Properate: Property Rating Using A Novel Instant Energy Modeling Approach

Arman Mottaghi, Alex Lavoie, Artur Akhmetgareev, Noura Seifelnasr

Abstract

For mass-scale home energy upgrades, we present 4 criteria that no existing Building Energy Modeling tool provides simultaneously – Reproducibility, Expandability, Auditability, and Lightness. Reviewing the limitations of the common "White Box" (e.g. HOT2000, EnergyPlus) and "Black Box" (e.g. Machine Learning, AI) energy modeling tools, our novel "Grey Box" energy modeler, RBEST, excels in delivering on all 4 criteria. We then showcase RBEST within the software suite, Properate, to demonstrate how it enables Virtual and Remote Energy Rating when conventional On-Site Rating may not be feasible. 6,917 buildings in the Climate Zones 4 through 7 of North America are assessed with Properate, then their accuracy is compared to On-Site Rating. The results show that while Properate introduces new flexibility and scalability in the process, it maintains an average weighted accuracy of 92% compared to On-Site Rating. The results can pave the path for community-scale home deep energy retrofits across Canada.

Background

With increasing declarations of climate emergency across the globe, communities and governments are looking for ways to understand their building stock and plan energy upgrades for them. Out of all building types, low-rise residential buildings have unique challenges in this regard; each building is unique and too small for programs that can operate at-scale. These challenges need to be addressed since the cumulative environmental impact of these buildings has been the largest amongst any building type [1][2].

The challenges are present from the "Energy Rating" (or "Energy Audits") stage — the very first step of a building energy upgrade. As such, many governments are looking into ways of scaling Energy Rating for homes [3].

An Energy Rating has been historically done by a visit to the building. The "On-Site Rating" can create a comprehensive image of one building; albeit, collecting accurate information from a home is time-consuming.

This creates a tradeoff: if an *Energy Rater* increases the precision of the data they collect from each home, they will end up assessing fewer homes in a given timeframe. Since time is the major factor in Energy Rating costs, each home ends up paying more as well.

The Energy Performance Certificates (EPCs) in the EU are an example of this tradeoff. Each EU country can set their own EPC procedure. Depending on the economical and environmental dynamics in each country, the method of EPC creation can widely vary. Most countries require a visit to the home for issuing the EPC, but not all. The data collection fidelity and the tools for processing the EPC is also up to the individual member states, creating wide variance in the costs [4].

In Canada, the Rating is done under the *EnerGuide For Homes* program. The data collection under this program requires On-Site data collection. The calculation for the Rating is typically done using the Canadian energy modeling software HOT2000.

This comprehensive process is one of the strengths of the EnerGuide program. It means that any rating generated by the program is backed by an energy simulation, making it auditable.

Amid substantial industry scale-up efforts, this program has managed to have approximately 65,000 EnerGuide Ratings in 2021, which is only 1% of the applicable build-stock in Canada. We believe that for mass-scale home energy upgrades, every home needs to have an Energy Rating that continuously gets updated.

For this purpose, we present improvements for both Energy Rating approaches and technology.

Table 1: Summary of different Energy Rating Approaches

	Rating Types		
	Virtual	**Remote**	**On-Site**
Time To Results	Instant	5 minutes	3-6 hours
Homeowner Involvement	**No**	Yes	Yes
Require Home Access	**No**	**No**	Yes
Purpose	Targeting	Qualification	Verification

Energy Rating Approaches

Relying only on On-Site rating makes it extremely difficult to scale-up Energy Ratings. Prior to the site visit, the Energy Rater does not have usable data from the home, so the economics and circumstances define which buildings get On-Site assessments. Added to the equity considerations, such an approach is inefficient.

To resolve the above limitations, our process adds two other Rating options:

1. **Virtual Rating** (or "Virtual Audits") does not require engagement of homeowners. It can be performed using available public data at the scale of neighbourhoods. The results of a Virtual Rating (VR) may not have the highest accuracy; however, such results may be used to direct efforts towards the homes that require the most attention.

2. **Remote Rating** (or "Remote Audits") requires the engagement of the homeowner. Remote Ratings (RR) can achieve higher accuracy than a Virtual Rating, with significantly less barriers compared to an On-Site Rating. Remote Ratings can be done in a DIY (Do It Yourself) fashion or through a professional. The process may involve a questionnaire, or some pictures of the home.

Aside from the above tooling needs, the connection between different Rating approaches is presently lacking in the industry. Remote Ratings can be done faster and more effectively with data from a Virtual Rating.

In the same way, significant speed-up is possible if On-Site Rating is done on the basis of the data available through Virtual/Remote Rating, eliminating many variables.

