Whole Building Estimates: Methodology

Table *of* Contents

Executive Summary

Whole Building Estimates is a data service that provides the real estate sector with the ability to obtain historical estimates of energy usage and carbon emissions for buildings[1]. The service takes a building's address along with several basic building features (e.g., building use type) as input, and then leverages machine learning (ML) and Measurabl's Quantum Database™ to produce monthly historical estimations. These monthly historical estimates can then be aggregated to obtain estimates for any building, which exhibits similar features to those in Measurabl's database, over any historical time period from 2017 onward.

Whole Building Estimates *at* Measurabl

Measurabl uses a three-step process to derive its building level energy and carbon estimates. First, energy usage estimates are computed using a ML model trained on Measurabl's database. The database includes time series energy usage records for tens of thousands of buildings and the various spaces within these buildings, across more than 100 primary use types. More specifically, the ML model is built on data from tens of thousands of buildings consisting of hundreds of thousands of monthly data points in the Measurabl database (see Appendix A for details on the ML model data). Measurabl applies a very strict data cleaning process to the underlying dataset to ensure maximum integrity and data quality of the data used for deriving Whole Building Estimates.

Next, the energy mix fraction for each building is computed based on similar buildings in Measurabl in terms of use type and location. The energy usage types that are trackable within Measurabl include electricity, fuel, and district (see Appendix C for further details). Finally, building energy usage estimates are transformed using the computed energy mix fractions and Measurabl's GHG Methodology to derive estimated carbon emissions.

[1]"Buildings" and "properties" can be used interchangeably throughout this document. A property may consist of multiple physical buildings.

The final estimates produced for a particular building represent the expectation of how much energy that building consumed as well as carbon it produced over the course of a specified historical time period.

One of the key strengths of Measurabl's models for Whole Building Estimates is the freshness of the training data. Utility data flows into Measurabl's database on a continuous basis, thus allowing us to update the training datasets on a monthly basis. After the end of each calendar month, the models are retrained to include any data that was added in the previous month. Utility data older than 2017 is discarded in order to keep estimates as current as possible.

Measurabl's Data

The ML model is trained to find patterns between usages and input features (e.g., building characteristics and weather) by learning what usage is likely for any combination of features in Measurabl's dataset - a process called "training." Once trained, the model can be applied to new buildings to estimate usage for any historical month, based on features from that new building.

Measurabl uses building features (see Appendix D for further details) that are universally available across all use types and locations and, at the same time are helpful in explaining the variance in usage patterns. The goal of the ML model is to accurately estimate the historical usage of buildings across a wide variety of use types, locations, sizes, etc.

The models rely on Measurabl's standardized, high-level use types (e.g. "Retail", "Residential", "Office", etc.), and their most granular subtypes (e.g. "Residential: Multifamily Housing","Residential: Senior Care Facility", etc.). The former allows the models to discern connections between the different subtypes that belong to the same general category, e.g., "Residential" in the example above. The latter means that whenever the most granular use type level is provided, the estimates will be based on averaging across this specific subtype. If only the highest use type level is provided, the average will be based on all buildings that share this less granular subtype.

Measurabl's time series data used to build the ML model consists of calendarized monthly usages for tens of thousands of spaces, spanning from 2017 onward. Those usages are sourced from actual bill/invoice data at the meter level, then calendarized and aggregated at the building level.

The data fed into the ML model for training purposes can be thought of as a table (Fig. 2), where each row consists of building ID, year and month as indices to each row, building features, highlighted in yellow, and monthly usage in MWh, for each month and year. All floor area values are converted to sq ft and all energy usage values are converted to MWh. The ML model's goal is to learn an algorithm that takes month, year, and all the features highlighted in yellow on the left-hand side as inputs, and outputs an estimated energy usage for building, month, and year.

