Manuscript Submitted to:
Environmental Science and Technology Word Count: **5402**

A New Estimate of Building Floor Space in North America

Jay H. Arehart [a,b*], Francesco Pomponi [b], Bernardino D'Amico[b], and Wil V. Srubar III[a,c]

[a] Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, ECOT 441 UCB 428, Boulder, Colorado 80309-0428 USA

[b] Resource Efficient Built Environment Lab (REBEL), Edinburgh Napier University, Edinburgh, UK

[c] Materials Science and Engineering Program, University of Colorado Boulder

[*] Corresponding Author, E: jay.arehart@colorado.edu

Keywords: floor space, building stock, North America, machine learning

Abstract

Floor space is a key variable used to understand the energy and material demands of buildings. Using recent datasets of building footprints, we employ a random forest regression model to estimate the floor space of the North American building stock. Our estimate for floor space in 2016 is 88,033 (+15,907 / -21,861) million m^2—which is 2.9 times higher than current estimates from national statistics offices. We also show how floor space per capita (m^2 cap^{-1}) is not constant across the North American region, highlighting the heterogeneous nature of building stocks. As a critical variable in integrated assessment models to project energy and material demands, this result suggests the following: (1) the North American building stock is more energy efficient than previously realized, suggesting that buildings are underutilized, (2) the embodied environmental impacts of buildings have been underestimated in comparison to operational impacts, and (3) the near-term demand for floor space and, consequently, the future demand for materials and energy have been largely underestimated.

1. Introduction

To meet mid-century targets, the building industry must adapt to reduce greenhouse gas emissions and limit maximum global temperature rise to less than 1.5 °C. If left unaltered, the construction industry alone would be responsible for up to 60% of the projected remaining carbon budget of 340 $GtCO_2$ [1]. This presents a challenge, given that the total gross area of the built environment is expected to increase 173% by 2050 [2], and buildings will drive a doubling of raw material consumption by 2060 [3].

Many models have been developed to estimate the future resource demand of the global building stock [4]. These models utilize either "top-down" or "bottom-up" approaches. In "top-down" methods, total resource consumption (*e.g.*, energy, material) and floor space are known or estimated for the entire building stock [5]. Useful statistics are then derived and reported per unit area of the total building stock. In a "bottom-up" approach [6–8], resource consumption is quantified per unit area (*e.g.*, kWh/m^2) for specific building typologies and multiplied by total gross areas of those building typologies within the building stock to estimate total resource consumption. Regardless of the approach, correct estimates of total building floor space are critical to ensure accurate quantification of current and future resource demand, since the results are directly proportional to the magnitude of those estimates. Floor space is typically estimated by using a floor space per capita for a particular region or per capita income level. Projections for regional or global floor space are estimated using the appropriate floor space per-capita estimates scaled by a projected population at specific income levels.

Two primary methodologies have been adopted to calculate a floor space per capita. The first methodology relies on government organizations to collect national-scale data, typically through surveys, on the number of occupants and building size, which are then aggregated into national-level statistics in a "top-down" fashion. For example, the Commercial and Residential Energy Consumption Surveys (CBECS and RECS, respectively) conducted by the US Energy Information Agency (EIA) utilize this methodology [9] to estimate floor space for both commercial and residential buildings. A similar methodology, relying on nationally published statistics of building permits issued, estimated that the residential building stock of the US was 21,846 million m^2 in 2010 [10]. From these estimates of floor space, floor space per capita are back-calculated (*e.g.,* square-meter of floor space/person or square-meter of floor space/USD) based upon the total population or gross domestic product (GDP) of the year the statistics were collected. An example of where this methodology has been applied is in the global-scale

EDGE model [6]. This model aggregates data from many national sources, including regions other than North America, to project regional floor space per capita based upon income levels, resulting in ranges between 31 m^2 cap^{-1} and 111 m^2 cap^{-1}. Likewise, another recent estimate of global floor space utilizes a total per-capita range between 7.56 m^2 cap^{-1} and 80.18 m^2 cap^{-1} [7].

The second methodology uses surveys of individual buildings to estimate floor space per-capita metrics for a particular building typology, then aggregating total floor space based upon the total population. For example, residential net floor space per capita was estimated for the US and Canada to range between 22.00 m^2 cap^{-1} and 50.98 m^2 cap^{-1}, depending upon the dwelling type, using a "bottom-up" modeling approach [11,12].

Each of these methods for calculating per capita floor space has inherent weaknesses. Survey methods at the regional or national level are not always transparent and may contain implicit or explicit biases. Likewise, uncertainties in the metrics are not always reported. Additionally, due to the resource intensity of capturing national-scale data, a limited number of surveys can be performed. For example, the most recent US EIA RECS surveyed only 5,600 of the estimated 118.2 million dwellings [13]. Another weakness of all methods is their inability to capture the underutilization of floor space. Surveys of occupied buildings will fail to capture unoccupied or underutilized building spaces. While capturing occupied buildings is certainly useful for modeling operational energy demand, it will underestimate the building stock's demand for materials and their associated embodied emissions. To date, there has been no "bottom-up" survey of all North American buildings to quantify total floor space, which is the aim of this work.

2. Aims and Objectives

Floor space estimates underpin many energy and material demand models, yet a single value is commonly used to estimate the extent of a region's building stock floor space. These estimates rely on small samples of the building stock, and so far, have failed to take advantage of recent advances made in deep learning and image classification. Thus, the present work has three objectives: (1) develop a methodology for using remotely sensed building footprint datasets to estimate total floor space of a building stock, (2) apply the methodology in the context of the of North American building stock, and (3) report the uncertainty of floor space per-capita estimates. We define the North America region to include the United States and Canada to align with the regional definitions used by other models (*e.g.*, International Institute for Applied Systems Analysis (IIASA) Global Energy Assessment).

3. Methodology

Using recently available remotely sensed satellite imagery [14–16], we propose a new method to estimate floor space and floor space per capita. This methodology section is divided into subsections to discuss the datasets, the calculation of geometric features, the machine learning model, validation of the model, and the limitations.

3.1 Dataset Description and Validation

Three open-source datasets published by Microsoft were used in the analysis to quantify the North American building stock: (1) US building footprints [14], (2) Canada building footprints [15], and (3) US building footprints with height attributes [16]. The first and second datasets were derived by extracting building footprints from satellite imagery using deep neural networks, while the third dataset is a subset of the first with the height attribute determined through interpolation of a digital terrain model.