In the industry, sometimes the scope of "Rating" (and "Labeling") is considered only for generating a simplistic overview (e.g. an A-E scale of efficiency). Meanwhile, the community-level insights are extracted through "building stock modeling"; Creating a disconnected system where top-down and bottom-up insights do not flow.

With connected Rating processes, a community may commission a Virtual Rating of all its homes and learn that homes built within specific years have the highest retrofit potential. The community may then target those homes with a Remote Rating program, through which homeowners can get qualified for incentives. Finally, an On-Site assessment, verifies the potential and quantifies the retrofit benefits.

For such Rating processes to be as successful as an On-Site assessment, the methodology and the scope of the Virtual and Remote Rating tools must be clear. For this reason, we believe that the On-Site level of comprehensiveness must apply to Remote and Virtual Rating.

Technological Requirements

To deliver Energy Rating, we have identified 4 key requirements that a Building Energy Modeling (BEM) software must deliver:

3. Reproducibility: One set of input conditions must always output the same results. This requirement is challenging to meet for models that rely on random sampling.

4. Expandability: The energy modeling tool must be able to handle "unseen" buildings. This is particularly important for low-income housing where not many buildings have been previously assessed due to the cost of On-Site Ratings. If past On-Site Rating data is being used as the basis for energy modeling, it may not represent low-income households well.

2. Auditability: The model outputs must be validatable at the component-level using a well-studied & common energy modeling tool, conforming to a standard such as ASHRAE 140. This requirement not only creates transparency in validating a result, but also can accelerate further On-Site Rating.

1. Lightness: The energy simulation workflow must be computationally light enough to be used for millions of homes. Constant introduction of new technologies and data means that comprehensive *digital twinning* of a community requires these simulations to run periodically on generally accessible computer hardware.

The above 4 constitute the *REAL requirements*. In the next section we provide examples of how these requirements may be reviewed.

Technology review

In many fields of science, simulation methods are classified within a "White Box" to "Black Box" spectrum [5][6]; the differentiating factor being the ways the simulation methods leverage physics & statistics. In the field of Building Science the same classification is evident.

White Box Modeling (Physics-Based)

The current industry tooling for building energy modeling heavily favours "White Box" building energy models.

The name, White Box, suggests that the modeling approach is reliant on mathematical equations, i.e. an observer can inspect the model and understand how it works.

The transparency however, comes at a computational cost. Since all building components are interconnected, there needs to be iterative computation to successfully simulate a building. *Figure 1* displays a conceptual diagram of such models.

EnergyPlus [7] is one of such White Box models. One component of EnergyPlus load balancing calculations is air flow. The air flow between different building zones (analogous to *m* in *figure 1*) are calculated iteratively "for each zone in an air loop until the convergence criterion is satisfied".

This convergence through iteration is done for each timestep (*n* in *figure 1*) of the simulation. A moderate number of timesteps may be over 50,000 in one year (every 10 minutes).

Such nested iterations (zones within timestep) create rising computation Time Complexity.

Multiple nested iterations may be present in one model for different purposes, adding to the computation required to reach convergence.

While the computational burden may not be significant for one building simulation, it becomes a challenge for upgrade planning and Remote Rating at the community level.

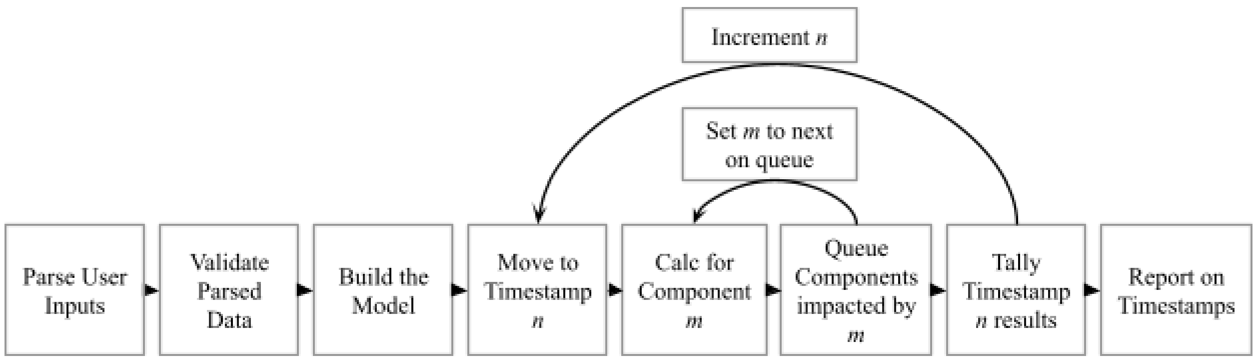

Figure 1: White Box Modeling

On our benchmarked case using HOT2000, a 4th generation Intel processor took 7 seconds to fully simulate a home. For a community of 50,000 homes, a single simulation run on such a computer will take 4 days to complete. Many rounds of simulation as well as iterative analysis of potential upgrades are necessary for a comprehensive community study, increasing the demand to millions of simulations which will require significant computation time. In short, the White Box models lack *Lightness*.

