GIS Data Extraction and Visualization to Support Urban Building Energy Modelling

Daniel Sobieraj[1], Xavier Mendieta[1], J.J. McArthur[1]
[1]Ryerson University, Toronto, Canada

Abstract

Urban building energy modelling is increasingly pursued to inform energy distribution and conservation strategies at the urban scale, but data availability is an impediment to such efforts. This paper presents a multi-campus investigation to quantify the availability and quality of GIS data for populating building archetypes for such models. Key findings are that GIS data is inconsistent and incomplete, but when combined with census data and satellite imagery, can provide adequate information to select and modify appropriate prototypical building archetypes for urban building energy models. The visualization inherent in GIS facilitates open-data deployment of model results at the city scale.

Introduction

With the urban population of the planet expected to grow from 3.9 billion (2014) to 6.3 billion by 2050 (United Nations, 2015), greenhouse gas emissions are also predicted to increase causing environmental damage. In response to this Canada has set several ambitious carbon reduction targets. The Canadian Federal Government has committed to reduce carbon emissions 17% by 2020 and 30% by 2030 (from 1990 baseline levels). These targets present a challenge since projected emissions with current measures are at least 17% above this defined baseline (Environment and Climate Change Canada, 2016) therefore there is a growing need to improve existing tools for strategic planning. Urban Building Energy Modeling (UBEM) has gained popularity due to its ability to simulate energy reduction interventions at larger scales. Previous studies such as that by Davila, Reinhart, & Bemis (2016) have established workflows to develop comprehensive urban building energy simulations, while others (Ballarini, Corgnati, Corrado, & Tala, 2011) have used a benchmarking approach to develop representative Energy Use Intensities (EUIs) and apply these to district scale models. Both approaches rely on establishing reference buildings (building archetypes) that are capable of representing the entire existing stock. Archetype buildings and UBEM models (UBEMs) require specific information about the building stock and researchers are starting to use data from Geographic Information Systems (GIS) to populate them. Within this paper, the adaptation of such workflows to develop UBEMs or archetypes are not proposed; rather the degree to which the publicly available GIS datasets are accurate sources to develop energy benchmarks and populate these archetypes and UBEM based is investigated.

This paper investigates the reliability of this data using a case study investigation of twelve university campuses across Ontario (Canada). GIS data for these campuses has been extracted, analyzed for completeness, and compared with information from other sources to evaluate its quality of GIS data. Further, the sensitivity of UBEM and archetypal model simulation to errors within this data is investigated through a study of buildings at Ryerson University, where detailed energy sub-metering data was available for empirical comparison.

Current Applications of GIS

There are many benefits to using Geographic Information Systems (GIS) for collecting and organizing data in the creation of building archetypes, energy benchmarks or UBEMs. In regions where there are large urban centers, GIS databases are extensive and are updated regularly (Davila & Reinhart, 2016). Such municipalities are most likely to have datasets containing building footprints, 3D massing, building occupancy type, and building construction period (Mastrucci, Baume, Stazi, Salvucci, & Leopold, 2014). These datasets are updated on a regular basis and are usually open source to a degree, providing the public with accessible and accurate data. In smaller cities, limited resources often preclude the development or update of such extensive GIS data sets. Despite this lack of data, building footprints are still generally available however additional data collection, such as report collection and manual visual analysis, is required (Davila et al., 2015).

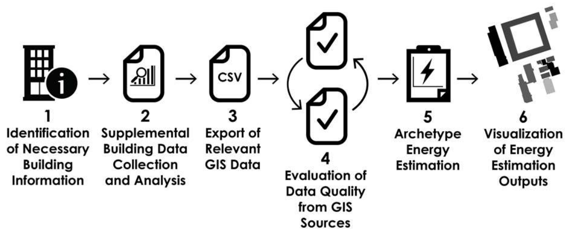
Besides being an effective platform for collecting data that is necessary for urban building energy modeling, GIS is a powerful visualization tool. GIS has been used in previous research to visualize results in 2D and 3D (Zhou & Li, 2006) and can be integrated with CAD and BIM data to create a seamless visualization (Döllner & Hagedorn, 2007), which can be directly exported and displayed in open source environments.

Given these benefits, GIS can be a significant enabler of carbon reduction at the municipal level and has proven to be beneficial in the analysis of multi-building energy consumption and interventions (Huang et al, 2015; Stoeglehner et al, 2016). However, this application is limited by data quality and availability. The case study presented investigates the extent to which data quality and availability is offered across various urban centers in Ontario (Canada).

