Benchmarking building energy performance: Accuracy by involving occupants in collecting data - A case study in Germany

Wederhake, Lars; Wenninger, Simon; Wiethe, Christian; Fridgen, Gilbert; Stirnweiß, Dominic

Journal of Cleaner Production 379 (2022) 134762

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Benchmarking building energy performance: Accuracy by involving occupants in collecting data - A case study in Germany

Lars Wederhake [a], Simon Wenninger [b,c,*], Christian Wiethe [b], Gilbert Fridgen [d], Dominic Stirnweiß [e]

[a] *credium GmbH, Katharinengasse 13, 86150, Augsburg, Germany*
[b] *Branch Business & Information Systems Engineering of the Fraunhofer FIT, Alter Postweg 101, 86159, Augsburg, Germany*
[c] *FIM Research Center, University of Applied Sciences Augsburg, An der Hochschule 1, 86161, Augsburg, Germany*
[d] *SnT – Interdisciplinary Center for Security, Reliability and Trust, University of Luxembourg, 29 Avenue John F. Kennedy, 1855, Luxembourg*
[e] *FIM Research Center, University of Augsburg, Alter Postweg 101, 86159, Augsburg, Germany*

ARTICLE INFO

Handling Editor: Mingzhou Jin

Keywords:
Energy quantification methods
Energy performance certificate
Data-driven methods
Building data collection
Building energy performance
Energy efficiency

ABSTRACT

Energy performance certificates (EPC) aim to provide transparency about building energy performance (BEP) and benchmark buildings. Despite having qualified auditors examining buildings through on-site visits, BEP accuracy in EPCs is frequently criticized. Qualified auditors are often bound to engineering-based energy quantification methods. However, recent studies have revealed data-driven methods to be more accurate regarding benchmarking. Unlike engineering methods, data-driven methods can learn from data that non-experts might collect. This raises the question of whether data-driven methods allow for simplified data collection while still achieving the same accuracy as prescribed engineering-based methods. This study presents a method for selecting building variables, which even occupants can reliably collect and which at the same time contribute most to a data-driven method's predictive power. The method is tested and validated in a case study on a real-world data set containing 25,000 German single-family houses. Having all data collected by non-experts, results show that the data-driven method achieves about 35% higher accuracy than the currently used engineering method by qualified auditors. Our study proposes a stepwise method to design data-driven EPCs, outlines design recommendations, and derives policy implications.

1. Introduction

1.1. Motivation

As an immediate consequence of the United Nations' Paris Agreement on Climate Change, the European Union (EU) has set up objectives across all energy end-use sectors in 2020 and aims at a decrease of 32.5% in energy use below 1990 levels by 2030 to tackle human-made climate change (The European Parliament and the Council of the European Union, 2018). Buildings in the EU consume 40% of the overall final energy for heating and cooling (The European Parliament and the Council of the European Union, 2018). A large share stems from residential buildings. In Germany, which holds the EU's largest building stock by country, residential buildings consume 22% of the country's total energy (Ballarini et al., 2014; German Federal Ministry for Economic Affairs and Energy (BMWi), 2018). Therefore, the energy efficiency of the residential building stock is a critical success factor in reaching climate goals in Germany and many other countries (Balaras et al., 2016; Fan et al., 2014).

However, energy retrofits occur rarely, and energy saving rates are far too low to meet EU targets (German Energy Agency, 2018). Uncertainty about the building's energy performance (BEP) and the resulting savings in energy costs after a retrofit are significant barriers to implementing energy-saving retrofits (Casals, 2006; Walter et al., 2014). Policymakers, e.g., in the EU, introduced energy performance certificates (EPC) to reduce that uncertainty by providing information on the BEP (The European Parliament and the Council of the European Union, 2002; Poel et al., 2007). EPCs primarily set out to benchmark buildings and perform comparative studies in national and international contexts (Droutsa et al., 2016). For successful benchmarking of buildings, the accuracy and frequency of the creation and renewal of EPCs are essential.