To resolve the computation limitation of White-Box models without changing the technology, researchers have done either of the following:

1. Scale the computation: In one example, Argonne National Laboratory in the US used a supercomputer to perform Energy Rating for 178,000 buildings [8]. Same can be done with on-demand cloud computing instances, which have become increasingly available over the past decade.

However, even at that scale, Energy Rating for an entire country would not be possible at a reasonable cost. The results may also not be instantly available.

2. Simplify the use-case: When computation is restricted, commonly, communities opt to make "archetypes" of the building stock, and study the archetype. Each archetype is meant to represent one common type of housing. By only performing analysis on the archetype, the researchers hope to make discoveries that are expandable to all the buildings the archetype is meant to represent.

Defining the archetypes may be considered part art part science by the industry professionals. The number and the features of the archetypes are defined according to the statistical analysis of building features, the project scope/budget, and the purpose of the analysis.

The issue with archetype studies is that they are "one size fits all" by definition, thus not suitable for Remote Rating.

3. **Relate to previously computed results**: When there is a non-exhaustive number of ratings available from the previous two methods, researchers may try to use the available results to guess the rating of unrated buildings. One way to do so is by finding the "nearest neighbour" of an unrated building. If there is a rated building that's sufficiently similar to the unrated building, the rating may be reusable for the unrated building. Examples of such an approach can be found in [9][10].

Inferring the relationship between the inputs and the outputs using statistics is a more dynamic way of *learning* from the rated building data. [11] calls these "surrogate models" (or "meta-models") in general and provides an overview. These models can vary greatly in their complexity. In general, these models tune themselves using an available set of "training data".

Particular to the residential Energy Rating, Surrogate models share some limitations with the White Box models; Surrogate models must become more complex as the fidelity increases. They face the so-called "curse of dimensions". In short, if they want to include more inputs from a building, they must have access to exponentially more training data.

Since our focus is on full-featured building energy models, below we discuss the comprehensive types of these Surrogates, the Black-Box models.

Black-Box Modeling (Statistics-Based)

One alternative to directly using the "White Box" models, are the "Black Box" models which use statistics. The advancement of data science and machine learning has prompted many research and development teams to attempt and bring such new technologies within the building science domain [12].

The results are considerable simplification of the building energy modeling architecture, boosting *Lightness*. The architecture simplification is apparent in *figure 2*

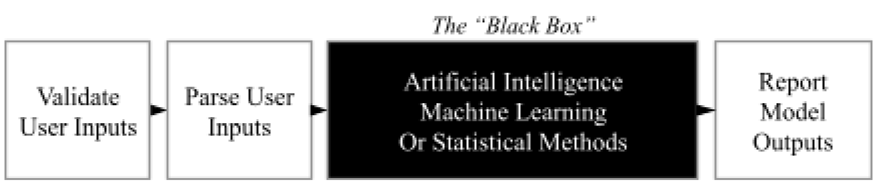

Figure 2: Black Box Modeling

This simplification, however, may come at a cost. While Black-Box models can be powerful tools for research, they often lack *Auditability*, *Reproducibility*, & *Expandability* for Remote Rating.

To study these *REAL requirements* we provide a simplified example. Let us assume a researcher decides on using polynomial regression to infer the EnerGuide score of a home based on the home's heated area. A regression model is a Black Box modeling approach.

The most apparent issue with this analysis as the name Black Box suggests is the opaqueness of how the model finds its results . Let us assume that the model is $0.035x^2 + 0.2x - 6.3$. Such a formula does not provide any way of reasoning about the results; therefore, it lacks *Auditability*. Actual Black Box models often employ complex neural networks and other Machine Learning tools which further restrict what can be learned from them.

Moreover, as the rudimentary *figure 3* example shows, the model also fails *Expandability*. Adding one new data point *D* between *B* & *C* changes the best regression of the graph, changing the outputs of the model.

Any Floor Area query bigger than C, despite not having any new evidence, has drastically changed. This “best fit” nature of the Black Box models makes them rather unpredictable.

We acknowledge that the provided case may be argued away as an example of overfitting on a small data sample; however, the example is valid as a simplified case of a critical issue with the Black Box models. In reality, the Black Box energy models have far more dimensions than the two in the example. As such, the actual models deal with sparse data across those dimensions where new data can substantially change the model’s behaviour.

