Energy & Buildings 203 (2019) 109434

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild

Effect of window-to-wall ratio on measured energy consumption in US office buildings

Luke Troup[a], Robert Phillips[a], Matthew J. Eckelman[a], David Fannon[a,b,*]

[a] *Department of Civil and Environmental Engineering, Northeastern University, 400 Snell Engineering Center, 360 Huntington Ave., Boston, MA, 02115, United States*

[b] *School of Architecture, Northeastern University, 151 Ryder Hall, 360 Huntington Ave., Boston, MA, 02115, United States*

A R T I C L E I N F O

Article history:
Received 14 March 2019
Revised 25 July 2019
Accepted 14 September 2019
Available online 16 September 2019

Keywords:
Window-to-wall ratio (WWR)
CBECS (commercial building energy consumption survey)
Regression analysis
building energy use
office buildings

A B S T R A C T

Windows can significantly affect building performance, and the window-to-wall ratio (WWR) describes the portion of an exterior wall that consists of windows. Previous research on the effects of WWR and efforts to determine optimal WWR generally rely on iterative simulation or algorithmic optimization. This study seeks to understand the effects of WWR in actual office buildings using survey data reported in the 2012 CBECS (Commercial Building Energy Consumption Survey). Both total annual energy use and four discrete end-uses (heating, cooling, lighting and ventilation) were characterized, and 32 categorical and numerical building characteristic variables were selected for linear regression analysis. Descriptive statistics for energy use intensity (EUI) in ~1000 office buildings across 6 WWR intervals show increased median total EUI with increasing fenestration, driven by increasing cooling loads. Total EUI generally decreases with year of construction, regardless of WWR. Single-variable regression finds consistent statistical significance for WWR on cooling, lighting, and ventilation energy use, but with a maximum goodness of fit R^2=0.04. In contrast, the single variable with the largest explanatory power is cooling degree days (R^2=0.22 for cooling energy use). Multi-regression modeling finds a maximum R^2 value of 0.34, with WWR appearing as a significant variable in the regression equations for cooling and lighting. In sum, the 2012 CBECS microdata for office buildings suggests that WWR is a significant predictor of energy use for cooling, and to a lesser extent lighting and ventilation, but to a much lower degree than has been found by purely simulation studies.

1. Introduction

According to the US Energy Information Administration (EIA), the commercial building sector consumed approximately 19% of the 2016 total US energy in 2016, which accounted for almost 18% of national carbon dioxide (CO_2) emissions [1]. The Lawrence Berkley National Laboratory estimated 34% of commercial building energy use in the US is window-related [2]. Although large interior lighting and cooling loads (*i.e.*, lighting, equipment, and people) define internal-load-dominated buildings like offices, office buildings' enclosure remains important; the proper design of fenestration and envelope in can reduce lighting and mechanical system energy use by 10–40% depending on climate [3]. In addition to cost and environmental benefits, the advent of new energy codes, green building standards, incentive programs, and certification systems motivates higher-performing building design. Improved fenestration performance addresses elevated energy conservation demands and provides additional benefits such as avoided environmental impacts, reduced economic harm and enhanced occupant comfort and productivity [4].

Because empirical data are not available prior to construction, design decisions, rating systems and even some energy codes rely on building simulation [5] to understand the performance of specific buildings. These models depend on detailed project information and accurate representations of complex building systems [6], and challenges like unpredictable occupant behavior, uncertainty regarding specific building characteristics and quality of construction, and model underspecification, complicate accurate estimates of energy use [7,8]. An alternate approach to building performance develops statistical models using large data sets of measured energy consumption and building characteristics from existing buildings, which offers opportunities to benchmark the relationships among building parameters as experienced in actual buildings, in all their messy reality.

* Corresponding author at: Department of Civil and Environmental Engineering, Northeastern University, 400 Snell Engineering Center, 360 Huntington Ave., Boston, MA, 02115, United States

E-mail addresses: d.fannon@northeastern.edu, d.fannon@neu.edu (D. Fannon).

https://doi.org/10.1016/j.enbuild.2019.109434
0378-7788/

The Commercial Building Energy Consumption Survey (CBECS), conducted by the EIA, is the only national-scale data source that includes building characteristics and measured energy consumption of the non-residential building stock [9]. Office buildings comprise the largest fraction of primary building activity in the recently-released 2012 CBECS—by both number of buildings and by total floor area—representing more than 16 billion square feet of commercial floor space. Since the dataset is representative of the national building stock, analysis across all US climate regions can isolate effects of specific parameters, including location, and provide analysis relevant for the entire country.

The present study tests the influence of window-to-wall ratio (WWR) on energy consumption in commercial office buildings using the 2012 CBECS data. Descriptive summary statistics examine relationships between WWR and total annual energy use intensity (EUI) and with end-uses directly affected by envelope performance, including heating, cooling, lighting, and ventilation. Because CBECS surveyed a representative sample of buildings, simple linear regression can evaluate statistical significance of WWR and other dataset building characteristics compared to end-use EUIs. Finally, effects of WWR in conjunction with other selected variable combinations are considered with multi-linear regression techniques and the results are validated against prototype building energy simulations.

2. Background

2.1. Window-to-Wall Ratio (WWR)

Windows shape aesthetics, admit daylight, afford views, and provide thermal enclosure in buildings. As a result, windows are regarded as one of the most important components of a commercial building envelope system [10–12]. The *window-to-wall ratio* (WWR) represents the proportion of exterior wall surface area that is *glazed* (consists of windows), and WWR affects many building attributes: window size establishes the physical/visual connection to the exterior, and determines the environmental impacts associated with material use [13]. WWR affects building energy use and occupant comfort through heat transfer (conduction), solar heat gain (radiation), air leakage or ventilation (infiltration) and daylighting (lighting demand offset) [4].

To quantify the effect of WWR on energy, the *ASHRAE Handbook of Fundamentals* provides a simplified, three-term equation to calculate the energy flow through window assemblies based on a combination of physical glazing properties (*e.g.*, solar heat gain coefficient (SHGC), visual transmittance (T_v) and thermal conductance (*U*-value)) as well as climatic conditions. Each term in the equation is a function of fenestration area, which underscores the effect of WWR on building energy use. That relationship is straightforward for thermal conductance, but of course the orientation and distribution of that fenestration area are critical to determining solar heat gain, and will affect lighting, which is not considered in this simple calculation. All three terms also depend on climate, which directly influences envelope, and especially window, performance [21–25]. Selecting the appropriate WWR—like many climate-responsive passive-design strategies—requires carefully balancing occasionally competing goals like aesthetics; thermal and visual comfort while reducing dependence on power-operated systems [26,27].

ASHRAE Standard 90.1, a model energy code, defines building characteristics to ensure minimum energy performance for each US climate zone, and establishes a maximum WWR of 40% for commercial buildings following the prescriptive path to compliance, and as the baseline for the performance path [28]. Numerous simulation-based investigations have attempted to determine the *optimal* WWR [14–20] under various circumstances, and some suggest WWRs less than 40% minimize energy use [29–32]. Proposals to reduce the maximum WWR to 30% during the 2013 revision of *ASHRAE 90.1* met some controversy and were ultimately rejected [33,34], but WWR remains an important topic in building design and operation.

2.2. The Commercial Building Energy Consumption Survey (CBECS)

Under the 1977 US Department of Energy (DOE) Organization Act, the EIA was mandated to “collect, evaluate, assemble, analyze, and disseminate data and information” relating to energy use and production [35]. The EIA developed comprehensive surveys covering three energy-use sectors (commercial, residential and manufacturing) which are required to be conducted at every four years [36]. CBECS collects Building Survey data directly from owners and operators—although some question the accuracy of such self-reporting [37,38]—and supplements as needed with energy data from the Energy Supplier Survey [39].

The fundamental purpose of CBECS is to provide comprehensive, statistically-representative energy use information about the US commercial building sector [36]. Uniquely, CBECS data pairs building characteristics with energy consumption data, and covers a broad range of commercial building types [38,40,41]. Survey respondents provide most characteristics by choosing generalized descriptions and categories without resorting to onerous measurement of physical parameters such as envelope *R*-value. Though this approach simplifies data collection, it limits the usefulness of survey responses for energy use calculations and simulations [42]. In any case, CBECS supports a broad range of activities and research, including many of the statistical studies cited in the next section. In addition, the dataset provides a historical record of building performance [40] and serves as the backbone for benchmarking EnergyStar performance in Portfolio Manager, which scores building energy use compared with averages [43].