In many countries, laws or ordinances define the energy

* Corresponding author. Branch Business & Information Systems Engineering of the Fraunhofer FIT, Alter Postweg 101, 86159, Augsburg, Germany
E-mail address: simon.wenninger@fim-rc.de (S. Wenninger).

https://doi.org/10.1016/j.jclepro.2022.134762
Received 8 July 2022; Received in revised form 13 October 2022; Accepted 14 October 2022
Available online 19 October 2022

Nomenclature

𝟙	Indicator function
i	Variable
N	Number of buildings in the data sample
ANN	Artificial neural network
BA	Basement available
BEP	Building energy performance
BoCY	Boiler construction year
BT	Building type
BuCY	Building construction year
BWM	Basement window material
CV	Coefficient of variation
ED	Exterior design
EPC	Energy performance certificate
EQM	Energy quantification method
ES	Energy source
FD	Facade damage
HDD	Heating degree day
ITI	Interspace thermal insulation
LS	Living space
MLA	Machine learning algorithm
MWF	Material of window frame
OWCT	Outer wall construction type
OWCY	Outer wall construction year
OWIP	Outer wall insulation placement
OWIT	Outer wall insulation thickness
OWM	Outer wall material
OWT	Outer wall thickness
PRI	Presence of roof insulation
R	Roofing
RPB	Rated power of the boiler in kW
RQ	Research question
SAP	Standard assessment procedure
SVM	Support vector machine
TWG	Type of window glazing
URS	Use of roof space
VI	Variable importance
WCY	Window construction year
WH	Water heating
YRC	Year of the last roof covering

quantification methods (EQM) to be applied when issuing EPCs. In the case of Germany, the Building Energy Act is the legal basis for norms like DIN 18599 (Federal Ministry of Justice Germany, 2020), while in the UK the Standard Assessment Procedure (SAP) is prescribed for EPCs (Crawley et al., 2019). Note that EPCs consider different energy end-uses for calculation depending on the country. For example, in German residential buildings, energy for space heating and hot water is considered, while the UK's EPCs additionally consider energy for lighting (Amirkhani et al., 2020; Federal Ministry of Justice Germany, 2020). These prescribed EQMs are under debate for their accuracy in practice (Cozza et al., 2020b; Zou et al., 2018), as they have been found with measured deviations of up to 287% between the actual energy consumption and the EPC value (Calì et al., 2016; Johansson et al., 2016). These so-called engineering EQMs rely on physical laws to calculate thermal dynamics and derive energy behavior, requiring detailed information on building components and measures (Zhao and Magoulès, 2012a). This is why qualified auditors collect data in most EU member states during on-site visits (Arcipowska et al., 2014). For example, they collect data on the thermal transmittance of a building's envelope. Exact measurements are necessary for the engineering EQMs to be accurate (Iribar et al., 2021). It is conceivable that occupants lack skills, knowledge, and perhaps the required tooling to collect such data accurately (Li et al., 2019).

In recent studies, EQMs using machine learning algorithms (MLA) – so-called data-driven EQMs - have been investigated with excellent prospects for real-world application in benchmarking (Seyedzadeh et al., 2020; Veiga et al., 2021; Wenninger and Wiethe, 2021). In contrast to engineering EQMs utilizing relationships from physical laws, MLAs learn from input data. These data may also represent non-physical measures such as the buildings' year of construction, which even occupants might be able to collect reliably.

Previous research on data-driven EQMs, though, focused on the prediction accuracy of MLAs on a given (often even simulated) dataset irrespective of accessibility of the building variables. To that end, Wei et al. (2018) ranked artificial neural networks (ANN), support vector machines (SVM), statistical regressions, and decision tree genetic algorithms as the most prevalent. In comparing data-driven EQMs, Zhao and Magoulès (2012a) identified that ANNs and SVMs are particularly well-suited to predict BEP. At the same time, ANNs are computationally less intensive than SVMs (Wei et al., 2018).

However and given this study's problem statement, central to the accuracy and effort for the frequent creation and renewal of EPCs in practice is the often neglected but error-prone step of data collection (Li et al., 2019), e.g., as in Walter and Sohn (2016), Chae et al. (2016), or Wenninger and Wiethe (2021). To evaluate the feasibility of the broad application of data-driven EQMs, the entire process of issuing EPCs including the design and operational phases must necessarily be considered. This study addresses this research gap by formulating its guiding research question (RQ) as follows:

RQ. How to systematically select building variables that occupants can reliably collect and contribute most to accurate EPC-based benchmarking?