The US building footprint dataset was evaluated for its accuracy for three different urban areas (Los Angeles County, New York City, and Denver) by Heris *et al.*[17]. The authors found it to have precision (positive prediction value) between 98.2% and 99.5% for these urban areas and recall (sensitivity) between 93% to 99% for buildings larger than 200m^2. In addition to the self-reported accuracy metrics which accompany the published dataset, these additional metrics show that the dataset accurately identifies true-positives and has very few false positives in urban areas. Yet the Microsoft building footprint dataset's accuracy has not been assessed rigorously for non-urban areas. Heris *et al.*, identified many false positives which occurred in areas of open water, high elevation, light colors (*e.g.*, snow and white sand), and bare ground. These false positives often have large footprint areas, making them problematic for assessing the extent of the North American building stock.

To identify and remove these false positives, we perform an analysis under the assumption that the vast majority of the building stock is located near publicly accessible roads. Thus, the further a building footprint is from a road, the higher its probability of being a false positive. To do so, we aim to identify a threshold at which a footprint can be rejected, using logistic regression models. The distance between a footprint's centroid and nearest road (metric of distance-from-road) is calculated for each building footprint of the dataset. Datasets for the North American roadways consisted of the US TIGER/Line shapefiles [18] and Canadian Open Roadway Data [19] which include both paved and unpaved roads. To ensure sufficient representation of buildings far from roads, we apply a base-10 log-transform to the distance-from-road metric. We then create a ground-truth dataset to identify the optimal distance-from-road threshold for which to exclude footprints. To ensure sufficient representation along the tail of the distribution, and an adequate balance between building and non-building footprints, we sample 150 buildings within each standard deviation of the log-transformed distribution (visualized in **Figure S1.1**). This sample comprises of 10 sampling buckets ranging from -2SD to +8 SD from the mean. We reviewed Google satellite imagery for these 1,500 footprints by "hand" to create the ground-truth dataset for which each building is classified as a building (true positive, *e.g.*, a cabin in a forest), or a non-building (false positive, *e.g.,* a snow field at high elevation). 408 of the 1,500 footprints (27.2%) are identified to be false positives, showing sufficient sample size and representation of non-building footprints for the following logistic-regression models. We use these ground-truth classifications as the outcome variable, and distance-from-road threshold status (set to 1 if footprint distance-to-road is less than the threshold, and 0 otherwise) as the predictor variable in a series of logistic regression models. Thus, we are able to test the discriminative ability for a multitude of distance-from-road thresholds, to find which threshold optimally segregated true building footprints from non-building footprints. We test threshold distances ranging between 32 and 10,000m, and evaluate model accuracy using the area under the receiver-operating characteristic curve (AUC) and pseudo-R-squared metrics (*i.e.*, Naglekerke and McFadden R-squared).

The threshold distance which maximizes the AUC (and most adequately predicted false positives in the ground-truth dataset) is 487m with an AUC of 0.81. This AUC is considered excellent[20] as the model at this threshold has an 81% chance of correctly distinguishing true building footprints from non-building footprints. Across all thresholds, the model with inclusion/exclusion criterion set to 487m features a global maximum in accuracy as illustrated in **Figure 1** (neither larger nor smaller thresholds improve the identification of non-building footprints). Thus, filtering on the distance-from-road metric adequately removes false positives from the dataset. It is important to note that this threshold is conservative, given that a considerable number of true positives are removed at the expense of filtering the majority of the less frequent false positives. For instance, large shopping centers surrounded by parking lots can have a distance from roads greater than 487m and are removed based upon this analysis. More stringent distance-to-road thresholds have similar model performance (AUC_{150m} = 0.77 and AUC_{300m} = 0.80). To be conservative, we utilize a 300m as a threshold for the results presented in **Section 4**. Additional results from other threshold distances, both more and less conservative, are included in **Supplements S5 and S6**.

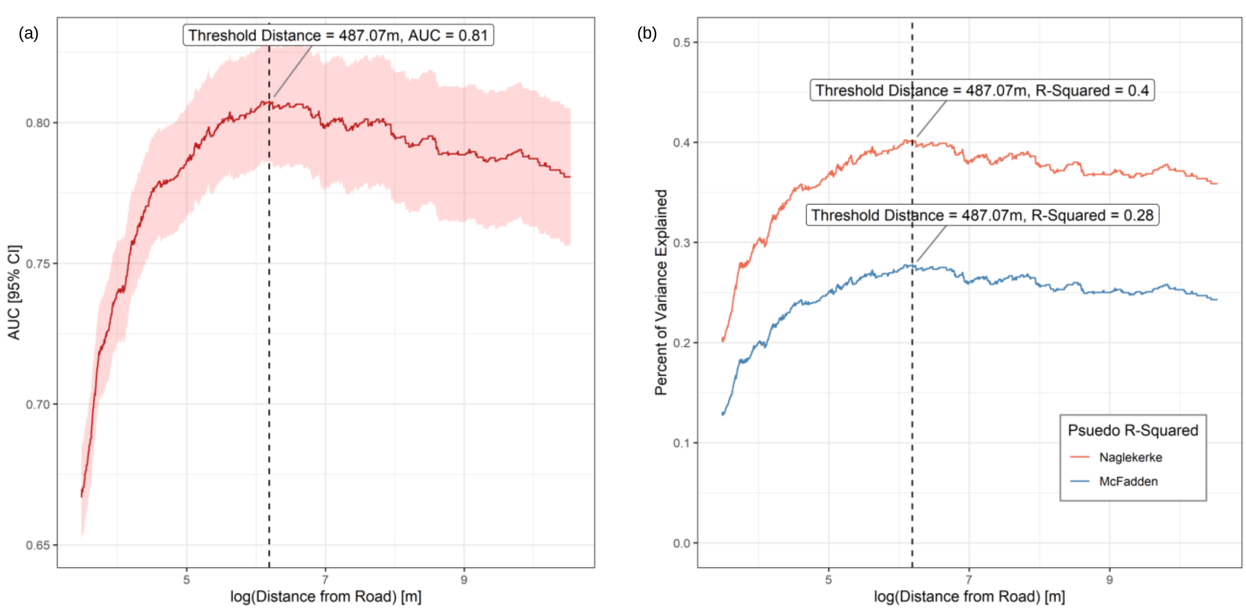


Figure 1. ***(a)*** *Area under the ROC curve (AUC) vs. log-transform of the distance from road with 95% confidence intervals.* ***(b)*** *Percent of variance explained using two pseudo R-squared metrics: Naglekerke and McFadden.*