Building ID	Year	Month	Use Type	Floor Area [sq ft]	Heating Degree Days	Cooling Degree Days	Latitude	Longitude	Country	Year Built	Monthly Usage [MWh]
1	2019	01	Office	7200	790	0	45.5	73.5	USA	1982	35.3
1	2019	02	Office	7200	740	0	45.5	73.5	CAN	1982	29.0
1	2019	03	Office	7200	320	0	45.5	73.5	CAN	1982	28.0
...	...	...	...	...	...	...	...	...	...	...	...
1	2020	12	Office	7200	353	1	45.5	73.5	CAN	1982	19.8
2	2019	01	Hotel	18000	220	2	51.5	0.1	UK	2011	87
2	2019	02	Hotel	18000	300	1	51.5	0.1	UK	2011	90
...	...	...	...	...	...	...	...	...	...	...	...
2	2020	12	Hotel	18000	190	0	51.5	0.1	UK	2011	85
3	2019	01	Mall	90000	34	102	19.4	99.1	Mexico	2000	140
3	2019	02	Mall	90000	33	99	19.4	99.1	Mexico	2000	145
...	...	...	...	...	...	...	...	...	...	...	...

Fig. 2: A representation of Measurabl's data used by the ML model

Measurabl's Machine Learning Model for Expected Energy Usage

Consider the relationship between energy usage and floor area across residential buildings (Fig. 3). While larger buildings use more energy on average, and a simple line based on a linear regression fit to the data will capture this trend, it will not be able to accurately explain the variation of usage for large floor areas.

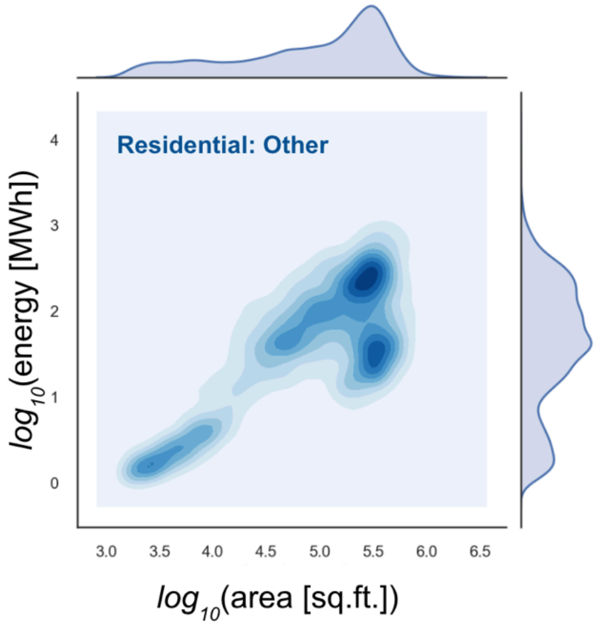

Fig. 3: Energy usage vs floor area for Residential properties in Measurabl (on a logarithmic scale)

This is where a more sophisticated non-linear model would be preferable. Decision tree-based ensemble models are better suited for modeling the complex relationships between building level characteristics and usage, due to fewer constraints about the statistical relationships between the building features, and thus, capturing more of the variance in the utility data.

Fig. 4 showcases a simplified example of a classification and regression tree used to determine the magnitude of the estimated usage based on multiple yes/no questions.