Methodology

A case study set of buildings has been developed in a six-step process to evaluate the quality of GIS data and the impact of using inaccurate data for energy simulation, as illustrated in Figure 1: (1) Identification of necessary building information; (2) supplemental building data collection and analysis; (3) export of relevant GIS data to populate building models; (4) evaluation of data quality from GIS sources; (5) archetype energy estimation to evaluate the impact of using inaccurate data; and (6) visualization of energy estimation outputs.

Figure 1: Methodology Flowchart



Identification of Necessary Building Information

The building information required to generate UBEMs was investigated by reviewing a variety of contemporary approaches to their development.

The use of archetypal or prototypical buildings (such as those developed by the US Department of Energy) to represent the building stock across the region of interest. These archetypes for energy analysis are based on building parameters that influence energy consumption, such as operation type, period of construction, building geometry and occupancy schedules. This reduces the level of effort significantly, as detailed building information collection is no longer required.

Davila, Reinhart, & Bemis (2016) developed an alternative workflow and a tool to generate UBEMs using archetype buildings and existing geospatial datasets. This approach involves importing GIS data into Rhinoceros 3D, simplifying building layouts, and defining heights, window to wall ratios, construction assemblies, thermal zones, and schedules based on developed templates.

The minimum required building information was thus that deemed to identify the most appropriate template or prototypical building model: the building geometry (at minimum the area, but preferably the footprint, height, window to wall ratio), the construction vintage, and the building type.

Supplemental Data Collection and Analysis

Several additional sources of data were obtained to both complement and validate the accuracy of the GIS data. These included mandatory energy reports, satellite imagery (combined with Daft Logic for measurement and visual review), and regional surveys including census data. The relationship between these information sources and data types available in each is illustrated in Figure 2.

Mandatory Energy Reporting (MER) was available from the Ontario Ministry of Energy Broader Public Sector database, first published in 2013. Current regulations require all provincially-funded entities such as post-secondary educational institutions to annually report

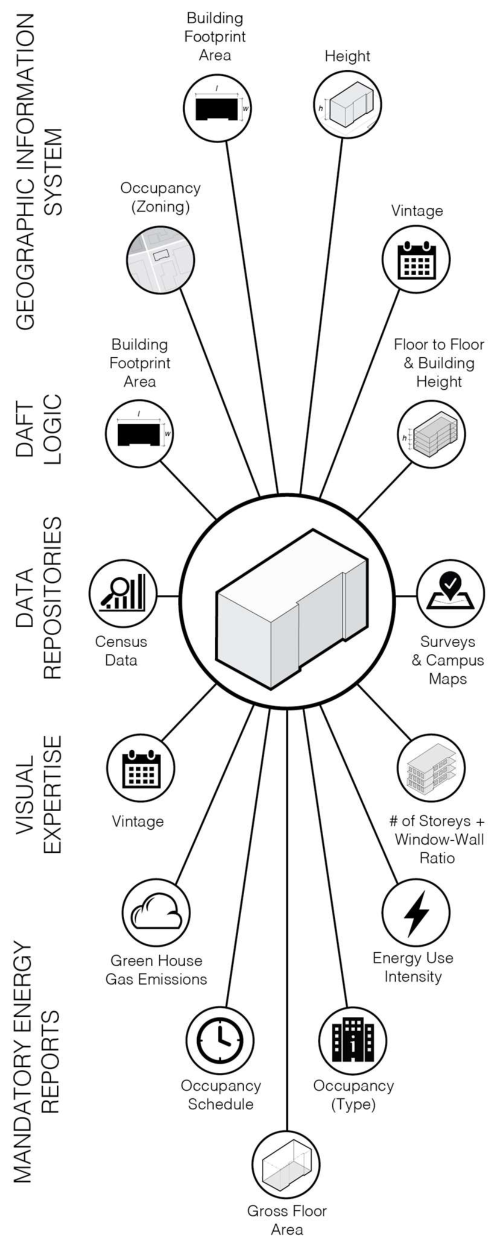


Figure 2: Data Sources for UBEM Input

their buildings energy consumption and Greenhouse Gas emissions. These reports include building areas, operation types, locations, and weekly operational hours. Previous research (Mendieta & McArthur, 2017) developed region-specific institutional energy benchmarks using this data set identified three building characteristics within this set that provided useful correlations with energy use intensity: building geometry, period of construction, and weekly occupancy hours.

Detailed building geometry information was available through satellite imagery and associated software. Visual inspection of aerial and street level maps allowed documentation of the number of building storeys, construction materials, window-wall ratio, and building architectural style, which in turn could inform or confirm building vintage. Within the satellite imagery analysis, *Daft Logic* - a satellite-based global positioning system for 3D measurements - was used to measure building floor plate areas, building heights, and individual floor-to-floor heights. Floor-to-floor heights were used to define the number of storeys and gross floor area (GFA) of the case study buildings. The GFA obtained using *Daft Logic* was then compared to the GFA provided by the mandatory energy reports. Buildings that had a 20% GFA difference were carefully inspected and if no logical reason was determined the building was discarded to avoid polluting the data.