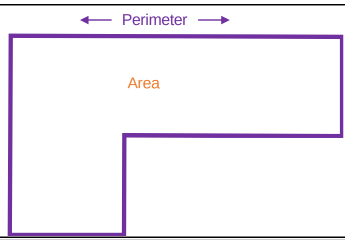
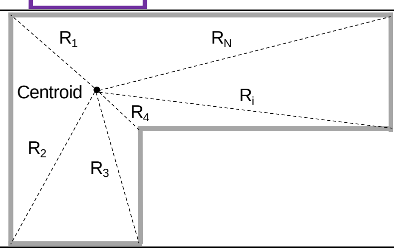
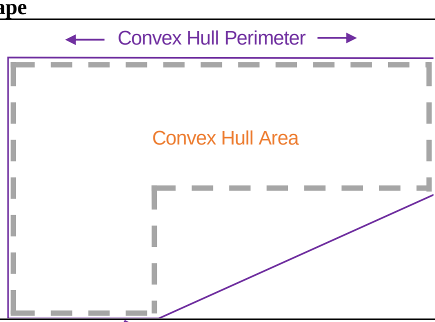
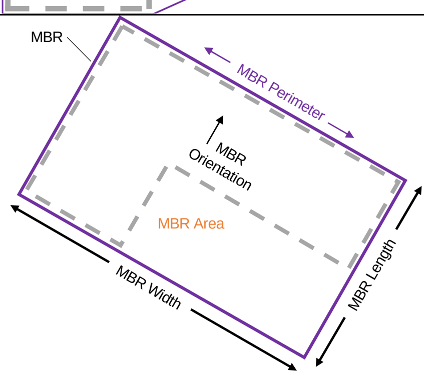
The third dataset which we use is a sample of the first dataset and consists of approximately 8 million footprints (6.4% of the total building stock identified by Microsoft) that were captured between 2014 and 2015.[16] The height attribute was determined through the interpolation of a digital terrain model. This dataset is considered the "training dataset" for the machine learning model and is deemed to be a representative sample of all buildings in North America. Additional outliers of each dataset are identified as buildings that are either extremely small, and unlikely to be inhabited (a footprint area smaller than 50 m^2), or extremely large (a footprint area larger than 10,000 m^2). Additionally, some outlier data points are identified in the dataset and removed. For example, the tallest building of the dataset is identified to be a water tower in Florida with a height of over 9 million meters. This building, and others with obviously egregious height errors are removed (n=12 of 7,993,302; *i.e.*, 0.00015%), based upon a visual inspection of the tallest 100 buildings identified in the dataset. Height detection of buildings in urban environments has typically been limited to the local scales (*i.e.*, cities) and regional and global estimates are a current focus of the remote sensing community [21].

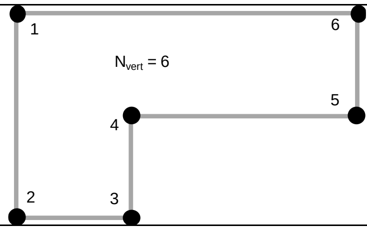
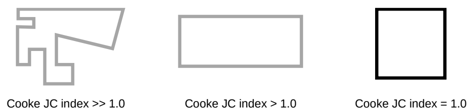
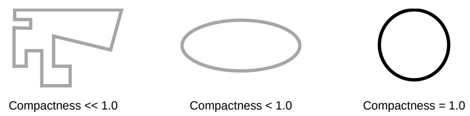
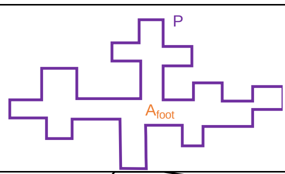
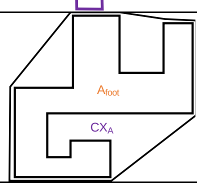
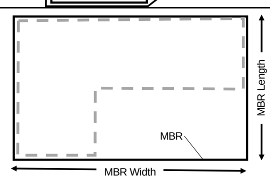
3.2 Geometric Features

Each building's footprint is defined by a GeoJSON string which describes the latitude and longitude of each of vertex of the footprint. From a building's footprint geometry, a large number of other metrics can be derived. We use 19 metrics ranging from the simple (*e.g.*, perimeter and area), to the complex (*e.g.*, compactness and fractality) as predictor variables for training a machine learning model. The metrics that are calculated for each building footprint are described in **Table 1**. The calculations are performed under four geographical projections depending upon the type of calculation (either distance or area) and region (Canada or United States). For area calculations, we use the USA Continental Equidistant Conic

(ESRI:102005) or Canada Albers Equal Area Conic (ESRI:102001) projections, while for distance or length calculations, we use the USA Continental Albers Equal Area Conic (ESRI:102003) or Canada Lambert Conformal Conic (ESRI:102002) projections.

Table 1. *Description of geometric characteristics of building footprint forms. Each metric is used as a predictor variable in the machine learning model which estimates building height.*

Variable	**Index**	**Notation or Equation**	**Description**
Size			
	Perimeter	P	Perimeter of the footprint.
	Area	A_{foot}	Area of the building footprint.
	Mean radius	$R_{mean} = \frac{1}{N}\sum_{i=1}^{N} R_i$	Mean distance from the building centroid to each vertex of the perimeter.
	Minimum radius	$R_{min} = \min(R_i)$	Minimum distance from the building centroid to each vertex of the perimeter.
	Maximum radius	$R_{max} = \max(R_i)$	Maximum distance from the building centroid to each vertex of the perimeter.
Shape			
	Convex hull perimeter	CX_P	Perimeter of the convex hull.
	Convex hull area	CX_A	Area of the convex hull.
	Minimum Bounding Rectangle		
	Perimeter	MBR_P	Perimeter of the minimum bounding rectangle.
	Area	MBR_A	Area of the minimum bounding rectangle.
	Width	MBR_w	Width of the minimum bounding rectangle.
	Length	MBR_l	Length of the minimum bounding rectangle.
	MBR Orientation	MBR_O	Orientation of the minimum bounding rectangle (MBR).

	Number of Vertices	n_{vert}	Number of vertices that make up the building footprint geometry.
	Cooke JC index	$\frac{P}{4\sqrt{A_{foot}}} - 1$	Measure of a footprint's shape efficiency with respect to a square [22].
	Compactness	$\frac{4\pi * A_{foot}}{P^2}$	Measure of a footprint's circularity or compactness [23].
	Fractality	$1 - \frac{\log(A_{foot})}{2 * \log(P)}$	Logarithmic ratio between the footprint area and perimeter [24].
	Concavity	$\frac{A_{foot}}{CX_A}$	Ratio of the footprint area to the area of the convex hull [24].
	Elongation	$\frac{MBR_l}{MBR_w}$	Ratio of the length of the MBR to the width of the MBR [25].