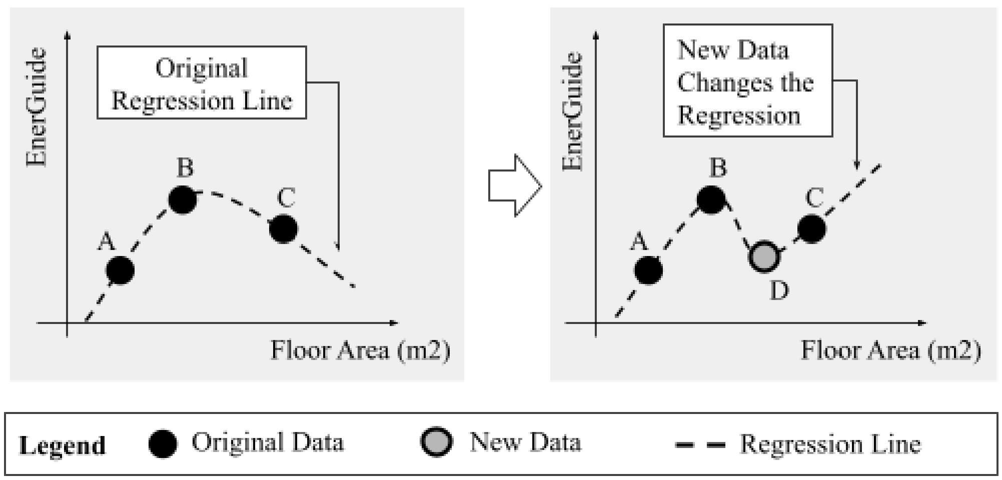

Figure 3: Rudimentary example of how new data impacts Black Box models.

Some Black Box models also fail *Reproducibility*. Examples are Monte Carlo based simulations [13]; or in general, any model that has random sampling or noise as a part of its activation.

Finally, the reliance of the Black Box models on past data means they are not flexible to changes in data collection or simulation methodology.

Let’s consider the example of EnerGuide. Over the recent years updates to the HOT2000 software were issued multiple times, often with changes that impacted simulation results. In one case, the climate data that powers the simulations was overhauled. Such changes can reduce the usability of the older versions’ data for a Black Box model.

This limitation also makes Black Box models unsuitable for Resilience analysis, where the climate of the future is of concern.

Grey-Box Modeling (Hybrid)

Grey-Box models are commonly referred to as the hybrid of the above approaches. The fashion in which the hybrid model is made can vary.

We present our version of a Grey-Box model, the Rapid Building Energy Simulation Tool (RBEST).

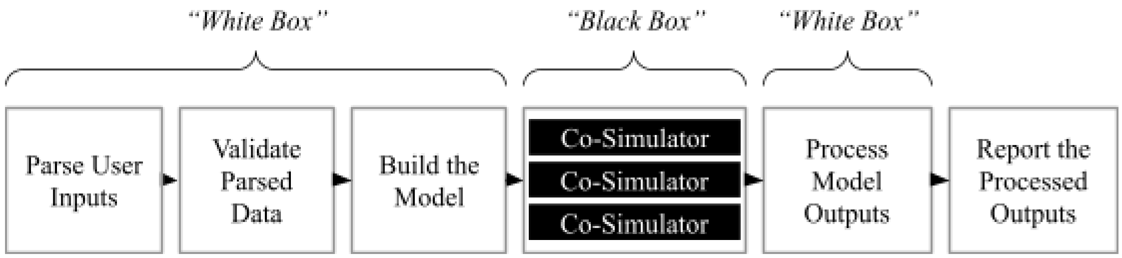

Figure 4: Properate's RBEST (Grey Box) Model Architecture

The design philosophy behind RBEST is mixing the advantages of the White Box modeling with the acceleration made possible by Black Box modeling. Hence, RBEST starts the process by making an energy model in a very similar fashion to a White Box model. RBEST parses the provided building components and prepares them for a simulation environment.

From that point, RBEST employs Black Box modeling to avoid the need for the computationally heavy parts of a White Box approach.

The goal of the Black Box module is to only produce the parameters that a White Box model would have computed through iterative calculations. That purpose significantly simplifies the Black Box model requirements. It also makes it trivial to detect errors in the Black Box model outputs.