This study addresses the RQ by a four-step method and its validation on single- and two-family houses. This building type represents the vast majority of the residential building stock in Germany (15.7 million (Federal Statistical Office of Germany, 2018)). In addition, its energy use is the largest by building type (German Energy Agency, 2016). Thus, it deserves particular attention. The study analyzes the accessibility of building variables, i.e., how simple it is to correctly collect a building variable (a characteristic of a building like the living space) and derive their variable importance (VI) for predicting BEP by an ANN, i.e., how much they contribute to accuracy. Based on this, the study derives a set of well-selected building variables, which are both easily accessible and important for accurate predictions of the BEP. Subsequently, the study tests whether the method allows for simplifying data collection (validated by letting occupants collect data) while achieving at least the accuracy of qualified auditors using their method. Specifically, it compares our method's prediction accuracy to an out-of-sample dataset, which contains real EPCs issued by qualified auditors applying engineering EQMs. While there is research on VI related to BEP in both residential to commercial sectors and across various types of energy sources and carriers (Ali et al., 2020b; Yuan et al., 2019), research on the accessibility of the identified variables, i.e., how simple it is to reliably collect the variables, is very scarce. Ali et al. (2020b) propose a data-driven approach for geographic information system-based building energy modeling applied to the Irish building stock. They combine engineering judgment and data-driven methods, i.e., various statistical approaches, to identify a subset of the most relevant building variables for their purpose. Yuan et al. (2019) use partial least squares regression and random forest to rank important features for predicting coal consumption of space heating in rural residences in China. Both works aim to increase prediction accuracy by identifying the most relevant variables, commonly referred to as “feature selection” in data science (Li et al., 2018) but do not use the information for any other purposes.

This study combines the perspectives of VI and accessibility for single- and two-family houses using data-driven EQMs to optimize accuracy in practice. Eventually, it derives recommendations based on our stepwise method with accessibility and VI to design accurate and simple data-driven EPCs.

This study strives to contribute in three ways. First, it presents a novel method for designing data-driven EQMs that ensures high practicability and accuracy by considering variable accessibility and variable importance and thereby differs from mainly accuracy-driven EQM development approaches. Second, it gives evidence that simple data-driven EQMs can be more accurate than engineering EQMs on single and two-family houses on real-world data. And third, it provides design recommendations for data-driven EPCs to support policymakers and subject matter experts from practice.

The remainder of this paper is structured in four sections, as graphically illustrated in Fig. 1. After the introduction, Section 2 introduces our method and especially the definition of an accessibility score before Section 3 depicts the datasets. Subsequently, Section 4 reports on and discusses the results before concluding and deriving the implications of our research. Section 5 concludes the paper.

1.2. Process steps for issuing energy performance certificates

The process of issuing EPCs entails the factors that influence their accuracy. This holds for engineering and data-driven EQMs (Wederhake et al., 2022). Often, issuing EPCs is associated only with the operational phase of collecting, calculating, and presenting the results for/of the EPC. However, this assumes that a design phase has already preceded the operational phase. For that reason, both the design and the operational phase in the process and its description are considered. Fig. 2 illustrates both phases, with one process step for the design phase and three for the operational phase. The outcome of the design phase is an EQM that can then be used during the operational phase to determine BEP as a target measure.

Different from engineering EQMs, collecting data for the diverse residential building stock to serve as training data is part of the design step of data-driven EQMs. This occurs before the selection of the MLA as well as before the identification and selection of input variables for training and prediction. To that end, Ali et al. (2020a) report that input variables have widely varying importance for predicting BEP, with only a few variables exhibiting substantial impact. That is why analyzing VI helps improve prediction accuracy with less demanding data requirements (Ali et al., 2020b; Zhao and Magoulès, 2012b). For further details on deriving VI, refer to Section 2.

Regarding data requirements, Chapman (1991) found that there are two main sources of error, which are interdependent. Increasing data requirements[1] (e.g., more granular information on building geometry and its materials) might reduce model error (model quality, error source B). However, on increasing the data requirements, the input error (data quality, error source A) tends to increase as well. Thus, error source A and B should be considered jointly to optimize the outcome, i.e., the accuracy of an EPC. While there are differences in the design phase between data-driven and engineering EQMs, either type should be designed considering both sources of error during the operational phase. According to the literature, the operational phase is a three-step process (Hardy and Glew, 2019; Li et al., 2019; Pasichnyi et al., 2019):

First, the necessary input data are collected and pre-processed (Fabbri and Marinosci, 2018; Hardy and Glew, 2019). For engineering EQMs, this involves on-site visits by the auditor to ensure high data quality (Arcipowska et al., 2014). Collecting data by occupants is thus only theoretically conceivable for engineering EQMs. In our study, this step involves occupants and their respective skillset instead. This different context, therefore, needs to be reflected in the design phase of the data-driven EQM. Second, the BEP is generated by software using an EQM. Third, the EPCs are issued, presenting the results (Pasichnyi et al., 2019).