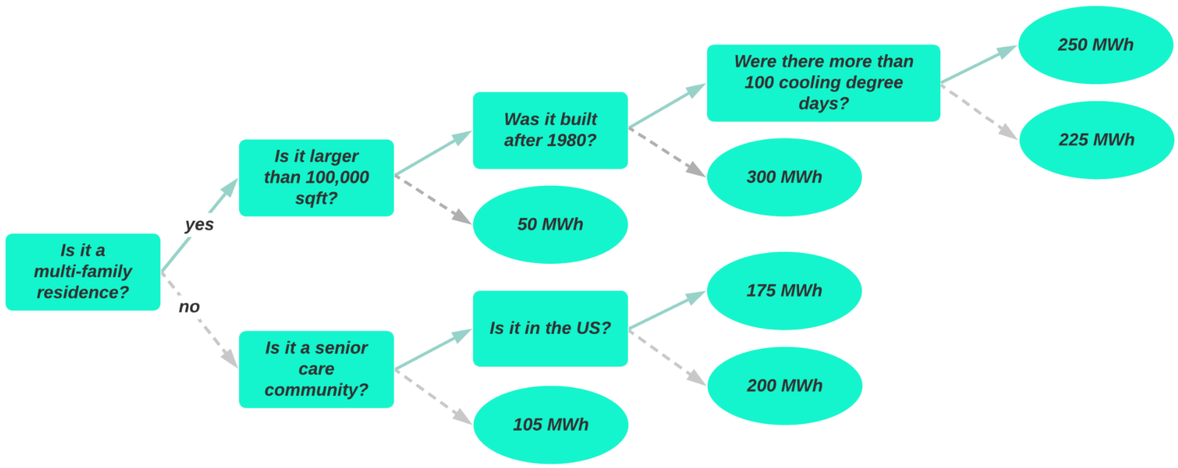

Fig. 4: An example of a classification and regression tree.

Model performance is measured on a set of buildings held out during model training (i.e., the test set) using standard metrics such as R^2. R^2 ("coefficient of determination") is an indicator of how much of the variance in the test set can be explained by the model. The closer R^2 is to 1 the better, where 1 means that 100% of the variance in the data is explained.

The model achieves a test set R^2 ~= 0.9. This means that the model is able to capture ~90% of the variance of the values it is attempting to predict. Thus, the energy estimates for buildings the model was not trained on should exhibit a high degree of accuracy when compared to actual consumption values, on average. Additional information regarding model performance validation, as well as a breakdown of performance metrics by building use type category can be found in Appendix E.

Additionally, a confidence label is returned alongside each estimate provided, which supplies further context into an estimate's expected accuracy (see Appendix F for further details).

As a final note, the ML model is optimized to produce estimates for all buildings <=1 million sq ft. Due to this optimization, Measurabl currently cannot provide estimates for buildings >1 million sq ft.

Measurabl's Methodology for Expected Carbon Emissions

Measurabl calculates total carbon emissions for each building by converting on-site energy usage to carbon emissions equivalent measured in MTCO2e, using local carbon factors. For more information on the local factors, please see Measurabl's GHG Methodology.

The carbon emission factors for natural gas are significantly (at times 50%) lower than the emission factors for electricity use in most regions. It's therefore important for the carbon estimates to account for a building's expected energy mix fractions for electricity and gas based on the best available data Measurabl has.

Thus, a building's expected energy mix fraction is estimated using the average energy mix fraction of similar buildings (those that have the same use type and share geographical proximity) found in Measurabl.

The estimated carbon emissions for each building are calculated using the following formula:

*Estimated Carbon Emissions = Estimated Energy **
*(electric fraction * carbon factor for electricity + gas fraction* carbon factor for gas)*

The applied emission factors can be found on pages 17-28 of Measurabl's GHG Methodology.

Appendix A - Machine Learning Model Data

The preprocessed data used to train the ML model for expected energy usage contains over 2,403,000 monthly data points from over 91,300unique buildings leveraging Measurabl's Quantum Database. These buildings come from 73 countries spanning the globe.

In order to enhance the accuracy of our model's estimations, we additionally leverage energy data from the California State Government's Building Energy Benchmarking Program and New York City Government's Local Law 84 Benchmarking Program.

Appendix B - Data Quality Preprocessing

Measurabl applies a very strict data cleaning process to the underlying dataset to ensure maximum integrity and data quality of the data used for deriving Whole Building Estimates. The main steps of this process include:

- Space-level data is aggregated at the building level, and only spaces with available, actual, utility data are included. This ensures that buildings' energy use intensity (EUI) is not underestimated due to spaces with missing data. Further, we exclude both buildings with data only from exterior meters and/or common space meters, and buildings that don't have electric meters in Measuabl's database.