The 2012 CBECS data contains 1119 total variables describing each individual building record. Individual variables are indicated here in caps (VARIABLE) using their CBECS names. 506 of the variables describe building characteristics or operation. 49 of these 506 description variables specify annual energy consumption by sources (*e.g.*, major fuel, electricity, natural gas, fuel oil or district heat) and end-uses (*e.g.*, heating, cooling, ventilation, water heating, lighting, cooking, refrigeration, computing, office equipment, and other uses) per source. The annual energy use variables are recorded by the amount of energy consumed (in kBTU), expenditures, and the quantity of the energy carrier used (*e.g.*, kWh, cubic feet, and gallons). CBECS analysts ensure the end-use breakdown accounts for the reported total building energy consumption by processing end-use variables with engineering models, cross-sectional regression and reconciliation [71].

Each building observation is weighted according to the proportion of the population it represents; the *final sample weight* (FINALWT) variable is the number of buildings in the entire US building stock represented by that single building survey record. CBECS is not a *census* but a sample survey of statistical estimates, which includes associated sampling errors measuring the accuracy of a particular estimate. To determine the accuracy of applying the survey estimate to the entire population, CBECS uses the jackknife method with the provided 197 *replicate weights* (FINALWT1–FINALWT197) to generate relative standard errors (RSE) between estimates and the full-sample estimate. The EIA recommends using the weights and/or calculating RSE when using the CBECS dataset [72], and we have followed that recommendation here.

2.3. Statistical prediction of building performance

Numerous studies have investigated different approaches to determine the influence of particular building characteristics on building energy use [44,45]. Statistical methods to forecast building energy consumption include regression [46–48], artificial neural networks (ANN) [49,50], random forest [51], clustering [52] and principal component analysis (PCA) [53–55]. Considering regression, both linear and non-linear multiple regression techniques have been used to predict energy for different building end-uses [8], and the early application of such models during the design phase was found to improve energy efficiency and reduce emissions [56]. The number of variables and the complexity of the regression equation can dramatically affect the model goodness of fit [57]. The manual selection process of stepwise regression can provide insight to individual variables and the influence of different levels [58].

Curiously, while WWR has been investigated extensively both globally and in the US for its influence on occupant well-being [59–67], controlled, statistical comparisons of alternative WWR designs in buildings generally rely on data generated simulations. For example, Yu et al. and Korolija et al. performed regression modeling on window assembly properties (*e.g.*, solar heat gain coefficient (SHGC), visual transmittance (T_v) and thermal conductance (U-value)) in combination with the WWR, climate and orientation in order to predict building energy use [68,69]. These analysis demonstrated the influence of climate region, because the statistical significance of window size and orientation is different in each location. In general, past analysis results indicate energy savings with increased SHGC and T_v and a heating energy savings with increased U-value [70].

The present study takes a parallel approach to determine the influence of WWR on building energy use, constructing regression models from data measured in actual buildings and recorded by the 2012 CBECS dataset, rather than from simulations. The goal of regression analysis is to quantify the influence of independent *predictor* variables (like the characteristics of office buildings) on the resulting dependent *response* variables (like energy consumption). Our regression models seek a coefficient for each building characteristic predictor variable that achieves the best linear fit to the surveyed annual energy use response variables: in short, the best possible prediction of energy use based on building attributes.

3. Methods

3.1. Observation & variable selection

To study of the influence of WWR on office building energy use, the 6720 commercial building observations in the 2012 CBECS dataset were subset to 1212 observations, where the ‘Principal Building Activity’ (PBA) is classified as ‘Office’, the ‘Percent Exterior Glass’ (GLSSPC) contained a reported value, and the ‘Window Glass Type’ (WINTYP) was *not* reported as having no windows. The GLSSPC variable, first introduced to the survey in 2003, is synonymous to WWR, with responses collected in six percentage intervals: less than 1%, 2–10%, 11–25%, 26–50%, 51–75% and 76–100%. WWR focuses only on the *amount* of glass; CBECS also contains data regarding the distribution or orientation of windows, specifically the categorical variables ‘Equal glass on all sides’ (EQGLSS) and ‘Glass on sides with most sunlight’ (SUNGLS).

3.1.1. Independent variables

We selected 28 independent building characteristic variables (including GLSSPC) that are relevant to energy performance in office buildings. Since this study focuses on windows, we excluded variables regarding details of mechanical systems, controls, and management systems. The 28 selected variables are divided into six sets to decompose the influence of groups of variables on building energy use. The *Initial Set* consists of ten variables capturing whole building and façade characteristics that influence energy use regardless of location. The *Envelope, Building*, and *Operational Sets* extend the *Initial Set,* adding potential energy-contributing building parameters. A *Windows Only Set* and an *Envelope Only Set* were isolated from the 28 variables to explore the façade effects independently. The selected variables and sets are shown in Table 1. Because building energy performance is driven by location-specific weather data [73], four climate variables recorded in CBECS (listed in Table 1) were also included as independent, or predictor variables, although of course these overlap inasmuch as they describe the same climate. Thus, we used 32 total independent, or predictor, variables in this analysis.

3.1.2. Dependent (Response) variables

This study considered annual building energy use, which CBECS records by energy carrier. These variables are summed to obtain a new, calculated measure of Total Annual Energy consumption (TOTBTU) which is not in CBECS Public Use Microdata File. To standardize comparisons, TOTBTU is divided by the building area (SQFT) of each building to yield the Total Annual Energy Use Intensity (TOTEUI) measured in kBtu/ft^2 · year. However, TOTEUI may mask conflicting trends among different energy end-uses arising from changes to WWR [74], and so disaggregated heating, cooling, ventilation and lighting energy use intensities (HTEUI, CLEUI, VNEUI and LTEUI, respectively) were also calculated and individually analyzed, yielding five total dependent, or response, variables listed in Table 1.

3.2. Data formating and filtering

Of the 32 independent variables 21 contain categorical data, while the remaining 11 contain integers (numerical) data. For example, GLSSPC is a categorical parameter because the six levels are descriptions of a building characteristic indicating a range of values, rather than a numerical value of the actual WWR. The CBECS variables are documented in the *Variable and Response Codebook* [75], and the details of the variables as used in this study—including notes on data formating—can be found in Table S1 of the Supplemental Information (SI).

Missing and null values in categorical variables present a challenge; to capture nonresponse or not applicable (NA) levels, we replaced all missing values with zero in these cases. We also replaced missing values in numerical variables with zero in cases where ‘not applicable’ can be interpreted the same as none (*e.g.*, ‘Number of Underground Floors’ (BASEMNT) and ‘Percent Daylight’ (DAYLTP)).

Two of the 32 independent parameters mask values outside the typical response ranges to preserve the confidentiality in the survey. ‘Number of Floors’ (NFLOOR) is explicitly reported for offices less than 14 floors, but buildings with more than 14 floors are reported in two tranches: with values of 994 for 15–25 floors, and 995 for more than 25 floors. Similarly, buildings built before 1946 report ‘Year of Construction’ (YRCON) as a single default value of 995. 214 office building observations meet both conditions and were removed from the dataset based on the lack of specific information about building height and age. Unfortunately, these 214 records constitute approximately 17% of the office buildings in raw dataset and by definition represent the older and taller buildings.

3.3. Statistical methods

We used a set of three increasingly sophisticated method of statistical analysis on the dataset, beginning with basic descriptive statistics, followed by simple linear regression, and finally multi-linear additive regression to test for relationships between WWR and energy consumption. Applying these techniques to the unique features of the CBECS dataset demanded particular methods, which are described below.

Table 1
Selected independent (Predictor) variable sets and dependent (Response) variables. The numerical hash symbol (#) denotes numerical variables, and all other variables are categorical.

