	Contract-specific

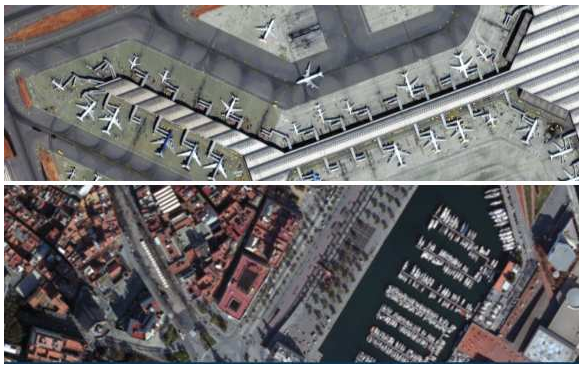

Figure 2: DigitalGlobe provides satellite imagery and higher-resolution aerial imagery, which can include multi-spectrum, for determining building footprints or material types.

Satellite and Aerial Imagery

'Satellite imagery tends to be high-resolution (relative to buildings), has high positional accuracy, global coverage, low cost, and multiple bandwidths allow multi-spectral disaggregation of material types more easily. Major disadvantages is insufficient algorithms for material classification accuracy, susceptibility to weather (e.g. cloud cover) and lighting conditions, and advances in computer vision are needed to make extensive use of this data.

Aerial imagery has very high resolution and texture information that's better for identifying materials. Major disadvantage is that objects have larger inner-class variations and hence are more difficult to extract.

National Elevation Data

Widely-available and free elevation data can allow the extraction of 2D building footprints into 3D building geometries for above-ground floors. One challenge is that resolution varies across different regions.

Street-level Imagery

Street-level imagery holds great promise for enabling urban-scale knowledge to inform building energy models and several related applications. Camera parameters are available (including both intrinsic and extrinsic parameters) to enable projection of options with geo-coordinates onto images. Major disadvantages include the manual labor or advanced algorithms necessary to reliably extract high-level information and extrinsic camera parameters have location-specific errors.

Cartographic Data

The United States Geological Survey (USGS) Earth Explorer is an example tool which provides various remote sensing (satellite or airborne imager, LiDAR), with most datasets free, and relatively convenient user interface for searching or downloading images. Major disadvantages include coverage varying significantly across datasets and very few images are from commercial satellites.

	National Elevation Dataset
Summary	Ground elevation data
Data type	Raster
Company	USGS
Website	http://ned.usgs.gov/
Temporal resolution	N/A
Spatial resolution	1/3, 1, and 2 seconds of arc; 1/9 arc-second and 1 meter for some areas
Measure accuracy	Mean square error is 1.55 m
Cost	Free
Format	Raster data
Building inputs	Main floor ground elevation
Area properties	Road surface elevation
Material properties	N/A
Coverage of US	High
Orientation	Aerial
Existing expertise	GIS software
Restrictions	Restrictions URL

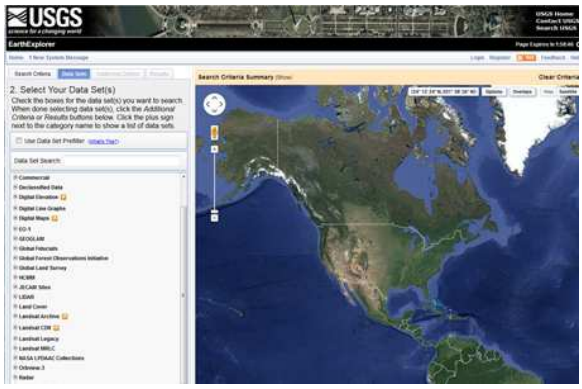

Figure 3: The US Geological Survey has an intuitive, web-based interface for exploring freely available georegistered datasets.

Figure 4: Street-level imagery has been used to determine building height, façade type, window-to-wall ratio, and building type. Google's StreetView currently has terms of use that prohibit saving or processing such imagery, but could be a valuable resources for computer vision experts and urban-scale building energy models.