- Outliers in the utility data can skew the models. Sometimes usage is entered in the wrong units, like in MWh instead of KWh, or the other way around. Other times, the floor area can be wrong, which may lead to unrealistic usage intensities. To avoid this issue, we remove invalid data as well as abnormally large and small usage intensities from the modeling data (i.e., approx. the top and bottom 15% for each building use type) after performing data inspections and reviewing usage intensity distributions.

- Use types that are not sufficiently represented (i.e., less than 20 buildings per Measurabl's most granular subtypes) within Measurabl's database are not included.

One of the key strengths of Measurabl's models for Whole Building Estimates is the freshness of the training data. Utility data flows into Measurabl's database on a continuous basis, thus allowing us to update the modeling dataset on a monthly basis. After the end of each calendar month, the model is retrained to include any data that was added in the previous month. Utility data older than 2017 is discarded in order to keep estimates as current as possible.

Appendix C - Energy Use Types

Energy, fuel and district are utility types defined by the Global Real Estate Sustainability Benchmark (GRESB) (see GRESB Methodology Documentation for more information).

Appendix D - Building Features

The building features utilized by the ML model are derived based on the following user inputs:

1. Building Location*
2. Primary Property Type*
3. Building Size (i.e., gross floor area) (sq ft or m^2)**
4. Year Built**
5. Time Period of Estimation (from 2017 to most recent complete month)*

**required input by user*

***optional input by user*

For best results, it is recommended that the user supply all inputs using actual data. However, in cases where the user cannot supply “Year Built” or “Building Size”, the ML model is still capable of producing estimations. When “Year Built” is not supplied, the ML model relies on an internal algorithm, created using training examples with missing “Year Built” values, to produce an estimate. When “Building Size” is not supplied, Measurabl’s Floor Area Estimates ML model is leveraged to supply an estimated gross floor area value for the user.

Additionally, Measurabl provides estimates for buildings assuming 100% building occupancy. In the future Measurabl may offer building occupancy percentage as an optional user input, which would be taken into account when making an estimate.

Appendix E - Machine Learning Model Performance

Performance Validation Framework

In order to validate model performance, Measurabl performs 5-fold cross validation. For each fold, the dataset is split into two parts: a training dataset and a test dataset. The training dataset is used to train the ML model, while the test dataset is used exclusively to validate model performance. In this way, performance results on test dataset estimations can be extrapolated to estimations for other buildings that are also unseen by the ML model.

While performing cross validation, Measurabl uses ~80% of the modeling dataset to create training data and ~20% to create test data. Note: Measurabl ensures that data in test and training sets come from non-overlapping sets of buildings.

ML Model Performance

Table 1 below showcases standardized performance metrics across several common building use type categories:

ML Model Performance Metrics (As of August 2025)				
	MWh			Properties
Use Type Category	R^2	MAPE	MdAPE	Count
All Use Types	0.87	26.50%	21.48%	91,366
Data Center	0.53	24.90%	21.09%	110
Education	0.89	21.59%	18.54%	1,890
Food Sales & Service	0.83	17.33%	13.93%	1,659
Healthcare	0.93	24.38%	21.64%	2,023
Hotel	0.90	18.77%	16.72%	1,597
Laboratory	0.66	37.75%	31.35%	472
Leisure	0.85	25.04%	19.83%	737
Manufacturing / Industrial Plant	0.55	52.47%	44.76%	1,320
Office	0.87	24.30%	21.80%	12,812
Public Services	0.87	23.74%	20.32%	520
Residential	0.87	25.12%	20.38%	46,223
Retail	0.83	28.16%	22.28%	10,813

Warehouse / Storage	0.71	40.09%	34.39%	8,782

Table 1: MWh = megawatt hours. R^2 = R-squared or coefficient of determination. MAPE = Mean Absolute Percent Error. MdAPE = Median Absolute Percent Error. All metrics computed using test data.