3.3 Machine Learning Model

With these predictor variables, the Python scikit-learn (v0.22.1) package [26] is employed to train machine learning models. Linear, ridge, support vector machine, gradient boosting, random forest, and Adaboost regression models are considered, using a 70/30 training-to-test split of the training dataset. Testing mean absolute error (MAE) and model training time are used to evaluate the performance of each model. Out of each of the six models, a random forest regression model has the highest performance. Random forest regression models are a commonly used ensemble learning method which predict a response (in this case building height) based upon the average prediction of many decision trees. A tree is formed by creating branches of nodes. Nodes are created at points when the input variables minimize the variance of the response variable. Subsequent nodes are then created along a branch and its length controlled by the depth of the tree. Controlling the depth of the tree is important to avoid the overfitting of the model. While useful for high dimension data, random forests are prone to overfitting and are difficult to visualize, although overfitting can be minimized by tuning the hyperparameters. Upon a large grid search of hyperparameters with 5-fold cross-validation, the random forest model found to have the lowest testing MAE consists of 100 trees, makes selections at each node using mean standard error, has a depth of 10 nodes, is limited to 9 estimators, and for a node to be created requires at least 9 samples. This model has a training mean absolute error (MAE) of 1.98 m, a training root mean square error (RMSE) of 3.49 m, a testing MAE of 1.98 m and testing RMSE of 3.52 m. There is negligible difference between the training and testing MAE suggesting that the identified random forest model is not overfit. This low MAE is a result of the error of predicting an individual building's height being insignificant in the context of the entire building stock.

While the structure of the forest is difficult to visualize, the predictor variable importance is summarized in **Figure 2**. Surprisingly, the orientation of the minimum bounding rectangle and fractality of a building footprint have the most predictive power within the random forest model. We suspect this result to be attributed to the fact that tall buildings are nearly always located in dense urban areas, while shorter buildings (*i.e.*, residential buildings) are often located along curved streets in suburbs. In the US and Canada, many city centers have grid-like road structures oriented along north-south meridians and east-west circles[27]. Thus, the orientation of the minimum bounding rectangle is useful for separating out buildings based upon their heights. Likewise, tall buildings are typically simple in their building footprint and do not consist of multiple wings. Complex building footprints (higher fractality) are often shorter, as it is less efficient to construct a tall building with a complex footprint. For additional details regarding the random forest model and predictor variables, see **Supplement S3** for a link to the code and data repositories.

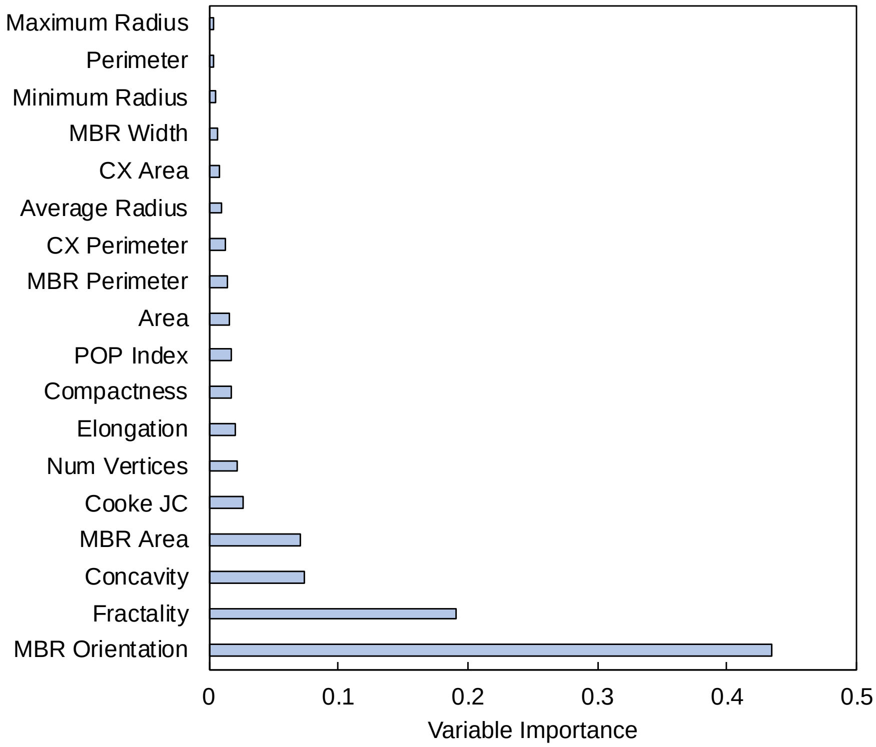


Figure 2. *Variable importance of the predictors used in the random forest regression model.*

This random forest regressor model is applied to all footprints in the US to predict the building height (h_i). To quantify uncertainty around each height prediction, the test MAE is added and subtracted to create three height predictions for each building footprint (upper, lower, and predicted). We then estimate the total floor space of North American buildings according to the following:

$$F_T = \sum_{i=1}^{N} A_{foot,i} * \left(\left\lfloor \frac{h_i}{h_{in}} \right\rfloor - 1\right) \qquad \textbf{(Eq. 1)}$$

where, F_T is the total gross floor space, $A_{foot,i}$ is the individual footprint area, h_i is the estimated height of each footprint, h_{in} is the interstory height, and N is the total number of polygons identified from the satellite imagery. h_i and h_{in} are visualized in **Figure 3** for a building with a pitched roof. The number of stories is determined by dividing the estimated height by the interstory height, rounding down to only count full stories, and subtracting one story. For the building in **Figure 3**, Equation 1 calculates the building to have two stories. One story is subtracted due to the height prediction being that of the maximum height. This value is determined by a visual inspection of the ground truth number of stories of 150 buildings against the predicted number of stories using Google street view imagery [28]. The buildings sampled are identified in **Supplement S3**. On average, the ground truth number of stories was determined to be 1.10 stories less than the only using the predicted height since the predicted height is recorded as the maximum height of the building, rather than the average height.

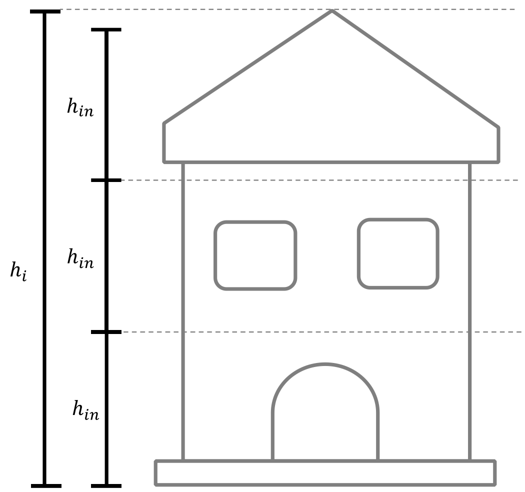


Figure 3. *Visual description of the height variables used by Equation 1.*

A floor space per capita is then estimated by dividing the total floor space by the total population of each region. It is assumed that the shape of each building is an extrusion of the building footprints and that only full stories contribute to floor space. This assumption ignores the fact that a building can have various heights for different parts of its footprint (*e.g.*, a building with wings may have various heights), or that a façade may be tiered with reduced floor space at higher stories. Because this is a large-scale regional analysis, and the optimal means of carrying vertical loads is through vertical elements[29], we consider the difference in floor space from buildings with discontinuous floor plates to be negligible. Furthermore, the optimal building form is cuboid in shape, meaning that it is an extrusion from a rectangular shape,[30] yet additional investigation is warranted for determining the extent to which vertically discontinuous buildings exist in the North American building stock.