The analysis of the process steps reveals these two interdependent error sources as a trade-off influencing the accuracy of EPCs as confirmed by Crawley et al. (2019). In the same vein, Arcipowska et al. (2014) and Poel and van den Brink (2009) further confirm that errors in the input data lead to inaccuracies in the calculated EPCs. For a detailed analysis of specific deviations from calculated EPC to observed values for residential buildings, refer to Cozza et al. (2020a).

Fig. 2 summarizes the process steps for creating EPCs and the error sources. In step 1, data quality is the most significant source of error. Notably, the accessibility of single variables is an essential factor influencing data quality because easily collectible variables are less susceptible to interpretation and are less error-prone (Strong et al., 1997). In step 2, the model quality underlying the EQMs is relevant for the accuracy of EPCs.

2. Method development

Fig. 3 presents the four-step method to address the underlying RQ adequately. To this end, this study examines the two central points – the potential for simplified data collection and more accurate predictions – in individual analyses that build on each other.

First, we analyze the accessibility and, therefore, implicitly the accuracy of the building variables that are typically collected for engineering EQMs as input variables. Here, we draw from a comprehensive and unique real-world dataset of 25,000 German single- and two-family houses. The raw dataset comprises 76 variables characterizing each building and the measured annual energy consumption. All data is collected by occupants. We present the dataset and pre-processing procedure to ensure high data quality in more detail in Section 3. After the initial data preparation, we introduce the *accessibility score* that measures the collectability by non-experts to determine a variable. We check each variable in the dataset for missing, undefined, or compared with correlated variables logically implausible values (e.g., insulation thickness equals 0 cm, but the presence of insulation is stated).[2] The accessibility score then represents the accessibility for a variable on a defined range of [0, 100], with higher scores indicating that it is easier to identify that variable as a non-expert. Equation (1) calculates the accessibility score in three steps: First, we count all incorrect values for a given variable, i.e., the higher the value, the more difficult it is to assess the variable correctly. Second, we take the logarithm of this value to allow for comparisons in different magnitudes. Because the logarithm is monotonously increasing, it still holds that the higher the value, the more difficult it is to assess the variable correctly. Third, we normalize the value to [0, 100] by subtracting and dividing from the maximum possible error. Here, we reverse the meaning, i.e., higher values indicate that the variable is less difficult to assess.

$$\textit{accessibility score}_i = 100 \cdot \frac{\log(N) - \log\left(\sum_{n=1,\ldots,N} \mathbb{1}_{\{value\ i\ incorrect\}}\right)}{\log(N)} \tag{1}$$

Here, $\mathbb{1}$ gives the indicator function that takes the value 1 if the condition is fulfilled and the value 0, else. N is the number of buildings in the sample, hence, it is also equal to the maximum possible error; i iterates over all variables.

Second, we investigate the VI for a data-driven EQM. Following

[1] E.g., Any type of data to be collected or measured. Requirements increase with the level of detail.

[2] Note, that this accessibility score is an indication, as there might be further incorrect values, which cannot be determined.

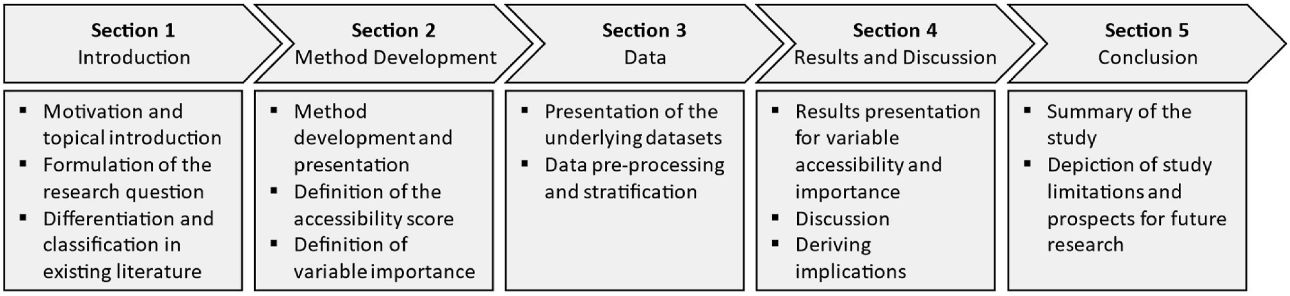

Fig. 1. Graphical illustration of the paper's structure split into five sections.