CONCLUSION

Urban-scale building energy modeling is a disruptive technology that is becoming increasingly tractable and beginning to be adopted by several organizations including electrical utilities. We have provided redacted excerpts of tax assessor's data for a county in Tennessee as an example of the useful data and scalability challenges associated with such information extraction. In an effort to overcome those challenges, the team has provided an overview of more scalable data sources and algorithms in the context of the use cases and workflow utilized to create a digital twin of 178,368 OpenStudio/EnergyPlus buildings in a utility's service area.

Future work will involve empirical validation and sharing building-specific, utility-scale energy, demand, emissions, and cost savings realizable through the deployment of more intelligent energy efficient technologies within the built environment. The authors also hope that comparison matrices for data sources and comparison of algorithmically-constructing building energy models with measured data will become more prevalent so that more direct comparison comparison among urban-scale modeling techniques can evolve into best practices.

	Google Street View
Summary	Street view images. Downloadable using Google Street View API
Data type	Image
Company	Google
Website	Developer URL
Temporal resolution	N/A
Spatial resolution	N/A
Measure accuracy	Location errors exist
Cost	Free
Format	jpg
Building inputs	Height, window-to-wall ratio
Area properties	N/A
Material properties	Road pavement materials (e.g., concrete, asphalt), building exterior materials (e.g., glass, concrete), parking lots (e.g., gravel, soil)
Coverage of US	High
Orientation	Multi-side
Existing internal software	Building height estimation
Existing expertise	OpenCV
Restrictions	25,000 API calls per day. Restrictions URL

NOMENCLATURE

- AutoBEM – Automatic Building detection and Energy Model Creation
- AutoGen – Automatic EnergyPlus file modi-

fier/Generator, worlds fastest building energy model creator utilizing text replacement for variable in EnergyPlus files; awarded by U.S. Copyright Office under registration number TXu 2-159-000.

- AutoSim – Automatic Simulator, (CR17-00072, UTB80000011) - worlds fastest buildings simulator for scalably distributing EnergyPlus files on High Performance Computing devices, simulating on virtual disk, and returning results for storage and analysis; awarded by U.S. Copyright Office under registration number TXu 2-141-960.
- ECM – Energy Conservation Measure
- EPB – Electric Power Board of Chattanooga, Tennessee
- EUI – Energy Use Intensity
- EV – Electric Vehicle
- GIS – Geographic Information System
- kWh – kilowatt-hours
- kW – kilowatt
- MLS – Multiple Listing Service
- MPI – Message Passing Interface
- NREL – National Renewable Energy Laboratory
- PV – photovoltaic (e.g. solar cells)
- SDK – Software Development Kit

REFERENCES

Chen, Yixing, Tianzhen Hong, and Mary Ann Piette. 2017. "Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis." *Applied Energy* 205:323 – 335.

Copeland, William, and Joshua R. New. 2019. "Digital Twin of a City Utility: Issues, science, implementation, and results." *Proceedings of the Better Buildings Summit, Data of the Future: Digital Cities*.

New, Joshua R., Mark B. Adams, Piljae Im, Hsiuhan Yang, Joshua Hambrick, William Copeland, Lilian Bruce, and James A. Ingraham. 2018. "Automatic Building Energy Model Creation (AutoBEM) for Urban-Scale Energy Modeling and Assessment of Value Propositions for Electric Utilities." *Proceedings of the International Conference on Energy Engineering and Smart Grids (ESG)*.

Parshall, Lily, Kevin Gurney, Stephen A. Hammer, Daniel Mendoza, Yuyu Zhou, and Sarath Geethakumar. 2010. "Modeling energy consumption and CO2 emissions at the urban scale: Methodological challenges and insights from the United States." *Energy Policy* 38 (9): 4765 – 4782.

Reinhart, Christoph F., and Carlos Cerezo Davila. 2016. "Urban building energy modeling A review of a nascent field." *Building and Environment* 97:196–202.

Strzalka, Aneta, Jrgen Bogdahn, Volker Coors, and Ursula Eicker. 2011. "3D City modeling for urban scale heating energy demand forecasting." *HVAC&R Research* 17 (4): 526–539.

Yuan, Jiangye, Joshua R. New, Jibonananda Sanyal, and Olufemi Omitaomu. 2015. "Urban Search Data Sources." *ORNL internal report ORNL/TM-2015/397*.