Building vintage data was primarily collected from the Toronto Survey and Mapping Services. This dataset was created in 2003 and many case study buildings were not included and thus assumed to be 2004-present vintage unless another information source indicated otherwise. Additional vintage data was collected from the Canadian Census database and confirmed, when necessary, with the visual inspection described previously. The data collected from all sources (GIS, Daft Logic, regional surveys, and visual expertise) was gathered and compiled with Ontario's BPS energy reports in a *.csv* file for evaluation, as discussed in the Results section.

Export GIS Data to Populate Building Models

GIS datasets are available for most universities included in Ontario's BPS mandatory energy reports. GIS facilitates the extraction of geometric characteristics, period of construction, and operation type of multiple buildings in a given location. In order to obtain these locations, mandatory energy reports for public universities were consulted (and used as supplemental information, as discussed later) to obtain building names and addresses. Postal codes and addresses were cross-referenced with the GIS data to identify the corresponding building footprints and ensure the correct organization of the available information for each building considered. Where a single address was reported to represent an entire campus, additional research was required at this stage to identify the correct GIS data.

Building footprints, occupancy types based on zoning categories, building heights and 3D massing were collected from the open source GIS datasets provided by municipalities. Data for building construction period in Toronto was obtained from the Toronto Survey and Mapping Services in a raster format that had to be translated into vector format. Building footprint, zoning category, building height and 3D massing data collected was in *shapefile* and vector format that was imported into ESRI's ArcMap. A table containing the names, postal codes, and addresses of the buildings to be analyzed was uploaded as a *.csv* file to a Location Hub Portal where this data was geocoded – a process where information is linked to addresses represented spatially as points with defined (x,y) coordinates. The Location Hub Portal exported the geocoded data in *.xls* file format for import to ArcMap.

A spatial join was performed in ArcMap to link the data points containing the building names and addresses to the *shapefiles* containing available building geometry, occupancy type based on zoning, and period of construction information. The data compiled in ArcMap was then exported as a table in *.csv* file format. Supplemental data that is collected manually by other means can later be imported into ArcMap to keep data organized and exported as a single table. All the available data that was collected for each of the campuses is displayed in Table 1.

Evaluation of Data Quality from GIS Sources

GIS datasets were extracted and carefully reviewed to evaluate if the information contained in Ontario's GIS are a reliable and effective source to collect data. To evaluate the quality of the available GIS datasets, the data collected from GIS was compared to data collected from supplemental sources.

Building footprint area, height, vintage, and occupancy type obtained from GIS was compared to data from Mandatory Energy Reports, aerial imagery through Daft Logic, census data, and visual expertise. The results of this comparison are presented in Table 2.

The quality of information found in GIS is dependent upon the source data. The quality of the source data relies on the accuracy of the method and tools that were used to collect the data. The quality of information when translating data from raster to vector format to be input to GIS is reliant upon the resolution of the raster source, such as satellite imagery and maps. A hierarchy of information sources was developed for each information type in order compare each source in that type to the primary.

The primary information source was selected based upon several factors. Data collected from sources that were up to date and updated regularly were favored to those that were less recent and not maintained. Sources that had data available in vector format were preferred over sources that provided information in rasterized formats due to errors in precision and positional accuracy when formatting from raster to vector format. Information sources that provided detailed data even at a small scale were chosen over sources that provided generalized data suitable for a larger scale, not at the individual building scale.