Building Set			
Abbreviation	*Description*	*Abbreviation*	*Description*
SQFT	Square footage (#)	RFTILT	Roof tilt
BLDSHP	Building shape	NFLOOR	Number of floors (#)
BASEMNT	Number of underground floors (#)	FLCEILHT	Floor to ceiling height (#)
ATTIC	Attic	YRCON	Year of construction (#)
RENOV	Any renovations	CUBEC	Percent open plan
DAYLTP	Percent daylight (#)		
Envelope Set			
Abbreviation	*Description*	*Abbreviation*	*Description*
WLCNS	Wall construction material	RFCNS	Roof construction material
RFCOOL	Cool roof materials	AWN	External overhangs or awnings
SKYLT	Skylights or atriums		
Windows Set			
Abbreviation	*Description*	*Abbreviation*	*Description*
GLSSPC	Percent exterior glass	SUNGLS	Glass sides most sunlight
WINTYP	Window glass type	TINT	Tinted window glass
REFL	Reflective window glass		
Operational Set			
Abbreviation	*Description*	*Abbreviation*	*Description*
FEDFAC	Federal complex	WKHRS	Total hours open per Week (#)
NOCC	Number of businesses (#)	NWKER	Number of employees (#)
OWNOCC	Owner occupied or leased to tenant(s)	SCHED	Light scheduling
OWNOPR	Owner responsible for operation and maintenance of energy systems		
Climate Variables			
Abbreviation	*Description*	*Abbreviation*	*Description*
CENDIV	Census division	HDD65	Heating degree days (base 65) (#)
PUBCLIM	Building america climate region	CDD65	Cooling degree days (base 65) (#)
Dependent (Response) Variables			
Abbreviation	*Description*	*Abbreviation*	*Description*
TOTEUI	Total annual site EUI (#)	LTEUI	Lighting annual site EUI (#)
HTEUI	Heating annual site EUI (#)	VNEUI	Ventilation annual site EUI (#)
CLEUI	Cooling annual Site eUI (#)		

The CBECS survey design uses sample weights (FINALWT), replicate weights (FINALWT1–FINALWT197) and estimation methodology (jackknife) to relate the sample to the total building population, and we used the *survey* package [76] in R to define these parameters for analysis. With a defined survey design, we fit ordinary least squares regression models and generalized linear regression models (GLM) to understand the relationships among the selected variables while accounting for the weightings associated with each building observation. The coefficient of determination (R^2) was used to measure of goodness of fit. The CBECS survey design models use likelihood ratio tests (LRTs) to calculate the R^2 because the dataset contains a relatively large number of observations and to account for weighting variables. A p-value < 0.05 (or a 95% confidence threshold) was used to test for statistical significance [77].

To evaluate significance or strong relationships between individual independent (predictor) variables and each response (dependent) variable, we fit simple regression models follow the form $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ where: β_0 is the intercept of the model, β_1 is the coefficient of the predictor variable X_1, ε is the residuals (or errors) of the model (assumed to be normally distributed), and Y is the response variable (*e.g.*, TOTEUI or HTEUI).

When considering multiple predictor variables for a single response variable, we used multivariate additive generalized linear regression. The model equation is in the form of $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \varepsilon$ where n is the number of independent variables considered. Additive multi-linear models assume the relationship between each predictor variable is only addition and evaluate the significance of each predictor variable and the R^2 goodness of fit for the entire model. We fit additive multi-linear regression models with all of the selected variable sets described in Table 1 for all five dependent (response) EUI variables. In addition to the full set of 32 variables, the window and envelope variable sets in Table 1 were also modeled in isolation. Although EUIs are normalized to building size, SQFT is evaluated with multi-linear analysis to ensure that a unique numerical variable is included in the regression models. We considered p-values as the criteria to eliminate predictor variables from multi-linear models to achieve a statistically significant reduced model with a minimum number of variables for each response variable.

As described previously, the CBECS survey includes many categorical variables, and in linear regression, categorical data requires *dummy coding* to expand each variable to multiple dichotomous variables, one for each level of the initial variable. The RENOV variable, for example, is coded as 1 = ‘Yes’ and 2 = ‘No’ which becomes two variables after dummy coding, RENOV1 and RENOV2, coded with binary values of 0 and 1. The general linear regression equation for these types of variables becomes $Y = \beta_0 + \beta_{1i}(I_{1i} \cdot X_{1i}) + \ldots + \beta_{ni}(I_{ni} \cdot X_{ni}) + \varepsilon$, where I is the binary indicator of each dummy coded level and i is the number of levels within the variable. When formated as categorical, the predictor variables use one level as the reference or *baseline* level which is included in the model intercept (β_0). The other levels in the variable become a change from the baseline accounted for in the regression coefficients. A potential challenge with categorical formated variables and dummy coding in regression models is the possibility of expanded categorical (especially binary) variables

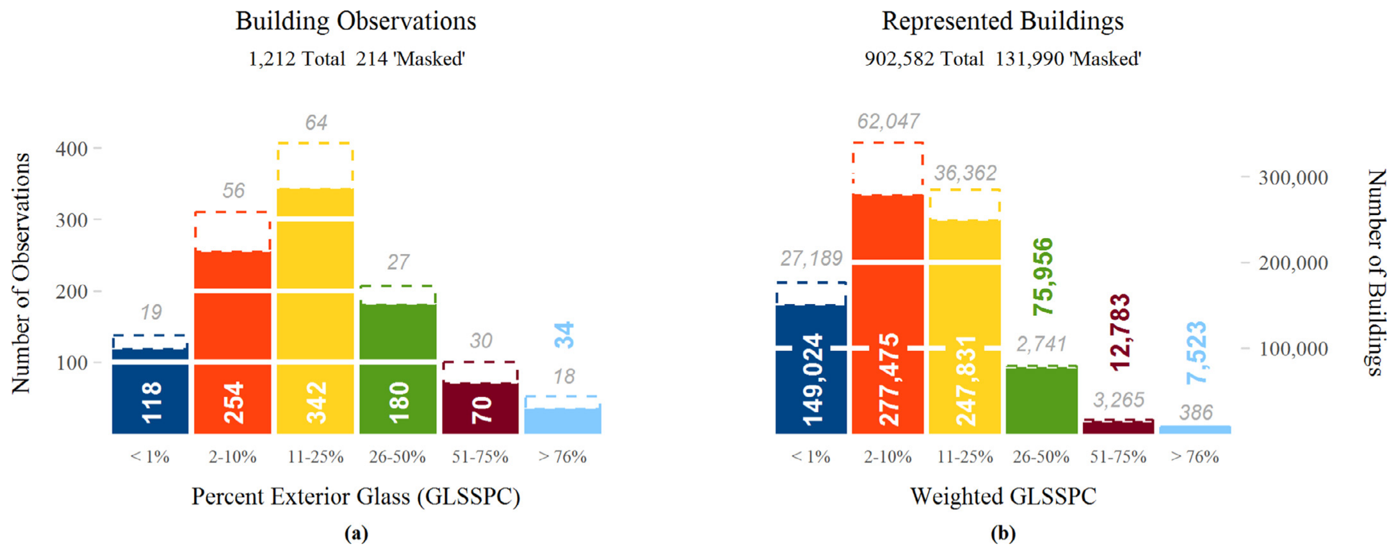

Fig. 1. Distribution of the number of CBECS (a) Building Observations and (b) Represented Buildings of the national building stock within each GLSSPC (WWR) interval. Larger rotated values in bold are the number of observations/buildings with masked values omitted. The dashed portion of each bar, labeled with smaller gray italics values is the number of observations/buildings where YRCON or NFLOOR contain a masked value.

becoming perfectly collinear, which prevents the R *survey* package from running, but this was not encountered.

4. Results & discussion

4.1. Descriptive statistics

The 1212 office building observations with a reported GLSSPC represent 902,582 office buildings of the entire national building stock. Of the 1212 observations, 76 have masked NFLOOR and an additional 138 have YRCON masked for confidentiality reasons as discussed above. Omitting these observations reduces the dataset to 998 observations. When considering the final sample weights (FINALWT) associated with each observation, the 998 CBECS building records represent 770,592 national stock buildings. We plotted the distribution of the number of office buildings for each GLSSPC level among the final office building dataset, and used descriptive and summary statistics to compare the GLSSPC to the five dependent annual energy use intensity (EUI) variables (TOTEUI, HTEUI, CLEUI, VNEUI, and LTEUI) to understand the distribution of buildings across the reported WWR. Fig. 1 shows the distribution of the CBECS building observations (a) and the national stock represented buildings (b) in each WWR (GLSSPC) interval within the office dataset. The larger bold, values listed vertically in each bar correspond to the 998 observations with the YRCON and NFLOOR masked values omitted. The smaller gray italic values indicate the number of observations and buildings removed because of these masked values.