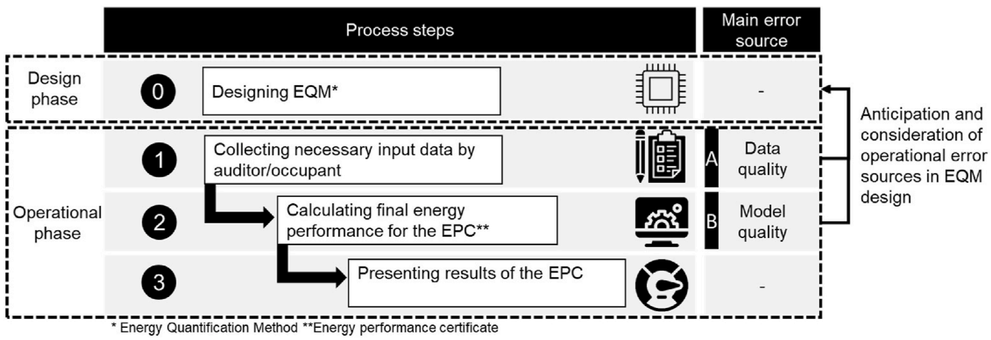

Fig. 2. Process of issuing energy performance certificates and its main error sources (adapted from Wederhake et al. (2022)).

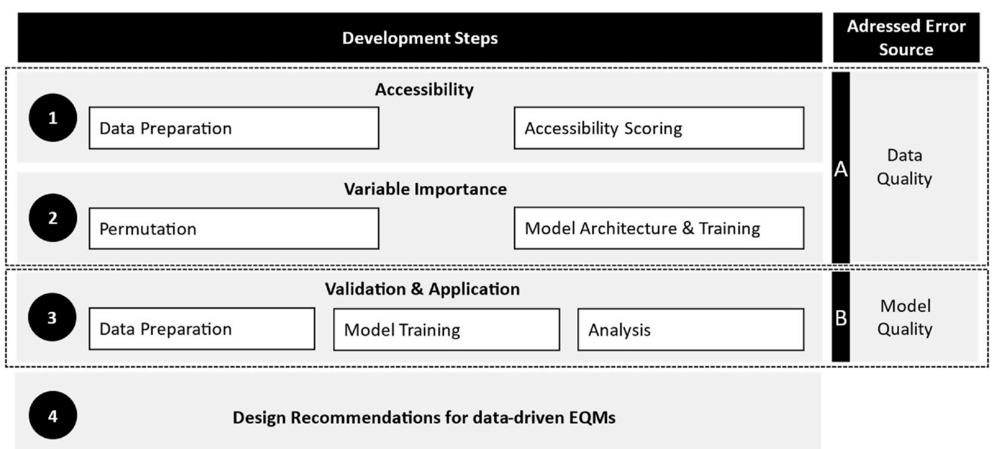

Fig. 3. Four-step research procedure of this study to derive guidelines for MLA-based EPCs.

standard data science practices (Kaymakci et al., 2021), we set up the data-driven EQM's model architecture and train the model with the training dataset. We then derive the VI for each variable to accurately predict the BEP. To ensure the generalizability of the results, we build on existing literature. Therefore, we set up the model architecture based on Buratti et al. (2014), who predicted annual energy performance in the residential sector using an ANN. We refer to Buratti et al. (2014) for implementation details to ensure reproducibility and only report on diverging implementation details of the data-driven EQM. We use the Adam optimization algorithm instead of back error propagation and set the learning rate to 10^{-4}. We train the model with batch sizes of 16 data points over 500 epochs while using an early callback to avoid overfitting when no substantial improvement in prediction accuracy has been achieved. To comply with state-of-the-art methods in the MLA domain, we apply tenfold cross-validation to train the models and evaluate their performance on the folds. To derive the VI, we use the MLA-agnostic method suggested by Breiman (2001). It permutes, i.e., randomly shuffles, each input variable and measures the decrease in prediction accuracy. The higher the decrease in accuracy after permutation, the more important the variable must have been for prediction, i.e., the higher the VI – similar logic as for sensitivity analysis. We additionally normalize the log-VI to provide a clearer presentation. After analyzing the accessibility and importance, we cluster each variable in the two-dimensional space represented by the accessibility score and the VI by the k-means algorithm (Likas et al., 2003). This allows deriving variables that are both easily accessible and highly important for accurate predictions of the BEP.