Table 2: Feature summary of different Energy Rating Technologies

Category	**White Box**	**Grey Box**	**Black Box**
REAL Requirements			
Reproducibility	Yes	Yes	Common
Expandability	Yes	Yes	No
Auditability	Yes	Yes	No
Lightness	No	Yes	Yes
Other Desirable Features			
Component-Level Outputs	Yes	Yes	Uncommon
Conversion between different models	To Grey & Black Box	To White Box & Black Box	No
Co-Simulation: Processing for multiple simulation paradigms, in one shot	No	Yes	Yes
Resilience Analysis: working with novel climate models and forecasted climate data	Yes	Yes	No

Another advantage of the Black Box module are Co-Simulators, which offer parallelization. We often find the need to employ multiple residential building energy simulation methodologies in one project; however, most methodologies are incompatible with each other.

For example, a model used for the energy rating of a building can be different from the model used for sizing the mechanical equipment. In other cases, homes that want to conform to the Passive House standard, also like to participate in EnerGuide incentive programs. Hence, needing two types of simulation.

Instead of making two separate models, RBEST uses single-shot multivariate simulation to give multiple results at once.

Once the Black Box model returns the computation result of one or multiple simulations, the results are fed back into a module resembling that of a White Box model. This final module converts and summarizes the inference results from the Black Box into building component-level simulation results. Finally, the module produces an energy label.

As *Table 1* shows, this Grey Box modeling passes all the *REAL Requirements*. RBEST provides the *Expandability* and *Auditability* of a White Box model at much faster speed.

On the benchmark computer introduced earlier, RBEST ran 6 orders of magnitude (1,000,000 times) faster than the White Box model. The benchmarking only involved the operation of RBEST. In the production environment, other components and the data transfer latency significantly add to the overall turnaround time.

The most challenging part of developing and validating RBEST was at the "seam" between the modules — at the transition between White Box and Black Box. The complex nature of the module communication made debugging the issues challenging.

Another byproduct of the RBEST design, is validation after input data parsing. This design choice was made to make the Black Box module more robust; however, it means that when anomalies arise in the validation step, they may not be understandable for the users. That is because RBEST finds the issues in what it has parsed, not in the user inputs.

Our resolution to this limitation is an error reporting subroutine which takes an RBEST error and translates it into something a user can understand and fix.

Properate

Properate is a property rating software suite which orchestrates user input collection, supplementary data compiling, building energy simulation, and retrofit planning.

Currently, Properate suite has the following functionality:

1. Maps: Shows the Energy Rating of every home in a neighborhood. If Remote Rating or On-Site Rating data is not available for a home, Virtual Rating results are displayed for that home.

2. Wizard: A questionnaire interface for Remote Rating. It can be operated by homeowners in a DIY fashion, or by professionals in a concierge fashion.

3. Studio: Made for professionals for On-Site or Remote Rating. Studio's interface allows for complete data collection from a building and the results provided have the highest fidelity.

4. API: Programmatic interfacing with the RBEST and estimators such as costing and GHG for third-parties.

RBEST powers the core instant building energy simulation functionality of Properate, while other technologies prepare the inputs and outputs. The following section provides a use-case of how Properate orchestrates all these technologies in the context of community-level energy labeling.

Case-Study

We put Properate to test for Remote Rating. White-Box modeling based on an On-Site Energy Rating is used as the "ground-truth" for the testing.

Methodology

We employed a modified blind-study methodology in which we gave Properate the following data points from 6,917 buildings. These buildings were also assessed by an On-Site EnerGuide evaluation. The On-Site evaluation results and the data points below were never examined by Properate prior to the case-study:

1. **Building Address**: Properate can use the address to extract properties such as weather station data, altitude, & surroundings (VR)
2. **Satellite Imagery** of the building location (VR)
3. **Property Records**: Building Square Footage, year built (VR)
4. **Building Massing**: number of stories and area per floor (VR and RR)
5. **Fuel types** of heating, and hot water systems (VR and RR)
6. **Upgrades** to the building over the past decade (VR and RR)

The goal of the study is to see how close Properate can simulate these unseen buildings when compared to an On-Site Rating.

We have noted the usage of each data point for RR (Remote Rating) and VR (Virtual Rating) in the above list. That's because not all data was available for every building. This provided an opportunity to study the effect of missing data on Properate's performance.

Data

The studied buildings are Detached and Attached Single-Family Dwellings (SFDs) located in the provinces of British Columbia and Alberta, in Canada. The defining characteristics of the data are as follows:

- **Age:** The homes were built between 1895 and 2022, encompassing roughly 130 years of construction. The data is heavily skewed towards homes built after 2011.
- **Location:** Most homes are located in Edmonton, the Lower Mainland, and Southern Vancouver Island. With other buildings spread out around BC. The data covers Climate Zones 4 to 7.
- **Definition**: Most buildings had at least one missing data point. That left 1351 homes which could have a Remote Rating done, mostly in Metro Vancouver. The rest could only be studied with their Virtual Rating.
- **Size**: The livable area of the homes ranges from approximately 9 m^2 (95 ft^2) to 1,300 m^2 (14,000 ft^2). Therefore, the data covers a wide range of residential building types from tiny homes to mansions. *Figure 5* plots a histogram.