3.4 Model Validation

To validate the random forest regression model's predictive power, the model is applied to all footprints in the US building footprints with height attributes [16] dataset. The total floor space is then calculated using both the reported height (from the training dataset) and predicted height (from the random forest regression model) with an interstory height of 3.6m. We assume that each building has an interstory height of 3.6m. This value is determined based upon a sensitivity analysis in which we considered various distributions of interstory heights. For the details of this sensitivity analysis, see **Supplement S2**. We find that using a single value of interstory height yields similar results to using other distributions. Interstory height has not been robustly measured in the North American building stock and would incrementally improve the present analysis.

3.5 Limitations

- A key assumption of this analysis is that the training dataset is representative of the entire North American building stock. While the data is taken from 44 states in the US, it has primarily coverage of urban areas rather than rural areas. Additionally, the error associated with quantifying building height using an interpolated digital terrain model is not expressed. While 150 buildings were visually checked for accurate prediction of number of stories using **Equation 1**, there still remains some uncertainty with the quality of the building height data used to train the model.
- Correctly estimating building height is an important component of the model, and this aspect can be improved upon as new methods for estimating building heights across large scales are developed [21]. While the quality of the comprehensive building footprint datasets [14,15] were validated for three metropolitan areas [17], they have not been validated across the entirety of North America. Thus, while we manually checked a small fraction of building footprints for their

accuracy and deemed them representative, it was not feasible to check the quality of the entire datasets due to their sheer size.

- In determining an appropriate distance to road threshold value, only 1500 building footprints were used. Additional ground-truth sampling of buildings far from roads might refine this criterion and provide further evidence for the best threshold distance.
- While the random forest model identified for predicting building heights in this study works well at the building stock scale, another model may be better suited to predict the height of an individual building. Furthermore, characteristics other than footprint geometry should be explored as predictor variables. While the random forest model used had sufficient predictive power for this study, it may not be the optimal model for predicting individual building heights, which would be useful for characterizing smaller-scale building stocks.

4. Results and Discussion

4.1 Estimate of Floor Space in North America

Across North America, our model estimates a total of 88,033 million m^2 of floor space with an upper bound of 103,940 million m^2 and a lower bound of 66,172 million m^2. Upper and lower estimates are associated with the error in estimating the height of an individual building footprint (see **Section 3.3**). When converting these estimates to floor space per capita, 242 m^2 cap^{-1} is predicted with an upper bound of 288 m^2 cap^{-1} and lower bound of 182 m^2 cap^{-1}. The distribution of floor space between the United States and Canada is 91.6% and 8.4%, respectively. Likewise, there is a 15% difference between the per capita floor space metrics for the two countries, with the US having 246 m^2 cap^{-1} and Canada having 210 m^2 cap^{-1}. **Figure 4a** and **4b** show the predicted floor space per capita estimate for each state or territory of the United States and Canada. Detailed results for these estimates are available in **Supplements 6 and 7**. No distinction is made herein between residential and non-residential buildings. However, the floor space could be subsequently disaggregated using high-fidelity data collected from other surveys. For example, the US HAZUS, estimates the US building stock to consist of 77.3% residential, 14.2% commercial, 3.0% public, and 5.5% agricultural and industrial [31].

The floor space per capita visualized in **Figure 4** show variation between administrative boundaries. For large-scale models that consider multi-national regions, using an average floor space per capita is appropriate. However, for sub-national analyses, floor space per-capita estimates vary greatly between states or territories, and especially, counties. For example, our analysis estimates that Denver county has a per capita floor space of 141 m^2 cap^{-1}, while some rural, sparsely populated counties have floor space per capita larger than 1000 m^2 cap^{-1}. We attribute this result to the large variation in economic activity between counties, limited land availability driving buildings to be smaller, and the disaggregation between the location of population centers and the location of buildings. While the metric of floor space per capita is commonly used for large-scale modeling purposes, caution should be taken when using this metric for analyses of sub-national building stocks. This analysis' primary aim was to estimate floor space per capita at the regional scale, so further investigation into individual counties was not performed, yet may yield interesting insights into the composition and heterogeneity of the North American building stock.

To elucidate whether floor space per-capita metrics are effective means of representing floor space, the correlation coefficient between population and floor space is computed for all administrative boundaries. A strong correlation at the state and province level is found (0.976), validating the use of floor space per capita for prognosticating the future demand and growth of floor space.

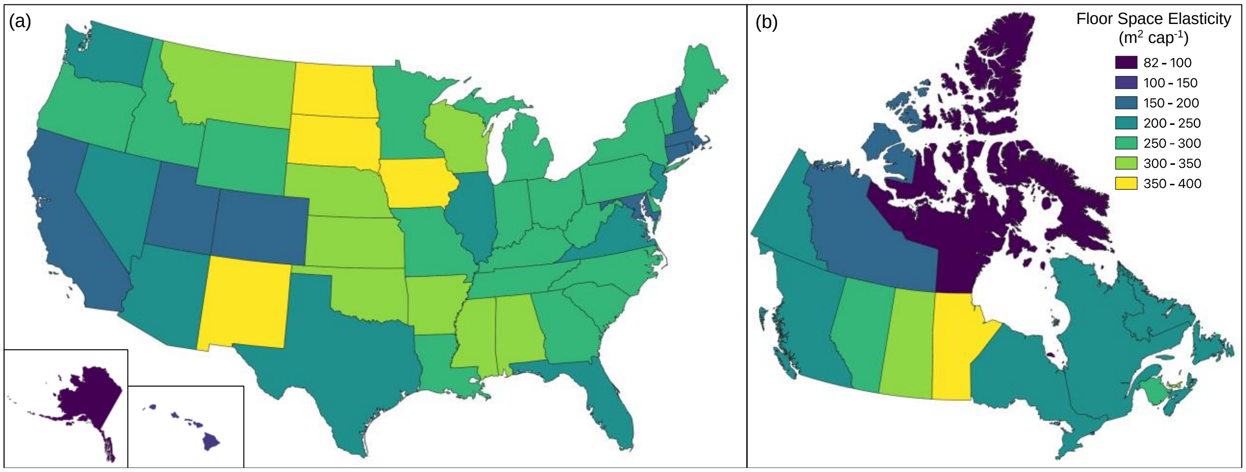


Figure 4. *(**a**) Floor space per capita (m^2 cap^{-1}) for each state in the USA. (**b**) Floor space per-capita (m^2 cap^{-1}) for each province and territory of Canada.*