The 11–25% WWR interval has the most observations while the 2–10% interval contains the greatest number of represented buildings. This indicates that while the masked (omitted) observations are relatively evenly distributed across glass percentage ranges in the CBECS data, the taller and especially older buildings represented generally have lower WWRs. The difference in distribution between *observations* and *represented buildings* emphasizes the importance of including the weights during any analysis of the CBECS data.

The overall average total EUI (TOTEUI) for the 1212 office buildings represented in the CBECS data is 130 kBtu/ft^2 per year. As seen in the end-use EUI break down in Fig. 2a, 20% of the total EUI is attributed to office equipment and computers, miscellaneous electricity use designated as ‘Other’ accounts for 10%, and an additional 5% labeled ‘Minors’ includes the sum of the smallest energy contributors (cooking, refrigeration, and water heating) in office buildings. Together these plug loads constitute to 35% of the total energy use in the CBECS office buildings and are not affected by WWR or envelope parameters. Heating energy at 32% is both the largest end-use affected by envelope, and in the sample broadly.

The operating energy is plotted for each GLSSPC interval in Fig. 2b, with each colored dot representing a single building observation, and the box plot showing the distribution. As WWR increases, the median TOTEUI increases, as does the interquartile range, which suggests buildings with higher WWR generally use more total energy. SI Table S2 includes similar boxplots and data tables of summary statistics for each end-use EUI *vs* GLSSPC. We would expect higher WWRs to be associated with (1) increased heating energy due to more conductive loss through larger windows (in cold climates and periods), (2) greater cooling energy use to counteract the conductive gains and additional solar radiation through increased window area (in warm climates and periods) and (3) decreased lighting energy because bigger windows let in more daylight. The descriptive statistics reveal that the median for heating and cooling EUI follow these expected trends but lighting median EUI also increases (details in SI Table S2), suggesting that, at least in this sample daylight is not displacing electric light even in buildings with large WWR.

Building age is a potential confounding factor, as building codes and window assemblies have dramatically advanced window thermal performance since the mid-20th century. Fig. 3 plots GLSSPC *vs* ‘Year of Construction’ (YRCON), with trendlines showing EUI performance within each WWR interval. In general, more recent YRCON corresponds to a lower TOTEUI, except for the 26–50% interval, which is relatively constant. The generally declining trend is most evident in buildings in the highest two WWR intervals, suggesting that over time, code requirements, and quality improvements in materials and construction have been effective at reducing energy use in buildings with highly glazed façades.

While consistent with expectations, and perhaps suggestive of relationships among variables, these simple descriptive statistics can be misleading, because they do not describe the strength of relationships between specific variables (most importantly

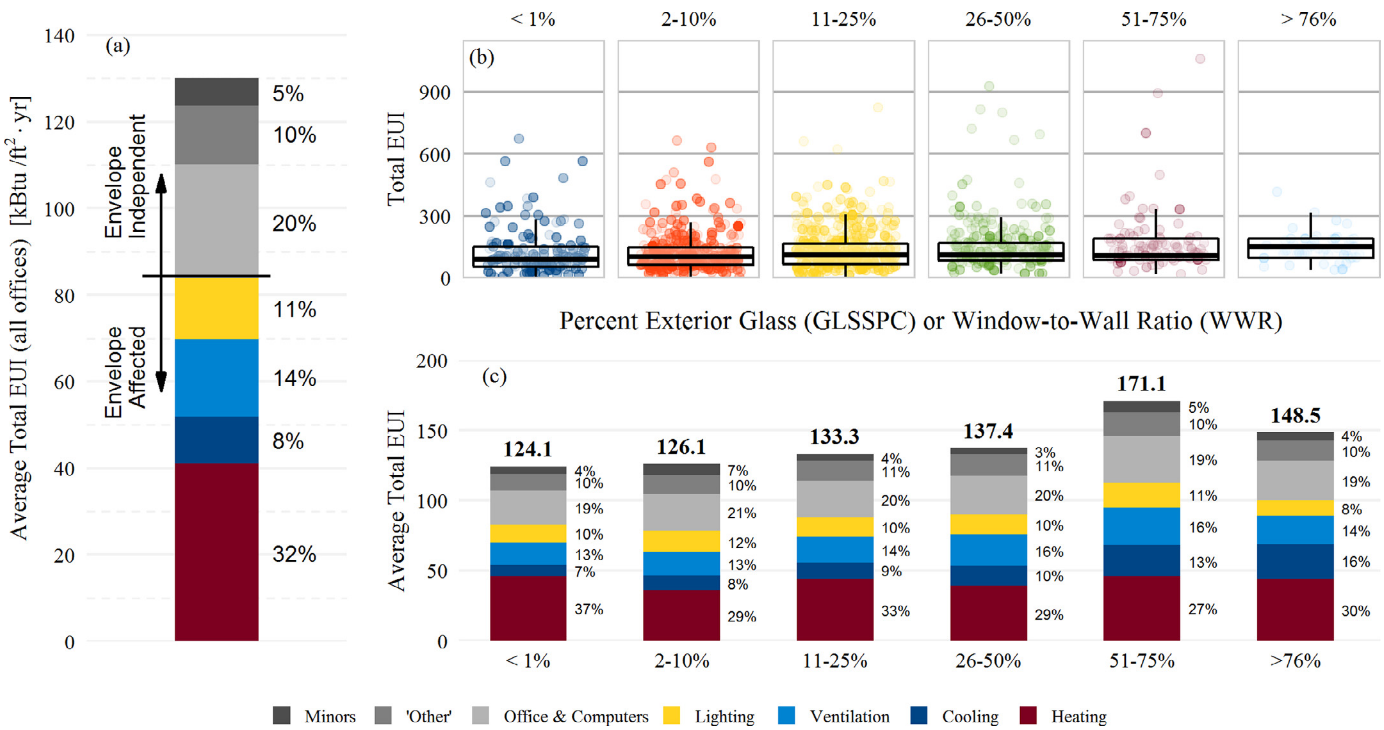

Fig. 2. Summary statistics and end use breakdown for Total EUI (TOTEUI) (a) Average of all 1212 office building observations. (b) Summary statistics by GLSSPC (WWR) interval and (c) End-use breakdown by GLSSPC. 'Other' is estimated miscellaneous electricity use and 'Minors' is a sum of the lowest contributors; cooking, refrigeration and water heating.