Simulation

Before simulating all buildings, we studied a random sample of 100 buildings from the dataset, looking for any issues that may adversely affect the accuracy. After flagging and addressing the issues, we performed a Virtual Rating of all buildings.

The subset of buildings that had all the necessary data points were simulated once again, this time with a Remote Rating, to measure the improvements.

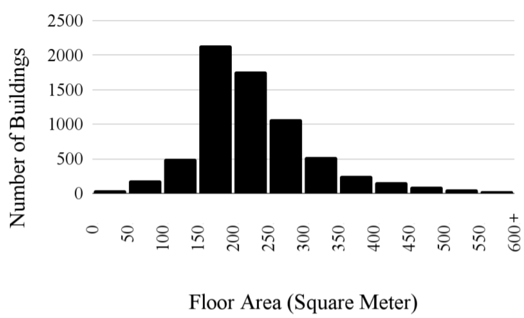

Figure 5: Histogram of Home Floor Areas

Error Definition

We measured the accuracy by comparing the EnerGuide ratings found by On-Site Rating and by Properate's remote Rating. The reason we selected this metric is its importance in Canadian Energy Rating. The EnerGuide score is the compound result of component-level computation, which means it can represent the overall energy model well. Achieving high accuracy in EnerGuide score estimates makes it easy to also accurately calculate other energy model outputs, such as GHG emissions.

Utilizing ASHRAE 14 simulation calibration accuracy measurement processes, we evaluated the result errors at 3 levels of complexity:

1. Mean Bias Error (MBE): Measures overlap of the On-Site Rating and Properate's remote Rating.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})$$

where:

n is the number of homes
y_i is Properate's estimated EnerGuide rating (GJ) for simulation i
$\widehat{y_i}$ is On-Site Rating EnerGuide score (GJ) for simulation i

2. Mean Absolute Error (MAE): In MBE, prediction overestimates and underestimates cancel each other out. MAE measures the absolute difference between the Properate's remote Rating versus On-Site:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y_i}\right|$$

3. Normalized Root Mean Square Error (NRMSE)[1]: In practice, large errors matter much more than small errors. This metric penalizes larger errors in an absolute fashion by squaring each error in Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}$$

To put RMSE in an understandable context, we then normalize it to get a percentage:

$$NRMSE = \frac{RMSE}{\bar{y}}$$

where:

$\bar{y}$ is the range of On-Site EnerGuide Rating.

In the same fashion, the previous error metrics, like MBE can be normalized as NMBE.

Accuracy Definition

As for how to define the "accuracy" of

---

[1] Our NRMSE is analogous to ASHRAE 14's CV(RMSE)

Properate, we simply use the normalized errors. We define the “Absolute Accuracy (AA)” as:

$$AA(\%) \;=\; 100\% - NMBE$$

and the “Weighted Accuracy (WA)”, which is meant to penalize large errors, as:

$$WA(\%) \;=\; 100\% - NRMSE$$

Analysis

The results of the simulations can be seen in *Table 2*, according to the error calculation methods introduced.

Table 2: Properate Accuracy Metrics

	VR	**RR**
n (# of homes)	6,917	1,351
$\bar{y}$ (rate range)	691 GJ	312 GJ
Mean Values		
MBE	31 GJ	1 GJ
MAE	44 GJ	16 GJ
RMSE	61	30
Normalized Values		
NMBE	4.5%	0%
NMAE	6%	4.5%
NRMSE	9%	8%
AA	94%	95.5%
WA	91%	92%

Exploring the three error measurement methods, it is clear that Properate has managed to estimate the EnerGuide score of most buildings with high accuracy.

The MBE of Virtual Rating (VR) is 31. The number being positive shows that Properate is generally overestimating the homes’ energy score. While the overlap of Properate and On-Site Rating is high, there is a gap of 30GJ.

Remote Rating practically eliminates the MBE. Properate’s predictions and the On-Site Rating have the perfect data overlap.

That being said, reviewing the other error measurements, Virtual Rating managed to maintain an accuracy of over 90% for both AA and WA despite its limited data points.

It must be noted that in this case, all 6 data points for each building given to Properate (as outlined in the Methodology section) were collected for Rating. This is the ideal process for Virtual Rating data collection.

In many other circumstances, Virtual Rating data may come from local government planning departments, tax authorities, or real estate databases. Then the data needs to be repurposed for Remote Rating, potentially impacting Virtual Rating’s accuracy. In that scenario, Remote Rating data collection gains another important purpose. Some of the data that Properate collects for Remote Rating is used to check the sensibility of the Virtual Rating data.