4.2 Comparison to Existing Floor Space Estimates

Our estimated floor space per capita are 2.4 – 3.0 times greater than other estimates for the North American region. As previously discussed, a limiting assumption of our model is that building stock height data has not been as robustly validated to ground truths as the other building footprint datasets. We test this assumption by randomly forcing a percentage of the building stock to only be single-story. For example, a large department store may have a relatively tall building height, yet only be single-story. Our model would predict it to be a multistory building, when in reality it is only single story. The results of this analysis are shown in **Figure 5**, with the computed floor space per capita compared against other estimates [6,7,13,32–34]. Other estimates for floor space align well with one another, often relying on the same foundational data sets. If we assume the building footprint datasets used herein comprehensively represent the North American building stock, then to arrive at the floor space per capita used by other models, every building would be required to be single-story (or 0% considered multi-story). In other words, the total area of building footprints is equivalent to these other estimates. We know this not to be true, which demonstrates that these other per-capita floor space metrics drastically and conclusively underestimate the total floor space in North America.

The extent of the underestimation of the North American building stock by many models necessitates a reevaluation of the methodologies used to estimate floor space per capita. Potential discrepancies between estimates of per capita floor space may be attributed to buildings being under occupied, having higher-than-expected rates of unoccupied buildings, or national surveys not being representative of the total building stock. Moreover, the complexity and uncertainty of commercial floor space may not be accurately captured by these models.

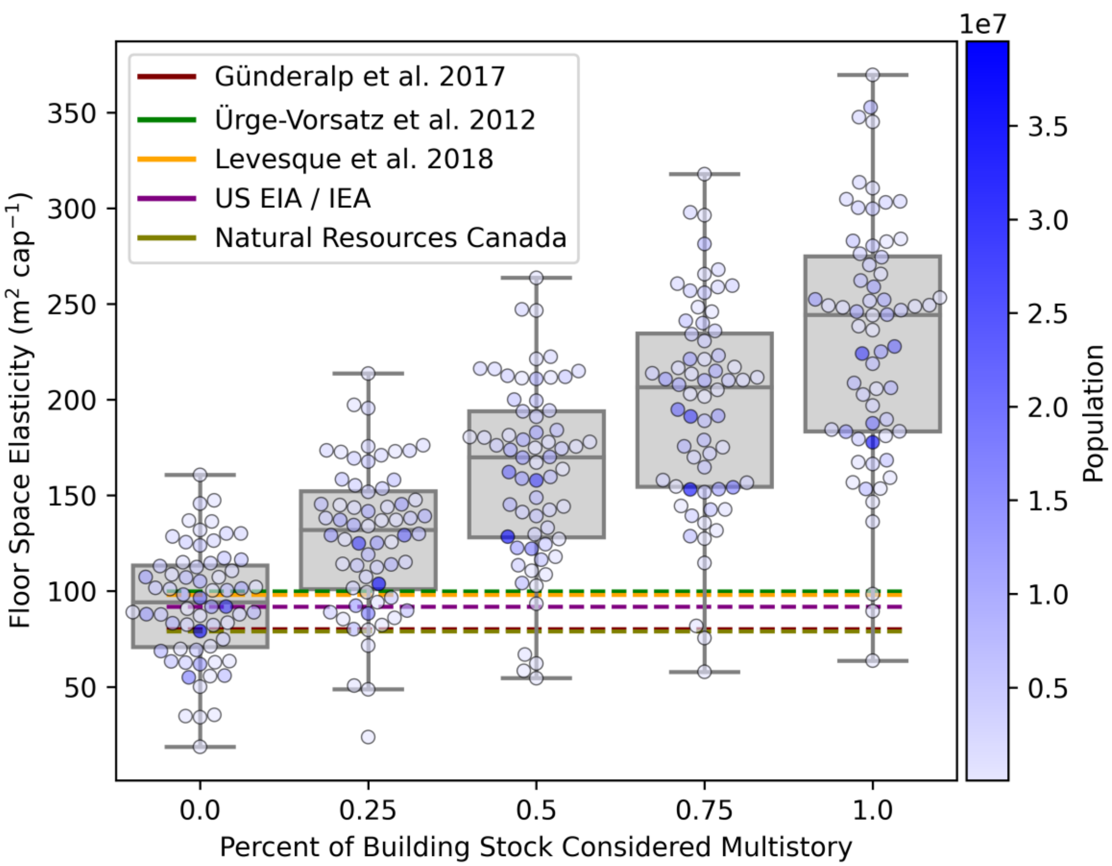


Figure 5. *Floor space per-capita for different administrative boundaries of the North American building stock compared against other estimates.*

4.3. Implications of Results

Floor space per capita is a critical variable in many global scale models, such as the MESSAGE [35], EDGE [6], and TIMER [36] models, which estimate the future energy demand of the building sector. Furthermore, the demand for construction materials[11,12], their availability for future reuse[37], and their potential to store carbon[38] also rely on this metric, as it is a key driver for projections. These resource demands are modeled using the Kaya identity methodology[39]. For example, residential energy demand can be modeled as:

$$E_{res} = \mathrm{h}\,\frac{p}{h}\frac{A}{p}\frac{E}{A} \qquad \textbf{(Eq. 2)}$$

where E_{res} is the total energy demand of a residential building stock, h is the number of households, (*p/h*) is the number of persons per household, (*A/p*) is the floor space per capita (or floor space elasticity), and (*E/A*) is the energy use intensity for a particular end-use (*e.g.*, space heating or space cooling). A similar approach can be taken for commercial buildings, using area divided by GDP as the use-intensity driver. With this modeling approach, the total resource demand of a building stock is directly proportional to the floor space per capita or floor space per unit of economic output. A second approach uses the floor space per-capita and simple physics-based models (*e.g.*, degree-day method) or regressions to estimate energy end-use demand for different building typologies[6]. In both of these approaches, the floor space per-capita metric is a critical component. While each model which uses the Kaya identity methodology has more complexity to it than the simple linear relationship presented in **Equation 2**, the total resource demand estimated by each analysis is directly proportional to the metric of floor space per capita. The results presented herein cause concern for the estimated resource consumption of the North American building stock, since the floor space estimates are 2.4 – 3.0 higher than the values used in other modeling efforts. There may be potential positive sides to this finding. For example, if the current underestimation is a result of underutilized floor space, then a significant opportunity exists to reduce the demand for new construction and its associated material and embodied emissions. With more floor space already available in the building stock, focus can shift from new construction to renovations and refurbishments.