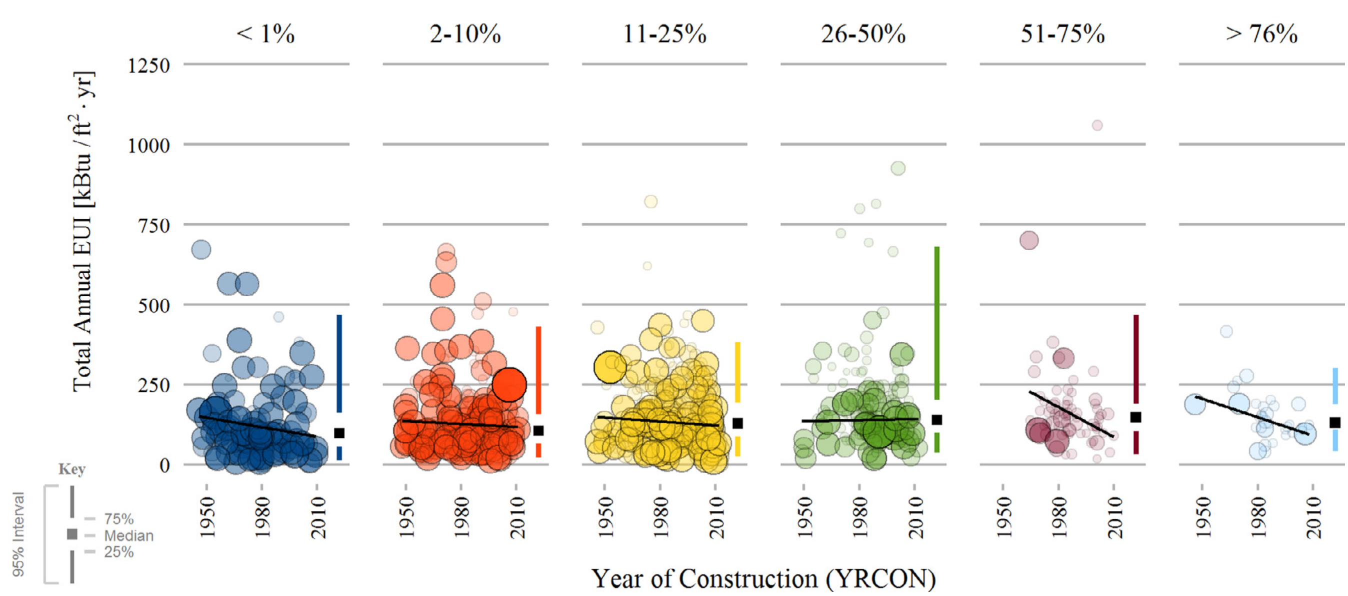

Fig. 3. Visual investigation of 'Total Annual EUI' (TOTEUI) *vs* 'Year of Construction' (YRCON) for each GLSSPC interval. Opacity and size of data points corresponds to sample weight (FINALWT) associated with each building observation. Vertical bars and black data point on the right of each GLSSPC interval are summary statistic plots showing the 95% confidence interval, inter-quartile range and median (key in lower left).

GLSSPC and energy), nor do they estimate our confidence in these relationships.

4.2. Regression analysis

4.2.1. Simple linear regression

All 32 independent variables were fit in separate simple regression models with each of the five EUI response variables to identify statistically significant parameters within each building energy end-use, and listed by end-use in Table 2. The full details of the results (R^2, coefficients, and p-values) for each single regression model are provided in Table S3 of the SI. For the categorical predictor variables, the results show significance (p-value) at each level. The goodness of fit measure (R^2) is evaluated for the entire model, regardless of the number of levels in a categorical variable, and the intercept of each regression model describes the reference level.

First considering WWR, the only statistically significant (p-value < 0.05) level of GLSSPC across all five EUIs is the < 1% GLSSPC interval, which is also the reference level, or intercept of the model. The total (TOTEUI) has only one additional significant

Table 2
Selected simple regression results. **Bold and underlined** R^2 value is the maximum of all models (CDD65) in CLEUI). ~~Struck through~~ p-values are insignificant (> 0.05). The model intercepts capture the reference levels in categorical variables and the intercept portion in numerical variables, denoted with (Ref) and (Int), respectively.

	Total EUI (TOTEUI)		Heating EUI (HTEUI)		Cooling EUI (CLEUI)		Lighting EUI (LTEUI)		Ventilation EUI(VTEUI))	
	R^2	p-value	R^2	p-value	R^2	p-value	R^2	p-value	R^2	p-value
RFCNS9 (Ref)	0.024	<0.001	0.023	<0.001	0.041	<0.001	0.017	<0.001	0.048	<0.001
RFCNS1		<0.001		~~0.140~~		<0.001		<0.001		<0.001
RFCNS2		<0.001		~~0.781~~		<0.001		<0.001		<0.001
RFCNS3		<0.001		~~0.087~~		<0.001		<0.001		<0.001
RFCNS4		<0.001		~~0.084~~		<0.001		<0.001		<0.001
RFCNS5		<0.001		~~0.102~~		<0.001		<0.001		<0.001
RFCNS6		<0.001		0.008		<0.001		<0.001		<0.001
RFCNS7		<0.001		0.005		<0.001		<0.001		<0.001
RFCNS8		<0.001		0.004		<0.001		<0.001		0.043
GLSSPC1 <1% (Ref)	0.008	<0.001	0.009	<0.001	0.040	<0.001	0.013	<0.001	0.014	<0.001
GLSSPC2 2–10%		~~0.372~~		<0.001		<0.001		<0.001		~~0.145~~
GLSSPC3 11–25%		~~0.121~~		~~0.987~~		<0.001		0.049		0.017
GLSSPC4 26–50%		0.011		~~0.835~~		<0.001		0.006		<0.001
GLSSPC5 51–75%		~~0.094~~		~~0.681~~		<0.001		<0.001		0.001
GLSSPC6 76–100%		~~0.096~~		~~0.669~~		<0.001		~~0.095~~		~~0.207~~
YRCON (Int)	0.004	0.027	0.009	~~0.080~~	0.001	~~0.420~~	0.000	~~0.772~~	0.001	~~0.105~~
YRCON		~~0.054~~		~~0.098~~		~~0.287~~		~~0.901~~		~~0.206~~
NFLOOR (Int)	0.009	<0.001	0.001	<0.001	0.002	<0.001	0.005	<0.001	0.020	<0.001
NFLOOR		<0.001		0.045		0.001		<0.001		<0.001
RENOV N/A (Ref)	0.018	<0.001	0.012	<0.001	0.003	<0.001	0.007	<0.001	0.012	<0.001
RENOV1 Yes		<0.001		<0.001		~~0.064~~		<0.001		<0.001
RENOV2 No		<0.001		<0.001		~~0.457~~		<0.001		<0.001
WINTYP (Int)	0.002	<0.001	0.006	<0.001	0.024	<0.001	0.004	<0.001	0.008	<0.001
WINTYP2		~~0.342~~		~~0.197~~		<0.001		~~0.077~~		0.013
WINTYP3		~~0.212~~		~~0.864~~		<0.001		~~0.764~~		~~0.283~~
SUNGLS (Int)	0.002	<0.001	0.004	<0.001	0.009	<0.001	0.001	<0.001	0.003	<0.001
SUNGLS1		~~0.206~~		~~0.606~~		~~0.625~~		~~0.435~~		~~0.222~~
SUNGLS2		~~0.502~~		~~0.628~~		0.001		~~0.213~~		~~0.067~~
SUNGLS3		0.024		<0.001		0.001		~~0.792~~		~~0.835~~
HDD65 (Int)	0.010	<0.001	0.155	~~0.870~~	0.165	<0.001	0.007	<0.001	0.001	<0.001
HDD65		0.013		<0.001		<0.001		0.008		~~0.208~~
CDD65 (Int)	0.006	<0.001	0.101	<0.001	**0.218**	0.006	0.004	<0.001	0.000	<0.001
CDD65		0.021		<0.001		<0.001		0.005		~~0.873~~

WWR range, 26–50%. The heating EUI (HTEUI) significant only at intercept and the lowest-glazed level 2–10%, which is quite surprising given the typically-direct relationship between window area and total envelope conductance. On the other hand, all levels of GLSSPC are significant in cooling EUI (CLEUI), which suggests a relationship between WWR and cooling energy. CLEUI also has the largest R^2 coefficient of determination, albeit only 0.04. In the lighting (LTEUI) model, all levels of GLSSPC are statistically significant except 76–100%, and in the ventilation (VNEUI) models levels are significant except 2–10% and 76–100%. The results suggest that the most pronounced effect of varying WWR on energy use in commercial office buildings will be on cooling loads, which is consistent with internally-dominated buildings. There appears to be a modest effect on lighting energy use, although only in less-glazed and not in the highest-glazed buildings, perhaps signaling that savings from daylighting have an initial benefit, but above a certain WWR, are overwhelmed by other energy costs.

Overall, these regression results for actual buildings stand in contrast to the simulation based results (typically performed for single, well-defined buildings) that have found strong correlations between WWR and heating and total EUI for office buildings in a variety of locations [31,63,79,80]. The lack of significance of WWR for *total* building energy use corroborates earlier work by Deng et al. that applied machine learning algorithms to CBECS microdata and also did not find correlation or inclusion of WWR in predictive artificial neural network models [78]. These findings emphasize the importance of considering a variety of buildings, locations, and use patterns when building predictive energy models.