Third, we validate and apply our findings to an out-of-sample dataset with EPCs issued by qualified auditors. We prepare the data from the initial dataset with the relevant variables identified at the end of step two and subsequently train an ANN with the same model architecture on the selected variables. After training, we predict the BEP of the out-of-sample dataset and analyze the results with the calculated values of the EPCs. Since we strive for real-world applicability and representativeness of our results, we apply post-stratification to our results. The principle of stratification describes sampling subpopulations independently, such that the relation between the numbers drawn from each subpopulation equals the actual relation over the whole dataset (Bowley, 1925). If the stratification takes place after the data collection, the term post-stratification is used. In this sense, the results are weighted to reflect the overall distribution regarding a specific characteristic. Throughout this study, we stratify the results regarding the buildings' years of construction.

3. Data

The target measure for our data-driven EQM is the BEP, as mentioned in Section 1.2. To calculate the target measure and subsequently train our model, we have four datasets at hand that were also previously used in research and described to explore various MLAs and procedures for BEP predictions (Wenninger and Wiethe, 2021): first, we have a training dataset comprising 25,000 single- and two-family houses from Germany registered from April 2007 to January 2014. Occupants collected the data in a survey in return for energy retrofitting recommendations. Consequently, even if they were non-experts, the occupants had an incentive to participate in the survey and provide as much input as possible and provide it correctly at the same time. The dataset contains 35 variables, e.g., information on the installed heating system, presence and thickness of insulation layers for walls and roofs, annual thermal energy consumption, and geographic location. Variables describing socio-economic effects and occupant behavior are not included in the dataset. Figs. 4 and 5 provide some descriptive statistics for the dataset. The survey questionnaire can be found in the appendix. This dataset serves as a basis for conducting the (development) steps outlined in Section 2 and training our data-driven EQM. Second, we have a separately collected validation dataset containing 345 additional single- and two-family houses. The data originate from two energy auditing companies employing qualified auditors from Baden-Württemberg, a federal state of Germany, and contain 42 variables. Qualified auditors originally collected the data during on-site visits and used them to create calculated EPCs. The auditors also included the measured BEP in the dataset. At the time of data collection, the auditors could only disclose 345 datapoints. Third, we use the distribution of classes of building construction year for single- and two-family houses of the entire German building stock from the German micro census (Federal Statistical Office of Germany, 2011). The micro census is a statistical and representative survey. We use this dataset for stratification to ensure the representativeness of our results. Fourth, we draw on data on the annual heating degree-days (HDD) for the last 50 years for 42 weather stations in Germany, published by Deutscher Wetterdienst (2020). We use this dataset to rectify weather effects from the measured BEP. In the following, we first describe how we derive the target measure for the training dataset.

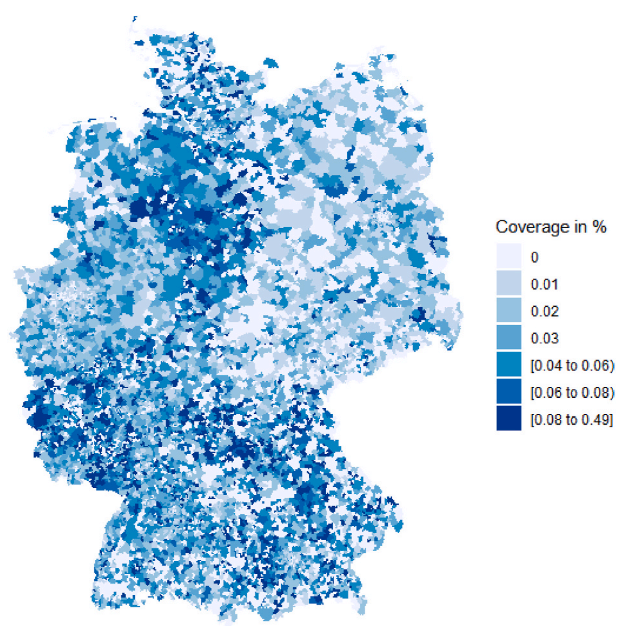

Fig. 4. Illustration of the local distribution of the buildings in the training dataset across zip codes in Germany (visualizations taken from (Wenninger et al., 2022a)).

We calculate the BEP by adjusting the measured energy consumption for weather effects to ensure comparability and robust results. For this purpose, we adopt the approach of Wenninger and Wiethe (2021) and assign the respective HDD (mean over the time the datasets were gathered) to each building based on the zip code, and approximate the effective building area with the living space.