Table 1: GIS Data Availability across Campuses

Post-Secondary Institution		Data Available											Note
Institution	Campus	Geographic Information System (GIS) Data					Mandatory Energy Report (MER) Data					Census Data	
Large Urban Centres (Pop. > 100,000)		Building Footprint Area	Height	3D Massing	Zoning (Type)	Vintage	Energy Use Intensity	Gross Floor Area	Greenhouse Gas Emissions	Occupancy Type	Occupancy Schedule	Vintage	
George Brown College	Casa Loma Toronto, ON	A	A	A	A	N	A	A	A	A	A	A	☐
George Brown College	St. James Toronto, ON	A	A	A	A	A	A	A	A	A	A	A	★☐
George Brown College	Waterfront Toronto, ON	A	A	A	A	A	A	A	A	A	A	A	☐
OCAD University	Toronto, ON	A	A	A	A	A	A	A	A	A	A	A	★☐
Ryerson University	Toronto, ON	A	A	A	A	A	A	A	A	A	A	A	★☐ ○
Sheridan College	Davis Brampton, ON	A	N	N	N	N	A	A	A	A	A	N	
Sheridan College	Hazel McCallion Mississauga, ON	N	N	N	A	N	A	A	A	A	A	N	
Sheridan College	Trafalgar, Oakville, ON	A	N	N	A	N	A	A	A	A	A	N	
University of Ottawa	Ottawa, ON	A	N	N	N	N	A	A	A	A	A	A	★◆
York University	Keele Toronto, ON	A	A	A	A	N	A	A	A	A	A	A	☐
Small Urban Centres (Pop. < 100,000)													
Loyalist College	Belleville, ON	A	N	N	A	N	A	A	A	A	A	N	■
St. Lawrence College	Cornwall, ON	A	N	N	A	N	A	A	A	A	A	A	■
Notes		☐	More accurate building vintage data available at a cost.										
		★	Building vintage data available for certain buildings.										
		■	Zoning / Land use and building footprints available on the municipality's interactive GIS map.										
		○	Mandatory Energy Reports provide Gross Floor area for entire campus and not by individual building.										
		◆	Building height and massing data available exclusively to Carleton University.										
		A	Data available.										
		N	Data not available.										

Information sources that provided detailed data even at a small scale were chosen over sources that provided more generalized data suitable for a larger scale, not at the individual building scale. While all the data in this case study was obtained through open data catalogues published by government entities, third party GIS datasets that are created by individuals or researchers are often available but should be analyzed for their precision and accuracy. It is therefore fundamentally important to revise and comprehend the quality of GIS data and evaluate its sources before any energy related analysis is performed.

Archetype Characterization and Energy Estimation

The collection of individual building characteristics such as wall components, HVAC efficiencies, air tightness, and occupancy schedules can become challenging and time consuming, therefore prescriptive requirements from ASHRAE standards and results of national surveys such as the Commercial Building Energy Conservation Surveys (CBECS) are typically used to characterize archetypes for energy analysis. These documents provide the relevant schedules, construction data, and representative zoning schemes for the majority of vintages and building types, however many post-secondary institutional buildings are not explicitly covered. The development of validated institutional prototypical building models requires participation from post-secondary buildings in Ontario and this would be a necessary area of future research. A set of 62 post-secondary classroom buildings that are part of Ontario's MER and that fall within ASHRAE's climate zone 6A were used to present the assessment of the quality of publically available GIS sources. The results and considerations from the analysis are discussed in the following sections.

In this preliminary study, prototypical buildings developed by the U.S. Department of Energy (DOE) were used to represent two Institutional Office/Research buildings part of Ryerson University. The buildings used for the case study were chosen due to availability of measured energy data. Appropriate prototypical-building models were selected based on the buildings area, vintage, and operation, all collected through three different approaches: (1) GIS and Census data; (2) Daft Logic, campus maps, and visual expertise; and (3) Ryerson space audits and construction data. The first approach used GIS data to define the GFA and operation of the building where Census data determined the vintage of both buildings, as this information was not available in the existing GIS dataset. The second approach defined the GFA of the buildings by using Daft logic to measure the buildings floor plate area and visual expertise to determine the number of floors and the buildings vintage. Campus maps set the operation type. The third approach is based on the results of space audits done by Ryerson University. Using net areas rather than GFA can influence the results of the case study, however this was the most accurate approach to represent the GFA of the buildings. Documents available to students of the university were used to determine the construction period of the buildings and the year when retrofits were done. The later was not used for the simulation of the energy models however it provides insight on variations of the results.

The selected prototypical-building models were simulated using Energy Plus. Construction assemblies were defined by the buildings vintage and are based on data collected from the results of the CBECS. A glassing ratio of 33% was established by visual expertise and was used for all models as none of the data sources provides this information. Thermal spaces, equipment loads, activity zoning, and HVAC systems were all predefined based on the archetype. Measured electricity and heating energy was available for both of the buildings in the case study, this information was used to compare the results from the simulation of each approach.

Visualization of Energy Estimation Outputs

After having completed the energy estimations for the buildings, the data was prepared for visualization. The data from the output of the estimations was synthesized and the energy use intensity and peak loads were extracted. Simulated EUI and peak load data was added to the table previously exported from ArcMap that contained data for the simulated buildings. The updated table was once again geocoded using a Location Hub Portal. The table exported from the Location Hub was then imported into ArcMap and spatially joined with the building footprints of the simulated buildings. Symbology was added to the data of the buildings being displayed. The result was a choropleth map in which lower values were displayed as light colors and higher values were displayed as darker colors. This process allowed the data to be exported in both raster formats (PNG and JPEG), or vector formats (Illustrator and PDF), which can be further edited in other programs. Data obtained in this study was exported to Keyhole Markup Language (.kml) format where it was then uploaded to Google Maps for visualization. This platform was selected based on ease of integration with other data formats and data entry, and of its open-source nature.