Considering other individual predictor variables, all nine types (levels) of 'Roof Construction Material' (RFCNS) are significant in all models except in the heating model, in which only four levels are statistically significant. The explanatory power of this single regression model is greatest for cooling and ventilation energy use (R^2 = 0.04–0.05). Similarly, most the responses (levels) of 'Any Renovations' (RENOV) are significant, as expected from upgrades to building equipment and envelopes. The categorical variables describing fenestration 'Window type' (WINTYP) and 'Glass on sides with most sunlight' (SUNGLS) are significant in the cooling model, indicating the importance of solar gain during the summer season, but have a small explanatory power with R^2 values of 0.024 and 0.009, respectively. The numerical variables 'Number of Floors' (NFLOOR), 'Heating Degree Days (base 65)' (HDD65), and 'Cooling Degree Days (base 65)' (CDD65) are all statistically significant, with the strongest correlation of any independent and dependent variable occurring between cooling degree days and cooling energy use (CLEUI), with a R^2 of 0.218. These data suggest that climate cannot be neglected, regardless of interior loads, and may indicate the importance of HVAC systems. In contrast to the visual results shown in Fig. 3, YRCON is not statistically significant in any of the simple regression models, with the exception of intercept of the TOTEUI model, and that analysis excludes the older, masked buildings.

4.2.2. Multilinear additive regression

The single-regression suggests important relationships between predictor and response but cannot capture combined effects or interactions of multiple predictor variables. Of the multi-regression models, CLEUI (containing all independent variables) has the highest R^2 of 0.37. Similarly, for models with envelope-only and windows-only variables, the CLEUI model has greatest R^2 of 0.33 and 0.30, respectively. Details of these three models are in Table S4

Table 3
Reduced multi-linear additive regression models listing the variables that exceeded a confidence threshold of 95% in their contribution to each model and are therefore significant predictors. **Bold and underlined** R^2 value is the maximum of all the EUI models.

Dependent variable and reduced model formula (Categorical Formated)	R^2
Total EUI (TOTEUI) CENDIV + HDD65 + SQFT + RFCNS + NFLOOR + RENOV + WINTYP + RFCOOL + TINT + AWN + BASEMNT + ATTIC + CUBEC + NOCC + OWNOCC + WKHRS + NWKER + SCHED	0.169
Heating EUI (HTEUI) HDD65 + PUBCLIM + SQFT + WLCNS + RFCNS + NFLOOR + RENOV + AWN + BASEMNT + ATTIC + CUBEC + OWNOCC + OWNOPR	0.256
Cooling EUI (CLEUI) CDD65 + SQFT + RFCNS + GLSSPC + NFLOOR + RENOV + SUNGLS + RFTILT + BASEMNT + ATTIC + OWNOCC + WKHRS + NWKER + SCHED	**<u>0.341</u>**
Lighting EUI (LTEUI) CENDIV + SQFT + RFCNS + RENOV + DAYLTP + TINT + AWN + RFTILT + ATTIC + CUBEC + FEDFAC + OWNOCC + OWNOPR + WKHRS + NWKER + SCHED	0.186
Ventilation EUI (VNEUI) SQFT + WLCNS + RFCNS + NFLOOR + YRCON + RENOV + WINTYP + DAYLTP + RFCOOL + ATTIC + CUBEC + NOCC + OWNOCC + WKHRS + NWKER + SCHED	0.232

of the SI. The best-fit predictor variable formulas of the reduced regression models are shown in Table 3. Details of the all the multiple regression models are given in Table S5 of the SI. Overall, cooling loads (CLEUI) show the best fit, with a 14-variable model giving a R^2 of 0.341, a marked improvement from the single variable regression models. In contrast, the Total EUI model, has the lowest R^2 of 0.169, suggesting that predictive regression modeling of offices using CBECS data is better carried out at the end-use level, rather than for whole-building energy use. Considering WWR, GLSSPC is included in the CLEUI multi-regression model and is significant for all five levels but is only significant for the 2–10% and >76% levels in the LTEUI model, and is not included at all in the TOTEUI, HTEUI and VNEUI models. This mirrors the single variable results for GLSSPC described in the previous section.

Considering other predictor variables, 'Any Renovations' (RENOV) and 'Owner Occupied or Leased to Tenant' (OWNOCC) are significant in all five models, while 'Roof Construction Material' (RFCNS), 'Total Hours Open per Week' (WKHRS) and 'Number of Employees' (NWKER) are significant for all models except heating. In the same way, 'Attic' (ATTIC) and 'Percent Open Plan' (CUBEC) are significant for all models except cooling. The significance of these variables is logical in terms of physical building performance; renovations can strongly affect energy efficiency, particularly if equipment is replaced or buildings are updated to modern codes and standards. The number and working schedules of the occupants directly influence building operations, particularly as people and their equipment are a significant source of heat gains in commercial office buildings and therefore drive energy consumption for cooling and ventilation in most climates and seasons.

5. Conclusion

A statistical investigation of the influence of window-to-wall ratio (WWR) on building energy use was performed using the CBECS dataset. Descriptive statistics suggested average total EUI increases with WWR, and disaggregated results showing the largest increase among cooling loads. However, linear regression analysis revealed that WWR (GLSSPC) is statistically significant at all WWR levels only when predicting cooling energy use (CLEUI), both when considered individually (single) and in conjunction with other variables (multi-linear additive). In both cases, the explanatory power of the regression models is low, with a maximum R^2 of 0.341 for the multi-regression model. WWR is also statistically significant at most levels of glazing when predicting lighting (LTEUI) and ventilation (VNEUI) energy use but was not found to be statistically significant for heating loads (HTEUI). The modest predictive power of the multi-variable regressions suggest that complex interactions of building, occupant, and climate characteristics influence energy use, and that these interactions are not captured in survey variables.

These results for a large sample of actual commercial buildings broadly corroborate previous studies based on building energy simulation, but with a much lower explanatory power, and some significant exceptions. This finding emphasizes the wide variation in actual (measured) building characteristics, operations, and energy performance that is not captured in simulations. These results also highlight the limitations of the CBECS data: while simulation models precisely define physical performance of all building elements (*e.g.*, thermal resistance), this type of information is not currently collected by CBECS. Several categorical variables in CBECS represent the geometry of the buildings and the distribution, and orientation of windows, which directly affect the solar gains and therefore loads, such as 'Glass on sides with most sunlight' (SUNGLS), but are often generic Yes/No and thus offer little explanatory power. The regression model results do indicate some connections among climate, glazing, and energy consumption, but the presence of occupancy characteristics like 'Number of Employees' (NWKER) in the best-fitting multi-regression models underscores the importance of *internal* loads and operations for predicting energy use in commercial office buildings. In sum, this analysis of measured energy consumption indicates that while it affects energy consumption—particularly for cooling—the influence of WWR is more complex and perhaps less influential on the energy use of US office buildings than previously thought.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgements

This material is based on work supported by the National Science Foundation under Grant No. 1455450.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.enbuild.2019.109434.

References

[1] U.S. Energy Information Administration, March 2017 Monthly Energy Review, Office of Energy Statistics, U.S. Department of Energy, Washington, DC, 2017 https://www.eia.gov/mer.

[2] J. Apte, D. Arasteh, Window-related Energy Consumption in the US Residential and Commercial Building Stock, Lawrence Berkeley Natl. Lab., 2008 https://facades.lbl.gov/sites/all/files/60146.pdf.

[3] Windows and Glazing | WBDG Whole Building Design Guide, (n.d.). https://www.wbdg.org/resources/windows-and-glazing.

[4] ASHRAE, 2017 ASHRAE Handbook - Fundamentals, I-P, ASHRAE, Atlanta, GA, 2017.

[5] I.A. Macdonald, Quantifying the Effects of Uncertainty in Building Simulation, University of Strathclyde Glasgow, Scotland, 2002.

[6] T. Catalina, V. Iordache, B. Caracaleanu, Multiple regression model for fast prediction of the heating energy demand, Energy Build. 57 (2013) 302–312, doi:10.1016/j.enbuild.2012.11.010.

[7] Y. Pan, Z. Huang, X. Zheng, Data processing and data mining on energy consumption database of commercial buildings in shanghai, ASHRAE Trans. 115 (2009) 382–388 http://ezproxy.neu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=40505702&site=ehost-live&scope=site.

[8] J.C. Lam, S.C.M. Hui, A.L.S. Chan, Regression analysis of high-rise fully air-conditioned office buildings, Energy Build. 26 (1997) 189–197, doi:10.1016/S0378-7788(96)01034-1.

[9] J. Micheals, Results from the 2012 commercial buildings energy consumption survey (CBECS) energy usage data validation study, (2015). http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_171491.pdf?_ga=2.136851318.772686931.1499546974-511417512.1499546974.