With Remote Rating, Properate’s accuracy increases further above 90%. The MAE of Properate’s Remote Rating is 16 GJ, meaning that on average a 16 GJ difference is expected between Properate and an On-Site Rating. This is a significant improvement from the Virtual Rating’s 44GJ. Looking at the normalized values however, the difference is not as pronounced. That is because, as mentioned, much of the Remote Rating data is in Metro Vancouver. Overall, the Remote Rated 1,351 homes have a narrower EnerGuide Score range.

Figure 6, provides a more granular level of error comparison, making the benefits of Remote Rating more understandable.

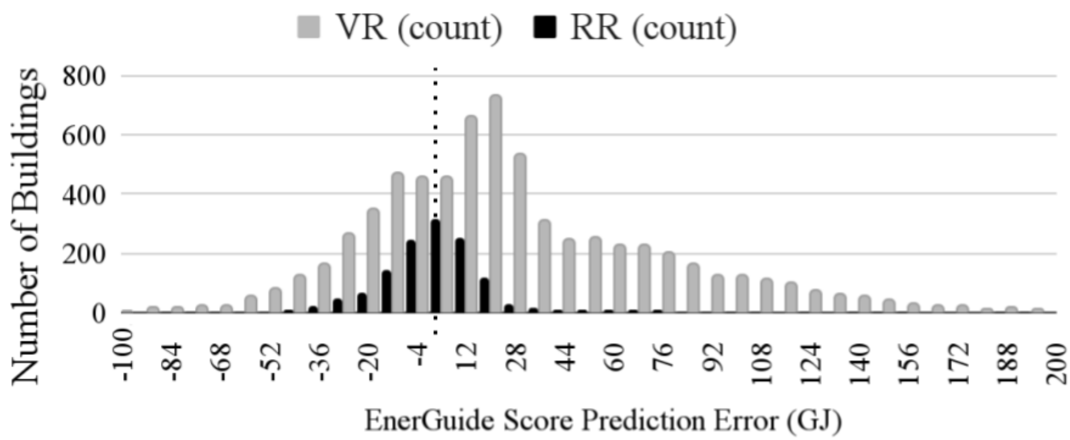

Figure 6: histogram of Properate's estimation accuracy for Virtual Rating (VR) & Remote Rating (RR)

Remote Rating's error distribution is more similar to a normal distribution, with a single peak that is aligned with the 0 error point (the dotted line) on the graph. Virtual Rating's error distribution is not as defined nor is it aligned with the 0 error point. MBE had already shown this difference between Remote Rating and Virtual Rating.

It also can be seen in *Figure 6* that there is, at times, a wide gap in Properate's estimation accuracy, especially for Virtual Rating. The larger errors are in part because of outliers.

Outliers

While the data collection process was for the purpose of Rating, there are still inaccuracies in the data. Some of such inaccuracies are evident when looking for outliers.

For example, 8 data points had a 3 digit number as their year built, e,g. *197* instead of *1974*. The year built errors were easy to identify and remedy; however, there are other inaccuracies that we could find only upon individual building review. The most common types of these inaccuracies were:

1. Rating being from the laneway home or part of a property, while Properate was given the main building address.

2. There have been re-development/renovations after some of the data points given to Properate.

3. The building areas were misidentified. A common example is misidentifying a first floor as a basement.

Data Skew

As we pointed out earlier, one of the data imbalances is a skew in the "year built" of the homes. To study the effects of data skew we sliced the data in such a way that it separates the overrepresented data points from the rest of the data.

Figure 7 compares the overrepresented slice of the data with the overall. In both cases, the error distribution is similar, except that the overrepresented data has a slightly smaller error.

The slight difference can be explained by highlighting that the overrepresentation comes from the homes being built recently. Newer homes are built under stricter and more elaborate building codes. Furthermore, there is less chance of interim modifications in the building, these factors have a positive effect on Properate's accuracy.

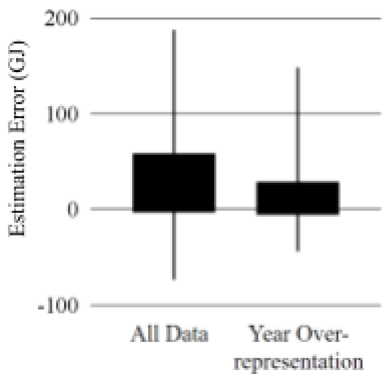

Figure 7: Error Rate Comparison Of Overrepresented Data with

With the difference being explainable, we don't find an adverse imprint from the skew in the results.