Basic Approach Comparison

To provide additional context for the ML model performance metrics (see Table 1), Measurabl implemented a Basic Approach for estimating building energy usage by leveraging building square footage and publicly available energy usage intensity (EUI) values. Specifically, the Basic Approach uses the following formula:

*Building Energy Usage = Building Sq Ft * Median EUI for Property Type*

Where:

- *Median EUI for Property Type: median EUI (MWh / sq ft) for the building's granular property type (based on: Energy Star Technical Reference: U.S. EUI by Property Type as of 04-2023)*

The Basic Approach represents a reasonable method for estimating a building's energy usage without Measurabl's Whole Building Estimates product. Table 2 showcases the ML model's percentage improvement over the Basic Approach for performance metrics in both log (i.e,. $\log_{10}$(MWh) and linear (i.e., MWh) space across several common building use type categories:

ML Model Percent Improvement Over Basic Approach (As of August 2025)			
	MWh		
Use Type Category	R^2	MAPE	MdAPE
All Use Types	6.10%	31.77%	31.15%
Data Center	N/A	N/A	N/A
Education	4.71%	44.36%	40.37%
Food Sales & Service	130.56%	33.50%	16.25%
Healthcare	-1.06%	-6.74%	-10.24%
Hotel	0.0%	9.28%	10.06%
Laboratory	34.69%	-1.53%	19.8%
Leisure	84.78%	11.08%	17.68%
Manufacturing / Industrial Plant	N/A	N/A	N/A
Office	3.57%	1.46%	6.20%
Public Services	64.15%	42.67%	47.34%
Residential	20.83%	41.31%	46.82%
Retail	13.70%	20.34%	16.33%

Warehouse / Storage	1.43%	26.4%	24.25%

Table 2: MWh = megawatt hours. R2 = R-squared or coefficient of determination. MAPE = Mean Absolute Percent Error. MdAPE = Median Absolute Percent Error. All metrics computed using the cross validation test data. Negative values indicate better performance by the Basic Approach. N/A values occur due to inability to obtain use type category EUI values from Energy Star Technical Reference: U.S. EUI by Property Type.

Appendix F - Confidence Labels for Machine Learning Model Estimates

Confidence labels and corresponding descriptions are returned alongside each estimate provided, which supplies further context into an estimate's expected accuracy.

Confidence Labels

Confidence labels can take on the following values:

1. High
2. Moderate
3. Low
4. Very Low

This label indicates how confident the ML model is at producing estimates generally for buildings with the property type of the building being estimated.

Confidence Label Descriptions

Confidence descriptions primarily take on the following forms:

1. For Energy Estimates: The average energy estimate error for properties in this category is anticipated to be <= {error percentage}%

2. For Carbon Estimates: The carbon estimate is generated from an energy estimate. The average energy estimate error for properties in this category is anticipated to be <= {error percentage}%.

The description is designed to convey that the ML model is 95% confident in asserting that the average error for buildings with the property type of the building being estimated is below the "error percentage" provided.

Generating Confidence Information

Confidence labels and descriptions are generated as follows:

1. Using the ML model and 5-fold cross validation, an estimate is made for each monthly energy usage value in the modeling dataset.
2. Monthly estimates from unique buildings are grouped, and the mean absolute percent error (MAPE) is computed for each building (Building-MAPE).
3. Building-MAPEs are then grouped based on granular property type information.
4. Within each property type group, the MAPE and standard error of the MAPE (MAPE-standard-error) are computed.
5. Again within each property type group, the MAPE's 95% confidence interval upper bound (MAPE-upper-bound) is computed using the formula:
    - MAPE-upper-bound = MAPE + 1.96 * MAPE-standard-error
6. The Confidence Description's "error percentage" is derived from the formula:
    - Confidence Description's "error percentage" = 1 - MAPE-upper-bound
7. Finally, Confidence Labels are assigned to each property type group by binning based on the following table:

Confidence Label	Confidence Description (Error %)
High	<=25%
Moderate	>25% – <=50%
Low	>50% – <=75%
Very Low	>75%