[10] L.M. Tucker, Sustainable Building Systems and Construction for Designers, 2nd edition, Fairchild Books, an Imprint of Bloomsbury Publishing Inc, LondonNew York, 2015.

[11] K.L. Haglund, Decision-making methodology & selection tools for high-performance window systems in US climates, BEST2 Strateg. Issues Build. Des. WB13e4 (2010) http://www.nibs.org/resource/resmgr/BEST/BEST2_052_WB13-4.pdf.

[12] V. Granadeiro, J.P. Duarte, J.R. Correia, V.M.S. Leal, Building envelope shape design in early stages of the design process: integrating architectural design systems and energy simulation, Autom. Constr. 32 (2013) 196–209, doi:10.1016/j.autcon.2012.12.003.

[13] H. Islam, M. Jollands, S. Setunge, I. Ahmed, N. Haque, Life cycle assessment and life cycle cost implications of wall assemblages designs, Energy Build 84 (2014) 33–45, doi:10.1016/j.enbuild.2014.07.041.

[14] F. Goia, Search for the optimal window-to-wall ratio in office buildings in different European climates and the implications on total energy saving potential, Sol. Energy. 132 (2016) 467–492, doi:10.1016/j.solener.2016.03.031.

[15] S. Didwania, V. Garg, J. Mathur, Optimization of window-wall ratio for different building types, (2011). https://www.researchgate.net/profile/Srijan_Didwania/publication/259921312_OPTIMIZATION_OF_WINDOW-WALL_RATIO_FOR_DIFFERENT_BUILDING_TYPES/links/0c96052e8be7617c6e000000.pdf.

[16] H. Alibaba, Determination of optimum window to external wall ratio for offices in a hot and humid climate, Sustainability 8 (2016) 187, doi:10.3390/su8020187.

[17] E. Ghisi, J.A. Tinker, An ideal window area concept for energy efficient integration of daylight and artificial light in buildings, Build. Environ. 40 (2005) 51–61, doi:10.1016/j.buildenv.2004.04.004.

[18] F. Ascione, N. Bianco, R.F. De Masi, G.M. Mauro, G.P. Vanoli, Design of the building envelope: a novel multi-objective approach for the optimization of energy performance and thermal comfort, Sustainability 7 (2015) 10809–10836.

[19] R.A. Mangkuto, M. Rohmah, A.D. Asri, Design optimisation for window size, orientation, and wall reflectance with regard to various daylight metrics and lighting energy demand: a case study of buildings in the tropics, Appl. Energy. 164 (2016) 211–219, doi:10.1016/j.apenergy.2015.11.046.

[20] X. Su, X. Zhang, Environmental performance optimization of window–wall ratio for different window type in hot summer and cold winter zone in china based on life cycle assessment, Energy Build 42 (2010) 198–202, doi:10.1016/j.enbuild.2009.08.015.

[21] P. Rathi, Optimization of Energy Efficient Windows in Office Buildings For Different Climate Zones of the United States, Kent State University, 2012.

[22] I. Susorova, M. Tabibzadeh, A. Rahman, H.L. Clack, M. Elnimeiri, The effect of geometry factors on fenestration energy performance and energy savings in office buildings, Energy Build 57 (2013) 6–13, doi:10.1016/j.enbuild.2012.10.035.

[23] M. Alwetaishi, Impact of glazing to wall ratio in various climatic regions: a case study, J. King Saud Univ. - Eng. Sci. (n.d.). doi:10.1016/j.jksues.2017.03.001.

[24] S. Banihashemi, H. Golizadeh, M. Reza Hosseini, M. Shakouri, Climatic, parametric and non-parametric analysis of energy performance of double-glazed windows in different climates, Int. J. Sustain. Built Environ. 4 (2015) 307–322, doi:10.1016/j.ijsbe.2015.09.002.

[25] M.Y. Pai, Siddhartha, Effect of building orientation and window glazing on the energy consumption of hvac system of an office building for different climate zones, Int. J. Eng. Res. Technol. 4 (2015), doi:10.17577/IJERTV4IS090754.

[26] S. Stevanović, Optimization of passive solar design strategies: a review, renew, Sustain. Energy Rev. 25 (2013) 177–196, doi:10.1016/j.rser.2013.04.028.

[27] E. Lee, S. Selkowitz, V. Bazjanac, V. Inkarojrit, C. Kohler, High-performance Commercial Building Façades, Lawrence Berkeley Natl. Lab., 2002 http://escholarship.org/uc/item/7d30b3rd.pdf.

[28] ASHRAE, ANSI/ASHRAE/IES Standard 90.1-2016: Energy Standard For Building Except Low-Rise Residential Buildings, I-P, ASHRAE, Atlanta, GA, 2016.

[29] P. Ma, L.-S. Wang, N. Guo, Maximum window-to-wall ratio of a thermally autonomous building as a function of envelope U-value and ambient temperature amplitude, Appl. Energy. 146 (2015) 84–91, doi:10.1016/j.apenergy.2015.01.103.

[30] E. Gratia, A. De Herde, Design of low energy office buildings, Energy Build 35 (2003) 473–491, doi:10.1016/S0378-7788(02)00160-3.

[31] Q. Yang, M. Liu, C. Shu, D. Mmereki, M. Uzzal Hossain, X. Zhan, Impact analysis of window-wall ratio on heating and cooling energy consumption of residential buildings in hot summer and cold winter zone in china, J. Eng. (2015), doi:10.1155/2015/538254.

[32] F. Goia, M. Haase, M. Perino, Optimizing the configuration of a façade module for office buildings by means of integrated thermal and lighting simulations in a total energy perspective, Appl. Energy. 108 (2013) 515–527, doi:10.1016/j.apenergy.2013.02.063.

[33] R.A. Athalye, Y. Xie, B. Liu, M.I. Rosenberg, Analysis of Daylighting Requirements Within ASHRAE Standard 90.1, Pacific Northwest National Laboratory (PNNL), Richland, WAUS, 2013, doi:10.2172/1092662.

[34] H. Sanders, L. Podbelski, Dynamic façades: solving the high performance building challenge without design compromise | building research information knowledgebase, in: BEST4 conf. Proc., BEST4 Technical Committee, National Institute of Building Sciences, Kansas City, MI, 2015.

[35] Department of Energy Organization Act, 1977.

[36] National Research Council, Effective Tracking of Building Energy Use: Improving the Commercial Buildings and Residential Energy Consumption Surveys, National Academies Press, Washington, DC, 2012.

[37] P.A. Mathew, L.N. Dunn, M.D. Sohn, A. Mercado, C. Custudio, T. Walter, Big-data for building energy performance: lessons from assembling a very large national database of building energy use, Appl. Energy. 140 (2015) 85–93, doi:10.1016/j.apenergy.2014.11.042.

[38] A. Swenson, M. Johnson, EIA's Commercial Buildings Energy Consumption Survey: Strengths and Weaknesses of the Data, ACEEE, Pacific Grove, CA, 2002 http://aceee.org/files/proceedings/2002/data/index.htm.

[39] U.S. Energy Information Administration (EIA), About the commercial buildings energy consumption survey, Commer. Build. Energy Consum. Surv. CBECS. (n.d.). http://www.eia.gov/consumption/commercial/about.php.

[40] D. Overbey, The cbecs and why it's important, Environ. Des. Constr. 16 (2013) 42–43 http://ezproxy.neu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=87046947&site=ehost-live&scope=site.

[41] U.M. Advisory, Utility management advisory | understanding cbecs – A key foundation of the EPA's portfolio manager, (n.d.). http://umadvisory.org/understanding-cbecs-a-key-foundation-of-the-epas-portfolio-manager/ (Accessed 11 July 2017).

[42] D.W. Winiarski, M.A. Halverson, W. Jiang, DOE's commercial building benchmarks-development of typical construction practices for building envelope and mechanical systems from the 2003 cbecs, in: Proc. 2008 ACEEE Summer Study, ACEEE, Pacific Grove, CA, 2008.

[43] ENERGY STAR, Learn the benefits, ENERGY STAR. (2019). https://www.energystar.gov/buildings/facility-owners-and-managers/existing-buildings/learn-benefits

[44] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: a review, Renew. Sustain. Energy Rev. 23 (2013) 272–288, doi:10.1016/j.rser.2013.03.004.