To ensure high data quality and consistent and valid results, we preprocess the raw dataset by deleting outliers and applying the two-stage LANG approach (Zhang et al., 2019) to test for semantic and syntactic data constraints. More precisely, we first remove faulty and contradicting data entries, e.g., if the outer wall or roof construction year is earlier than the building's construction year. We count each entry removed for the respective variable, which caused the discard to calculate the accessibility score. Second, we remove variables that do not contribute to energy consumption, e.g., identification numbers. Here, we do not add an accessibility score because these variables do not bear explanatory power. The resulting dataset contains 10,220 buildings (40,8% of the total dataset) with 26 variables (25 independent variables) to train our model. For details, refer to Table 3 in the Appendix.

Last, to ensure representativeness, we stratify our results such that they resemble the distribution of the entire German building stock for single- and two-family houses in terms of classes of building construction year as given in the micro census. In our case, the subpopulations are the classes of building construction years from the micro census. Fig. 6 depicts the percentage shares of the classes of building construction year both in the micro census and in our training dataset to ensure transparency.[3]

We proceed analogously to clean the validation data. Some training and validation dataset variables describe the same building properties but differ in the respective scale or the nominal specification. To meaningfully validate our ANN on the same data used for training, we adapt the validation data from the respective variables to the training data. We eventually discard each variable only present in one dataset, resulting in 14 variables.

4. Results and discussion

4.1. Accessibility and variable importance

This section presents the results to our RQ. We first present the accessibility score and VI results. We then compare our data-driven EQM's prediction accuracy to the qualified auditors', using the Coefficient of Variation (CV) (Amasyali and El-Gohary, 2018) as the performance evaluation measure, before we derive policy implications. To this end, we first train an ANN on the entire prepared dataset (all 25 independent variables) and evaluate the VI. Fig. 7 displays the results of the accessibility score and VI calculated according to Equation (1) and the method suggested by Breiman (2001); the variables' names are given in Table 1 with further information in the Appendix.

[3] Note, that the census only provides data in more aggregated form to avoid conclusions most effectively. As a result, eight classes of building construction year are available for the distribution. We merged the two most recent classes because there are too few data points thus avoiding distortion of results after scaling.

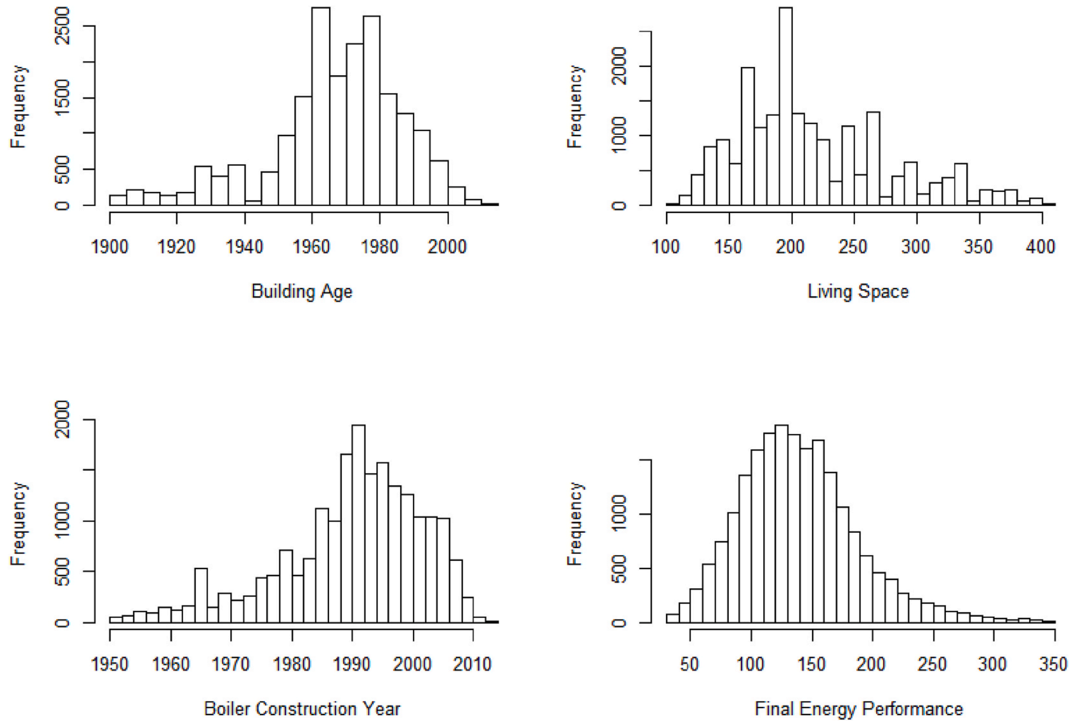

Fig. 5. Descriptive statistics for the training dataset represented as histograms (visualizations taken from (Wenninger et al., 2022a)) – Living Space in [m^2] and Final Energy Performance in [kWh/m^{2}a].