Additionally, there is a lack of data that describe building stocks in the Global South. As these economies continue to develop, it is expected that their floor space per capita will also increase[7,35,40]. This expected increase is based upon estimates for North America and other developed economies. By underestimating the floor space per capita of higher-income level countries, the current projections for global floor space growth—especially in the Global South—may be vastly underestimated. This is concerning given that most of the growth in floor space is expected to occur in the Global South[7].

Bottom-up models for operational energy demand which utilize the Kaya identity methodology are often validated using top-down estimates. This validation suggests that the floor space per capita used is appropriately scaled for occupied, conditioned spaces. The results presented by our analysis utilizes a bottom-up approach, which considers all building footprints in North America, regardless of if they are unoccupied or unconditioned. To explore this discrepancy, we consider two scenarios. The first is that operational energy demand is appropriately modeled using floor space per-capita metrics derived from the US EIA and validated with top-down estimates. When considering the results from our analysis, this would mean that only up to one-third of the building stock is occupied and conditioned, an implausible scenario. An alternative scenario is that the building stock is much more energy efficient than previously realized, due to an extent of underutilization. While some unconditioned buildings (*e.g.*, agricultural and industrial) are included in our analysis, but excluded in the US EIA's analysis, they only contribute 5.5% of the total building stock[31], which is not enough to rectify the discrepancy observed.

Regardless, the fact that the building stock is 2.4 – 3.0 times larger than expected causes concern when modeling material projections and embodied carbon emissions, since all buildings will have this demand for material, regardless of whether or not they are fully conditioned. This underestimation of embodied carbon emissions is worrisome as much of the attention in the past decades has been paid to reducing operational emissions, when in fact embodied emissions are more significant than realized. By further characterizing the North American building stock, using bottom-up approaches, we will gain a better understanding of where the opportunities might lie to reduce life cycle energy demand and carbon emissions.

Understanding floor space per capita is essential to quantifying global resource demand. We present in this work a novel method for quantifying the floor space of the North American building stock. The results call for the reevaluation of how floor space per-capita metrics are calculated for building stocks throughout the world. The methodology for estimating floor space using satellite imagery and machine learning can be applied to other regions of the world, specifically the Global South, as high-quality data becomes available. These insights will improve systems-scale models for predicting global energy and material demand, existing material stocks in the built environment, and the carbon storage potential of the global building stock. In addition, newfound estimates of floor space per-capita metrics will aid in identifying and prioritizing building-related interventions required to minimize greenhouse gas emissions from the building sector.

Supporting Information

Details on distance from road threshold; sensitivity analysis of interstory height; link to code repository; validation of Equation 1; floor space per capita by USA county; additional results for different distance from road threshold values.

References

(1) Müller, D. B.; Liu, G.; Løvik, A. N.; Modaresi, R.; Pauliuk, S.; Steinhoff, F. S.; Brattebø, H. Carbon Emissions of Infrastructure Development. *Environ. Sci. Technol.* **2013**, *47* (20), 11739–11746. https://doi.org/10.1021/es402618m.

(2) International Energy Agency. *Transition to Sustainable Buildings: Strategies and Opportunities to 2050*; International Energy Agency: Paris, 2013.

(3) OECD Publishing. *Global Material Resources Outlook to 2060: Economic Drivers and Environmental Consequences*; Paris, 2019.

(4) Lanau, M.; Liu, G.; Kral, U.; Wiedenhofer, D.; Keijzer, E.; Yu, C.; Ehlert, C. Taking Stock of Built Environment Stock Studies: Progress and Prospects. *Environ. Sci. Technol.* **2019**, *53* (15), 8499–8515. https://doi.org/10.1021/acs.est.8b06652.

(5) Schaffartzik, A.; Mayer, A.; Gingrich, S.; Eisenmenger, N.; Loy, C.; Krausmann, F. The Global Metabolic Transition: Regional Patterns and Trends of Global Material Flows, 1950–2010. *Global Environmental Change* **2014**, *26*, 87–97. https://doi.org/10.1016/j.gloenvcha.2014.03.013.

(6) Levesque, A.; Pietzcker, R. C.; Baumstark, L.; De Stercke, S.; Grübler, A.; Luderer, G. How Much Energy Will Buildings Consume in 2100? A Global Perspective within a Scenario Framework. *Energy* **2018**, *148*, 514–527. https://doi.org/10.1016/j.energy.2018.01.139.

(7) Güneralp, B.; Zhou, Y.; Ürge-Vorsatz, D.; Gupta, M.; Yu, S.; Patel, P. L.; Fragkias, M.; Li, X.; Seto, K. C. Global Scenarios of Urban Density and Its Impacts on Building Energy Use through 2050. *Proc Natl Acad Sci USA* **2017**, *114* (34), 8945–8950. https://doi.org/10.1073/pnas.1606035114.

(8) Langevin, J.; Harris, C. B.; Reyna, J. L. Assessing the Potential to Reduce U.S. Building CO2 Emissions 80% by 2050. *Joule* **2019**, S2542435119303575. https://doi.org/10.1016/j.joule.2019.07.013.

(9) US EIA. *2015 RECS Square Footage Methodology*; Washington, DC, 2017; p 23.

(10) Moura, M. C. P.; Smith, S. J.; Belzer, D. B. 120 Years of U.S. Residential Housing Stock and Floor Space. *PLoS ONE* **2015**, *10* (8), e0134135. https://doi.org/10.1371/journal.pone.0134135.

(11) Marinova, S.; Deetman, S.; van der Voet, E.; Daioglou, V. Global Construction Materials Database and Stock Analysis of Residential Buildings between 1970-2050. *Journal of Cleaner Production* **2020**, *247*, 119146. https://doi.org/10.1016/j.jclepro.2019.119146.

(12) Deetman, S.; Marinova, S.; van der Voet, E.; van Vuuren, D. P.; Edelenbosch, O.; Heijungs, R. Modelling Global Material Stocks and Flows for Residential and Service Sector Buildings towards 2050. *Journal of Cleaner Production* **2019**, 118658. https://doi.org/10.1016/j.jclepro.2019.118658.

(13) US EIA. *Residential Energy Consumption Survey (RECS) - Energy Information Administration*; 2017.

(14) Microsoft. US Building Footprints. 2018.

(15) Microsoft. Canadian Building Footprints. 2019.

(16) Microsoft. Microsoft Buildings Footprint Training Data with Heights. ESRI March 1, 2019.