[45] R. Evins, A review of computational optimisation methods applied to sustainable building design, Renew. Sustain. Energy Rev. 22 (2013) 230–245, doi:10.1016/j.rser.2013.02.004.

[46] A.P. Kaskhedikar, Regression tree-based methodology for customizing building energy benchmarks to individual commercial buildings, M.S., Arizona State University, 2013. http://search.proquest.com/docview/1433075181/abstract/70CA50872EC4476APQ/1.

[47] J.S. Hygh, J.F. DeCarolis, D.B. Hill, S. Ranji Ranjithan, Multivariate regression as an energy assessment tool in early building design, Build. Environ. 57 (2012) 165–175, doi:10.1016/j.buildenv.2012.04.021.

[48] R. Sullivan, S. Nozaki, R. Johnson, S. Selkowitz, Commercial building energy performance analysis using multiple regression procedures, ASHRAE Semi-Annu. Conf., 1983.

[49] H. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (2012) 3586–3592, doi:10.1016/j.rser.2012.02.049.

[50] M. Yalcintas, U. Aytun Ozturk, An energy benchmarking model based on artificial neural network method utilizing us commercial buildings energy consumption survey (CBECS) database, Int. J. Energy Res. 31 (2007) 412–421, doi:10.1002/er.1232.

[51] A. Kaskhedikar, T.A. Reddy, G. Runger, Use of random forest algorithm to evaluate model-based eui benchmarks from cbecs database, ASHRAE Trans 121 (2015) 17+ http://libraries.state.ma.us.ezproxy.neu.edu/login?gwurl=http://go.galegroup.com.ezproxy.neu.edu/ps/i.do?p=AONE&sw=w&u=mlin_b_northest&v=2.1&it=r&id=GALE%7CA461945334&asid=2b696d77641b9db4893cc816d4b44b13 .

[52] X. Gao, A. Malkawi, A new methodology for building energy performance benchmarking: an approach based on intelligent clustering algorithm, Energy Build 84 (2014) 607–616, doi:10.1016/j.enbuild.2014.08.030.

[53] J.C. Lam, K.K.W. Wan, K.L. Cheung, L. Yang, Principal component analysis of electricity use in office buildings, Energy Build 40 (2008) 828–836, doi:10.1016/j.enbuild.2007.06.001.

[54] E. Wang, Z. Shen, K. Grosskopf, Benchmarking energy performance of building envelopes through a selective residual-clustering approach using high di-

mensional dataset, Energy Build 75 (2014) 10–22, doi:10.1016/j.enbuild.2013.12.055.

[55] J.C. Lam, K.K.W. Wan, S.L. Wong, T.N.T. Lam, Principal component analysis and long-term building energy simulation correlation, Energy Convers. Manag. 51 (2010) 135–139, doi:10.1016/j.enconman.2009.09.004.

[56] S. Asadi, S.S. Amiri, M. Mottahedi, On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design, Energy Build 85 (2014) 246–255, doi:10.1016/j.enbuild.2014.07.096.

[57] D.H.W. Li, S.L. Wong, K.L. Cheung, Energy performance regression models for office buildings with daylighting controls, Proc. Inst. Mech. Eng. Part J. Power Energy 222 (2008) 557–568, doi:10.1243/09576509JPE620.

[58] S.S. Amiri, M. Mottahedi, S. Asadi, Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S, Energy Build 109 (2015) 209–216, doi:10.1016/j.enbuild.2015.09.073.

[59] J. Karlsson, A. Roos, Annual energy window performance *vs*. glazing thermal emittance — the relevance of very low emittance values, Thin Solid Films 392 (2001) 345–348, doi:10.1016/S0040-6090(01)01055-0.

[60] D.H.W. Li, E.K.W. Tsang, An analysis of daylighting performance for office buildings in Hong Kong, Build. Environ. 43 (2008) 1446–1458, doi:10.1016/j.buildenv.2007.07.002.

[61] K.T. Chan, W.K. Chow, Energy impact of commercial-building envelopes in the sub-tropical climate, Appl. Energy. 60 (1998) 21–39, doi:10.1016/S0306-2619(98)00021-X.

[62] R. Azari, Integrated energy and environmental life cycle assessment of office building envelopes, Energy Build 82 (2014) 156–162, doi:10.1016/j.enbuild.2014.06.041.

[63] S.K. Alghoul, H.G. Rijabo, M.E. Mashena, Energy consumption in buildings: a correlation for the influence of window to wall ratio and window orientation in Tripoli, Libya, J. Build. Eng. 11 (2017) 82–86, doi:10.1016/j.jobe.2017.04.003.

[64] M.N. Inanici, F.N. Demirbilek, Thermal performance optimization of building aspect ratio and south window size in five cities having different climatic characteristics of Turkey, Build. Environ. 35 (2000) 41–52, doi:10.1016/S0360-1323(99)00002-5.

[65] M.K. Urbikain, J.M. Sala, Analysis of different models to estimate energy savings related to windows in residential buildings, Energy Build 41 (2009) 687–695, doi:10.1016/j.enbuild.2009.01.007.

[66] M.-L. Persson, A. Roos, M. Wall, Influence of window size on the energy balance of low energy houses, Energy Build 38 (2006) 181–188, doi:10.1016/j.enbuild.2005.05.006.

[67] D. Arasteh, S. Selkowitz, J. Apte, M. LaFrance, Zero Energy Windows, Lawrence Berkeley Natl. Lab, 2006 http://escholarship.org/uc/item/2zp5m6x8.pdf.

[68] Z. Yu, W.L. Zhang, T.Y. Fang, Impact of building orientation and window-wall ratio on the office building energy consumption, Appl. Mech. Mater. 409–410 (2013) 606–611 doi:10.4028/www.scientific.net/AMM.409-410.606.

[69] I. Korolija, Y. Zhang, L. Marjanovic-Halburd, V.I. Hanby, Regression models for predicting uk office building energy consumption from heating and cooling demands, Energy Build 59 (2013) 214–227, doi:10.1016/j.enbuild.2012.12.005.

[70] J.W. Lee, H.J. Jung, J.Y. Park, J.B. Lee, Y. Yoon, Optimization of building window system in asian regions by analyzing solar heat gain and daylighting elements, Renew. Energy. 50 (2013) 522–531, doi:10.1016/j.renene.2012.07.029.

[71] U.S. Energy Information Administration (EIA), Estimation of energy end-use consumption, Commer. Build. Energy Consum. Surv. CBECS. (2016). https://www.eia.gov/consumption/commercial/estimation-enduse-consumption.php.

[72] U.S. Energy Information Administration (EIA), User's Guide to the 2012 CBECS Public Use Microdata File, U.S. Department of Energy, Washington, DC, 2016 https://www.eia.gov/consumption/commercial/data/2012/pdf/user_guide_public_use_aug2016.pdf.

[73] J. Hensen, R. Lamberts (Eds.), Building Performance Simulation For Design and Operation, Spon Press, London, New York, 2011.

[74] N. Lechner, Heating, Cooling, Lighting: sustainable design methods for architects, 3rd ed, John Wiley & Sons, Hoboken, N.J, 2009.

[75] U.S. Energy Information Administration (EIA), Variable and response codebook, (2016). https://www.eia.gov/consumption/commercial/data/2012/xls/2012microdata_codebook.xlsx.

[76] T. Lumley, survey: analysis of complex survey samples, 2018. https://CRAN.R-project.org/package=survey.

[77] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, 6th edition, John Wiley and Sons, Inc, Hoboken, NJ, 2014.

[78] H. Deng, D. Fannon, M.J. Eckelman, Predictive modeling for us commercial building energy use: a comparison of existing statistical and machine learning algorithms using cbecs microdata, Energy Build. 163 (2018) 34–43, doi:10.1016/j.enbuild.2017.12.031.

[79] H. Samuelson, S. Claussnitzer, A. Goyal, Y. Chen, A. Romo-Castillo, Parametric energy simulation in early design: high-rise residential buildings in urban contexts, Build. Environ. 101 (2016) 19–31, doi:10.1016/j.buildenv.2016.02.018.

[80] C. Marino, A. Nucara, M. Pietrafesa, Does window-to-wall ratio have a significant effect on the energy consumption of buildings? A parametric analysis in Italian climate conditions, J. Build. Eng. 13 (2017) 169–183, doi:10.1016/j.jobe.2017.08.001.