Results and Discussion

Table 1 displays all the available data that was collected for each of the campuses. In the following section the analysis of this compiled file is presented with the objective of providing insight on the suitability and accuracy of using GIS as a primary source for data collection in Ontario.

Limitations of Data Availability

Currently, the greatest limitation of GIS is the lack of data availability. As evident in Table 1, large urban centers (e.g. Toronto) contain GIS datasets with the widest information availability, particularly compared with medium-sized (Ottawa, Brampton, Mississauga) and small (Belleville, Cornwall) urban centers. Building footprint and zoning data was available from almost every municipality's open source data catalogues. However, building height data and vintage was less commonly available and only found in large urban centers. Gross floor area was not a GIS dataset that was available from municipalities' open source data catalogues and thus had to be estimated.

The collection of datasets not included in municipal open data catalogues can be very difficult and time consuming. In this case study, building construction period data was collected partially from a dataset from the City of Toronto's Survey and Mapping Services and partially from census data available from Statistics Canada. The census data was obtained in a table format that needed to be further cleaned and refined before importing to GIS. Although this information was available to the public and contained most of the information for the buildings in the case study, it was not the most detailed data and did not include all the buildings inspected. A more current dataset for construction periods for the City of Toronto was available from the Municipal Property Assessment Corporation at high cost, and was found in an energy mapping study to misclassify 30% of properties.

Proprietary sources include not only corporations but also post-secondary institutions. For Ottawa, the only available building footprint data from open source catalogues was for the buildings on the University of Ottawa campus, while nearby Carleton University possessed an additional dataset containing 3D building massing and height but this data was only available to students and faculty of Carleton University and was not available for purchase.

Trends in Data Quality

The quality of building footprint area varied greatly between campuses located in large and small urban centers. GIS building footprint area information was on average more accurate (based on satellite imagery measurements) for campus buildings in small urban centers (99.1-99.5% accuracy) than those in large urban centers (18.2-99.1% accuracy). The high accuracy of the former data can be related to the increased granularity of available information facilitated by the smaller number of buildings. On the larger campuses, GIS data was often at the tax lot level and resulted in more homogeneous data reporting (one building as a proxy for multiple adjacent buildings), resulting in oversimplification and – in several cases – assignment of a total building area to each constituent building on the tax lot.

Within large urban centers, GIS building footprint area accuracy also varied. In this case study, with the Greater Toronto Area, the quality of data decreased the farther the campuses were located from the dense core of downtown Toronto. The accuracy of GIS building footprint area data varied by as much as 80.9% between campuses located in downtown Toronto (George Brown – Waterfront Campus, 99.1%) and the farthest campus in the Greater Toronto Area (Sheridan – Trafalgar Campus, 18.2%).

The two primary methods whereby GIS footprint data is created can help explain the variance in data accuracy of the GIS building footprint data. The first method determines building footprint by software that traces building regions from photogrammetric imagery and digitizes them into polygons (Lee, Lee & Lee, 2008).

This method is entirely dependent upon the resolution of the images and can be inaccurate when compared to other, more current methods. The second method uses Light Detection and Ranging (LiDar) technology with photogrammetric imagery. This method can provide data with an error of 33.18% when detecting building

Table 2: Accuracy of GIS Data Compared with Other Sources

Post-Secondary Institution		Data Accuracy							
Institution	Campus (# of study bldgs.)	Building Footprint Area	Gross Floor Area		Building Height	Occupancy Type	Building Vintage		
Large Urban Centres (Pop. > 100,000)		GIS vs Aerial Imagery (%)	GIS vs MER (%)	MER vs Aerial Imagery (%)	MER vs Aerial Imagery (%)	GIS vs MER (%)	GIS vs Census Data (%)	GIS vs Visual Expertise (%)	Census Data vs Visual Expertise (%)
George Brown College	Casa Loma (4)	47.2	35.0	7.7	82.8	100.0	N/A	N/A	50.0
George Brown College	St. James (6)	76.6	21.2	17.6	88.2	0.0	16.7	16.7	33.3
George Brown College	Toronto (1)	99.1	9.6	8.9	97.2	0.0	N/A	N/A	0.0
OCAD University	Toronto (2)	66.1	49.9	35.5	76.9	50.0	N/A	0.0	50.0
Ryerson University	Toronto (22)	74.0	N/A	N/A	81.7	0.0	5.0	9.0	36.0
Sheridan College	Davis (6)	26.3	N/A	74.8	N/A	0.0	N/A	N/A	N/A
Sheridan College	Hazel McCallion (1)	N/A	N/A	91.6	N/A	0.0	N/A	N/A	N/A
Sheridan College	Trafalgar (10)	18.2	N/A	77.9	N/A	90.0	N/A	N/A	N/A
University of Ottawa	Ottawa (7)	79.0	N/A	77.9	N/A	0.0	N/A	N/A	28.6
York University	Keele (11)	77.4	53.3	82.3	87.6	100.0	N/A	N/A	54.5
Small Urban Centres (Pop. < 100,000)									
Loyalist College	Belleville (1)	99.1	N/A	75.4	N/A	0.0	N/A	N/A	N/A
St. Lawrence College	Cornwall (2)	99.5	N/A	10.4	N/A	100.0	N/A	N/A	0