(17) Heris, M. P.; Foks, N. L.; Bagstad, K. J.; Troy, A.; Ancona, Z. H. A Rasterized Building Footprint Dataset for the United States. *Sci Data* **2020**, *7* (1), 207. https://doi.org/10.1038/s41597-020-0542-3.

(18) US Census Bureau. TIGER/Line Shapefiles https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html (accessed Feb 5, 2020).

(19) 2016 Census Road Network File - Open Government Portal https://open.canada.ca/data/en/dataset/57d5ffae-3048-4a19-9b4c-eab12f6322c5 (accessed Jul 6, 2020).

(20) Hosmer, D. W.; Lemeshow, S.; Sturdivant, R. X. *Applied Logistic Regression*; John Wiley & Sons, 2013; Vol. 398.

(21) Li, X.; Zhou, Y.; Gong, P.; Seto, K. C.; Clinton, N. Developing a Method to Estimate Building Height from Sentinel-1 Data. *Remote Sensing of Environment* **2020**, *240*, 111705. https://doi.org/10.1016/j.rse.2020.111705.
(22) Kouskoulas, V.; Koehn, E. Predesign Cost-Estimation Function for Buildings. *Journal of the Construction Division* **1974**, *100* (4), 589–604.
(23) Li, W.; Goodchild, M. F.; Church, R. An Efficient Measure of Compactness for Two-Dimensional Shapes and Its Application in Regionalization Problems. *International Journal of Geographical Information Science* **2013**, *27* (6), 1227–1250. https://doi.org/10.1080/13658816.2012.752093.
(24) Basaraner, M.; Cetinkaya, S. Performance of Shape Indices and Classification Schemes for Characterising Perceptual Shape Complexity of Building Footprints in GIS. *International Journal of Geographical Information Science* **2017**, *31* (10), 1952–1977. https://doi.org/10.1080/13658816.2017.1346257.
(25) Burghardt, D.; Steiniger, S. Usage of Principal Component Analysis in the Process of Automated Generalisation. In *Proceedings of 22nd international cartographic conference*; 2005; pp 9–16.
(26) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
(27) Boeing, G. Urban Spatial Order: Street Network Orientation, Configuration, and Entropy. *Appl Netw Sci* **2019**, *4* (1), 67. https://doi.org/10.1007/s41109-019-0189-1.
(28) Google. *Google Earth Pro v7.3.3*; Google, 2020.
(29) Steadman, P. Why Are Most Buildings Rectangular? *Arq* **2006**, *10* (2), 119–130. https://doi.org/10.1017/S1359135506000200.
(30) D'Amico, B.; Pomponi, F. A Compactness Measure of Sustainable Building Forms. *R. Soc. open sci.* **2019**, *6* (6), 181265. https://doi.org/10.1098/rsos.181265.
(31) FEMA. *HAZUS-MH Software*; US FEMA, 2019.
(32) Ürge-Vorsatz, D.; Petrichenko, K.; Miklos, M.; Staniec, M.; Labelle, M.; Ozden, E.; Labzina, E. *Best Practice Policies for Low Energy and Carbon Buildings. A Scenario Analysis*; Center for Climate Change and Sustainable Policy (3CSEP); Global Buildings Performance Network, 2012.
(33) IEA; UN Environment Programme. *2018 Global Status Report: Towards a Zero-Emission, Efficient and Resilient Buildings and Construction Sector*; 2018.
(34) US EIA. *Commercial Building Energy Consumption Survey (CBECS) 2012*; U.S. Energy Information Administration: Washington, D.C, 2012.
(35) Grubler, A.; Wilson, C.; Bento, N.; Boza-Kiss, B.; Krey, V.; McCollum, D. L.; Rao, N. D.; Riahi, K.; Rogelj, J.; De Stercke, S.; Cullen, J.; Frank, S.; Fricko, O.; Guo, F.; Gidden, M.; Havlík, P.; Huppmann, D.; Kiesewetter, G.; Rafaj, P.; Schoepp, W.; Valin, H. A Low Energy Demand Scenario for Meeting the 1.5 °C Target and Sustainable Development Goals without Negative Emission Technologies. *Nat Energy* **2018**, *3* (6), 515–527. https://doi.org/10.1038/s41560-018-0172-6.
(36) Daioglou, V.; van Ruijven, B. J.; van Vuuren, D. P. Model Projections for Household Energy Use in Developing Countries. *Energy* **2012**, *37* (1), 601–615. https://doi.org/10.1016/j.energy.2011.10.044.
(37) Stephan, A.; Athanassiadis, A. Towards a More Circular Construction Sector: Estimating and Spatialising Current and Future Non-Structural Material Replacement Flows to Maintain Urban Building Stocks. *Resources, Conservation and Recycling* **2018**, *129*, 248–262. https://doi.org/10.1016/j.resconrec.2017.09.022.
(38) Churkina, G.; Organschi, A.; Reyer, C. P. O.; Ruff, A.; Vinke, K.; Liu, Z.; Reck, B. K.; Graedel, T. E.; Schellnhuber, H. J. Buildings as a Global Carbon Sink. *Nat Sustain* **2020**, *3* (4), 269–276. https://doi.org/10.1038/s41893-019-0462-4.

(39) Ürge-Vorsatz, D.; Cabeza, L. F.; Serrano, S.; Barreneche, C.; Petrichenko, K. Heating and Cooling Energy Trends and Drivers in Buildings. *Renewable and Sustainable Energy Reviews* **2015**, *41*, 85–98. https://doi.org/10.1016/j.rser.2014.08.039.
(40) International Energy Agency. *Technology Roadmap Energy Efficient Building Envelopes*; 2013; p 68.

Acknowledgments

JA and FP gratefully acknowledge the financial support received for this research by Edinburgh Napier University, through the project "Carbon Storage of the Built Environment: Exploring the Theoretical Potential (CaSBE)" - Grant No. N452-000F. Additionally, JA gratefully acknowledges the financial support from the Temple Hoyne Buell Architectural Fellowship, the College of Engineering's Global Enrichment Fund, and the Living Materials Laboratory (LMLab) at the University of Colorado Boulder. This work represents the views of the authors and not necessarily those of the sponsors. The authors acknowledge and thank Chris Arehart for his input and insight regarding the methodology for dataset validation. Additionally, the authors are grateful to the three anonymous referees for their review and comments that helped improve the clarity and significance of the manuscript.

Author contributions

All authors conceptualized the research. JA developed the methods and performed the analysis. JA, FP, and BD contributed to the discussion and interpretation of the results. All authors contributed to the writing and editing of the manuscript.

Competing interests

The authors declare no competing interests.

For Table of Contents Only