Representativeness

The year built and region of the data available to us were not homogenic. So the results may not be broadly representative of the built stock in the provinces of BC and AB. That said, establishing representativeness appears to be highly likely with more data from underrepresented regions and years.

Barring the age and region imbalances, using Confidence Interval statistical analysis shows that for the roughly 2 million homes in BC and AB, given the sample size of the data in this study there is a 1.55% margin of error with a confidence level of 99%.

Next Steps

To our knowledge, this is the largest study of its kind for Remote Rating. The learnings from this study also reveal the following areas for further research:

1. Acquire more Rating data to make the results broadly representative across Canada.

2. Study the building *digital twins* with other co-simulators beyond EnerGuide. Co-simulation can substantially facilitate access to more specialized energy modeling activities such as mechanical system sizing and climate adaptation.

3. Perform sensitivity analysis on inputs to understand how inaccurate inputs can lead to outliers in accuracy results.

Conclusion

We used Properate to remotely perform Energy Rating on 6,917 buildings in the Climate Zone 4 through 7 of North America then compared the results with On-Site Rating from those buildings.

For 1,351 of the buildings that had 6 generally-available data points, the results show an Absolute Accuracy of 95.5% and a Weighted Accuracy of 92%.

When the number of data points given to Properate were reduced to what was available for all 6,917 buildings, the result was Absolute Accuracy of 94% and a Weighted Accuracy 91% for all buildings. The decrease in accuracy being modest, shows the flexibility of Properate when dealing with missing data.

Given the limitations of the source data, the results may not be generalizable to all homes in the studied Climate Zones; however, our analysis of the data imbalances and features indicate high confidence in the statistical significance of the results for the studied building types.

References

[1] ACEEE, "Home Energy Efficiency Policies: Ratings, assessments, labels," ACEEE Policy Brief, 2018. Accessed: Oct. 01, 2022. [Online]. Available: https://www.aceee.org/sites/default/files/pdf/topic-home-energy-assessment.pdf

[2] Natural Resources Canada, “Survey of Household Energy Use,” Office of Energy Efficiency, 2015. Accessed: Oct. 01, 2022. [Online]. Available: https://oee.nrcan.gc.ca/corporate/statistics/neud/dpa/showTable.cfm?type=SH&sector=AAA&juris=CA&rn=4&page=2

[3] K. B. Baum, “Ottawa’s next push on ‘green homes’ is to require an energy audit before you can sell,” *The Globe and Mail*, Feb. 26, 2022. Accessed: Oct. 10, 2022. [Online]. Available: https://www.theglobeandmail.com/business/article-ottawas-next-push-on-green-homes-is-to-require-an-energy-audit-before/

[4] A. Arcipowska, F. Anagnostopoulos, F. Mariottini, and S. Kunkel, *Energy Performance Certificates across the EU*. the Buildings Performance Institute Europe (BPIE), 2014.

[5] J. Langevin *et al.*, “Developing a common approach for classifying building stock energy models,” *Renewable and Sustainable Energy Reviews*, vol. 133, p. 110276, Nov. 2020, doi: 10.1016/j.rser.2020.110276.

[6] S. A. Mottaghi and R. Mora, “Model-based coupling of air and hydronic system operation in a high performance academic building,” Dissertation/thesis, British Columbia Institute of Technology, 2017.

[7] U.S. Department of Energy, “EnergyPlusTM Version 22.2.0 Documentation: Engineering Reference,” Sep. 2022.

[8] “Argonne supercomputing resources power energy savings analysis,” *Argonne National Laboratory*. Accessed: Oct. 10, 2022. [Online]. Available: https://www.anl.gov/article/argonne-supercomputing-resources-power-energy-savings-analysis

[9] Institut Wohnen und Umwelt GmbH, *The joint EPISCOPE and TABULA Website*. https://episcope.eu/welcome/ (accessed Jan. 04, 2023).

[10] Clean Air Canada Partnership, “Cost Benefit Analysis Tool - NRCan, Lucas Coletta Training,” *YouTube*. Jun. 15, 2021. Accessed: Jan. 04, 2023. [Video]. Available: https://www.youtube.com/watch?v=fRvgQhzIA5U

[11] P. Westermann and R. Evins, “Surrogate modelling for sustainable building design – A review,” *Energy and Buildings*, vol. 198, pp. 170–186, Sep. 2019, doi: 10.1016/j.enbuild.2019.05.057.

[12] “ASHRAE - Great Energy Predictor III,” *Kaggle*. https://www.kaggle.com/competitions/ashrae-energy-prediction/data (accessed Oct. 10, 2022).

[13] N. Metropolis and S. Ulam, “The Monte Carlo Method,” *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949, doi: 10.2307/2280232.