footprint area (Prerna & Singh, 2015). While acceptable for building energy simulation at the urban scale as this error can cancel over multiple buildings, at the building scale, validation of such data is required.

Comparison of GIS data to supplemental data proved that while building geometry data was of value, the GIS data obtained for building construction age was less accurate than census data. This could be a consequence of using outdated data from a source in a raster format that had to be translated to vector format in GIS. Errors in data that is outdated are to be expected along with errors in positional accuracy, scale, and projections when formatting from a raster to vector format. Another factor that will affect the accuracy of the building age is the method in which the data is created. The data could have been collected from visual expertise or tax assessment data that was later generalized for areas the size of city blocks. Alternatively, building construction period data may have been created using LiDAR. In previous research LiDAR derived building attributes predicted building construction age with an average error of 16.8 years (Tooke & C. Coops, 2014).

Building height data was only available for campus buildings in Toronto, however this information type was on average the most accurate in comparison to all the other information types. The accuracy of the building heights from GIS data in comparison to aerial imagery varied by only 20.3% with the campus with the most accurate height being 97.2% and the campus with the lowest accuracy in height being 76.9%. The increased accuracy of building height data in comparison to building area footprint data is relevant to the fact that LiDAR is a system that is meant to measure distance and range rather than building regions and therefore is more accurate at measuring building heights (Lee et al., 2008). It is possible to measure building height from one or more aerial or photogrammetric images, however this information is relatively inaccurate.

Despite post-secondary institutions recording the unique addresses of all building on campus, GIS datasets may not recognize that these buildings are individual. For example, George Brown College reported the addresses of four individual buildings on their Casa Loma campus. However, the City of Toronto's building massing dataset only recognized two individual buildings because it had grouped the three addresses that were adjacent to one another as a single building footprint. This is an issue because buildings that are grouped together in GIS will not only display the wrong areas and heights, but they can often be of different building occupancy types and have different construction dates. This may be due to the inability of the software to differentiate individual building footprints in a dense building cluster. The accuracy of building massing data varies greatly from city to city and cities should ensure that their method of collecting building shape data is as accurate as possible.

Occupancy use types were categorized based on zoning categories because the exact occupancy use type of specific buildings is not available. The issue with creating building archetypes based on occupancy use types that are categorized in this way is that zoning categories generalize the building types in an area. For example, an archetype based on the Institutional Education zoning does not recognize that there are many different occupancy use types that fall under this zoning, such as classrooms, residences, and administration offices. More importantly, some post-secondary educational buildings, especially those in large urban centers were located in areas zoned for commercial residential buildings. The occupancy hours of a commercial residential building are very different from those of a classroom. The differences in the factors that inform the archetype model will make a significant difference in the results of energy simulations and will be discussed in the following section.

Archetype Energy Simulation Results

The variables that determined which prototypical energy model was appropriate to use were chosen due to their influence on energy consumption. The influence that vintage, operation, and area can have on energy use has already been discussed throughout this paper. This section focuses on the comparison of measured energy with simulated energy by using the three above-mentioned approaches. The results presented in Table 3, show that the most accurate method was the one that used space audits and existing reports provided by Ryerson University. The outcomes of this approach demonstrate that simulated energy was 14% and 29% below the measured energy data. This underestimation of the energy consumption can be related to occupancy schedules, as these buildings are not only used for administrative purposes and some rooms are accessible every day of the week; measured area, as the area used is defined by internal space audits and not GFA; and equipment loads, as some rooms contain lab equipment that is not typically seen in administrative offices.

The illustration below shows the significant limitation of reliance on unvalidated GIS data. Inadequate GIS zoning information classified the operation type for both buildings as commercial/residential. This variable alone would show significant variation on the buildings' energy consumption as it becomes impossible to identify the most appropriate prototypical building for simulation. To continue this investigation, however, both buildings were assumed administrative offices to evaluate the influence of the other variables. Census data was used to determine the vintage in this approach, the information collected was incorrect as both these buildings were known to be built before 1980. Finally, the geometry of both buildings was shown to be the same. The inaccuracy of this approach overestimated the energy use by 30% and 228% (Figure 3).

The results from the energy models based on Daft Logic, visual expertise, and campus maps were promising though some complications were identified. The most significant issue with this approach is that conditioned spaces are difficult to differentiate from unconditioned spaces as the GFA is obtained through multiplying the

Table 3: Energy simulation results using different approaches for data collection

Variables	Model Based on GIS / Census Data		Model Based on Daft Logic / Visual Expertise / Campus Map		Model Based on RU Campus Data	
Building Name	GER	COP	GER	COP	GER	COP
Vintage	Post-1980	Post-1980	Pre-1980	Pre-1980	Pre-1980 - R	Pre-1980 - R
GFA (sqft)	37,985	37,985	32,980	4,391	27,441	6,870
WWR (%)	33	33	33	33	33	33
Operation Type	Commercial/ Residential	Commercial/ Residential	Institutional Office/Research	Institutional Office/Research	Institutional Office/Research	Institutional Office/Research
Simulated Operation Type	Administrative Office M	Administrative Office M	Administrative Office M	Administrative Office S	Administrative Office M	Administrative Office S
Simulated Prototype GFA (sqft)	53,638	53,638	53,638	5,502	53,638	5,502
Simulated Prototype EUI (kWh/sqft)	20.77	20.77	19.14	24.73	19.14	24.73
Simulated Energy (EUI*GFA)	789,108	789,108	631,179	108,571	525,172	169,866
RU Measured Energy	607,897	240,678	607,897	240,678	607,897	240,678
RU Measured vs. Simulated (kWh)	181,211	548,430	23,282	-132,107	-82,725	-70,812
RU Measured vs. Simulated (%)	30	228	4	-55	-14	-29

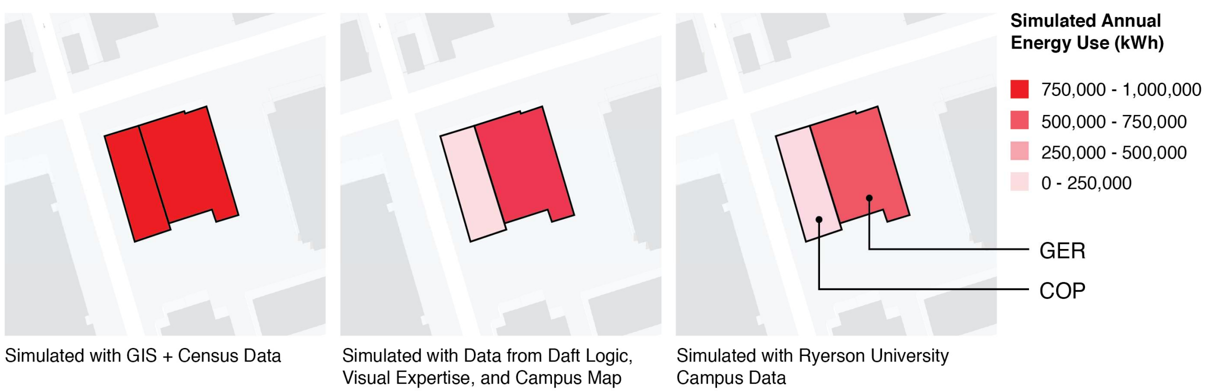


Figure 3: Comparison of Energy Simulation Data

roof area by the number of observed floors.

The operation type can be determined by looking at campus maps, however, determining the function of a building was not very intuitive and this approach would be challenging for non-educational buildings. The vintage was easily determined based on visual inspection due to the broadly defined construction periods (pre-1980 vs. 1981-2004). A more granular vintage determination would be much more challenging to address and would require more consistent reporting of building construction data. The underestimation of 55% by using this approach is more than likely related to the GFA measurement as it was determined to be 36% smaller when compared to the measured area obtained with the space audits.

Visualization

The visualization of energy consumption across the campus was easily executed using the methodology developed and the Google map created using the available sub-metered energy data is presented in Figure 4.

Conclusion

This paper has identified the following potential for GIS as a facilitating tool for UBEM development:

1. Within urban contexts, GIS provides a wealth of information to facilitate the collection of key building data, particularly related to the building geometry, however validation is critical outside the dense core areas.
2. Lack of specific building information and consolidation of addresses is a key limitation of current GIS data sources.
3. When validated with other sources of information, GIS and other sources are effective for collecting, cleaning, and synthesizing data to populate

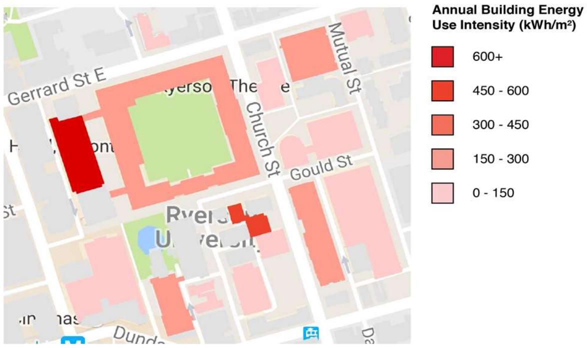


Figure 4: Visualization of RU Campus Buildings' EUIs

archetype models. This is further enhanced when data analytics techniques are used to synthesize this information and remove outliers.

4. GIS interfaces easily with spreadsheets, facilitating scaled multi-building calculations based on representative archetypal building performance and visualization of results can be quickly developed and deployed on open-format platforms for ease of dissemination.

Current limitations and areas for future research:

1. Pre-existing archetypes limited, any subdivision based on GFA, and do not account for building shape or period.
2. Building shape currently a manual determination, pattern recognition algorithms will be of value to automate this process.
3. Full sub-metering for selected buildings was not available so energy evaluation was based on electricity only.
4. Lack of DOE prototypical buildings for post-secondary limited application.
5. More data needs to be available: building height, building age, occupancy type and GFA.
6. The data that already exists (Building footprints, building age, building height) needs to be more accurate and maintained regularly.
7. Data should be made available and open source by institutions to help accelerate research in the field. Institutions should not only take from but they should also contribute to open data repositories.

Acknowledgements

The authors acknowledge financial support from Ontario Centres of Excellence Voucher for Innovation and Productivity), Arup, and Ryerson University Faculty of Engineering and Architectural Science.

References

Ballarini, I., Corgnati, S., Corrado, V., & Tala, N. (2011). Improving Energy Modeling of Large Building Stock through the Development of Archetype Buildings. *12th Conf. of International Building Performance Simulation* 2874-2881.

Council of Ontario Universities. (2015). *Inventory of Physical Facilities of Ontario Universities.*

Döllner, J., & Hagedorn, B. (2007, January). Integration of GIS, CAD, and BIM Data with Service-Based Virtual 3D City Models.

Davila, C., & Reinhart, C. F. (2016). Urban Building Energy Modeling: A Review of a Nascent Field. *Building and Environment*, 196 - 202.

Davila, C., Reinhart, C., & Bemis, J. (2016). Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets. *Energy*, *117*, 237 - 250.

Environment and Climate Change Canada. (2016). *Canada's Second Biennial Report on Climate Change.* Quebec: Government of Canada.

Huang, Z., Yu, H., Peng, Z., & Zhao, M. (2015). Methods and tools for community energy planning: A review. *Renewable and sustainable energy reviews*, *42*, 1335-1348.

Lee, D. H., Lee, K. M., & Lee, S. U. (2008). Fusion of Lidar and Imagery for Reliable Building Extraction. *Photogrammetric Engineering & Remote Sensing* , *74* (2), 215-225.

Mastrucci, A., Baume, O., Stazi, F., Salvucci, S., & Leopold, U. (2014). A GIS-Based Approach to Estimate Energy Savings and Indoor Thermal Comfort for Urban Housing Stock Retrofitting. *BauSIM* 190-197. Germany-Austria: IBPSA.

Mendieta, C. and McArthur, J. (2017). Development of Energy Benchmarks Using Mandatory Energy and Emissions Reporting Data: Ontario Post-Secondary Residences. *International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering*, *11*(2), 57-60.

Prerna, R., & Singh, C. K. (2015). Ealuation of LiDAR and image segmentation based classification techniques for automatic building footprint extraction for a segment of Atlantic County, New Jersey. *Geocarto International* , 694-713.

Stoeglehner, G., Narodoslawsky, M., Erker, S., & Neugebauer, G. (2016). Processes and Tools for Integrated Spatial and Energy Planning. *Integrated Spatial and Energy Planning*. 91-110. Springer International Publishing.

Tooke, T. R., & C. Coops, N. (2014). Predicting building ages from LiDAR data with random forests for building energy modeling. *Energy and Buildings*, 603 - 610.

United Nations. (2015). *World Urbanization Prospects, 2014.* New York: United Nations.

Zhou, G., Z., T., M., C., & Li, C. (2006). Customizing Visualization in Three-Dimensional Urban GIS via Web-Based Interaction. *Urban Planning and Development*, *132*, 97-10