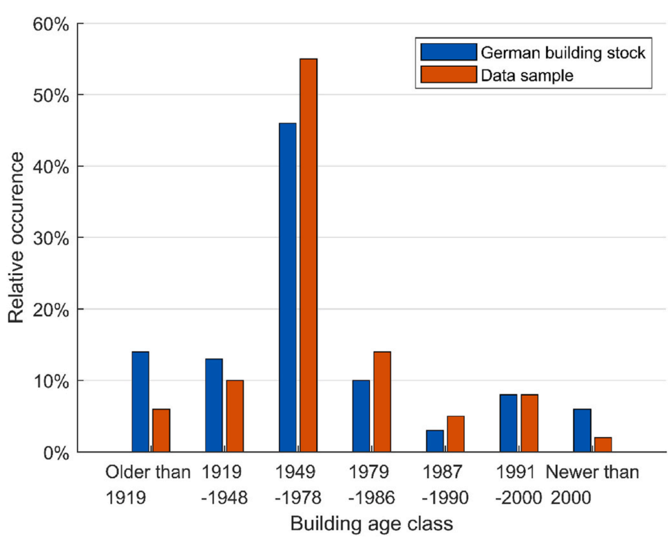

Fig. 6. Classification of buildings' construction year, real-world distribution in Germany, and distribution in the training dataset.

The horizontal axis depicts the normalized log$_2$-VI, while the vertical axis depicts the normalized log$_2$-accessibility score. Again, the higher the accessibility score, the easier it is for a non-expert to collect the variables on his/her own. The colors indicate the individual clusters resulting from the k-means clustering algorithm. The clusters are distributed across the space spanned by accessibility score and VI. Also, the clusters are clearly distinguishable from one another. The variables located in the lower-left area (colored blue) are supposedly difficult to collect correctly. They expose low accessibility scores and a relatively low VI. The power of the heating boiler (given in kW) stands out with the lowest accessibility score but ranks with a relatively high VI. Reading the specification label on the boiler should be simple and easy in theory. However, the labels often do not state the power in the unit [kW] but in [kWh/min] or in [m^3/min]. In that way, technically less skilled occupants might face difficulties identifying it correctly. From an implementation perspective, this variable is particularly interesting because, although it improves accuracy due to a high VI, it is difficult to collect and prone to error. Uniform or easy-to-read values on the boilers could provide better accessibility here and further enhance data-driven EPCs by occupants. The cluster in the upper-right area (colored red) contains variables that are supposedly easy to collect but at the same time, have a high VI. It is unsurprising to find variables such as living space or energy sources in this cluster. Occupants can easily obtain data on living space from rental contracts or central documents of the properties. On the other hand, the variables in the upper-left area (colored green) are not so easy to classify, as they are located in a relatively wide range of VI and accessibility scores. They have a higher accessibility score than the blue cluster and lower VI than the red cluster. Therefore, comparing the VI and accessibility score of these variables with findings in previous studies is especially relevant as focusing solely on prediction accuracy may not be sufficient (Wenninger et al., 2022b). Here, some results partially diverge from the literature. For instance, the outer wall thickness is directly proportional to its insulation capacity (Asdrubali et al., 2015; Enshen et al., 2005), yet it ranks low in VI. This is possibly due to the data-driven EQM learning underlying dependency structures between several variables. It thus implicitly derives information from outer wall thickness without further relying on it. To this end, the results do not suggest that outer wall thickness alone is unimportant for the BEP. Instead, it reflects that the additional information gain is limited. This deduction of further information may thus constitute an advantage of the data-driven EQM, especially for otherwise hard-to-access variables. The same might hold for other building characteristics, such as the effect of buildings' thermal mass on energy consumption which is discussed controversially in literature (Al-Sanea et al., 2012). Depending on occupancy and climate conditions, a building's thermal mass might have differing effects (Reilly and Kinnane, 2017). Further, regionally varying regulations, standards, or the availability of materials can complicate the correct determination and thus diminish gains in prediction accuracy.

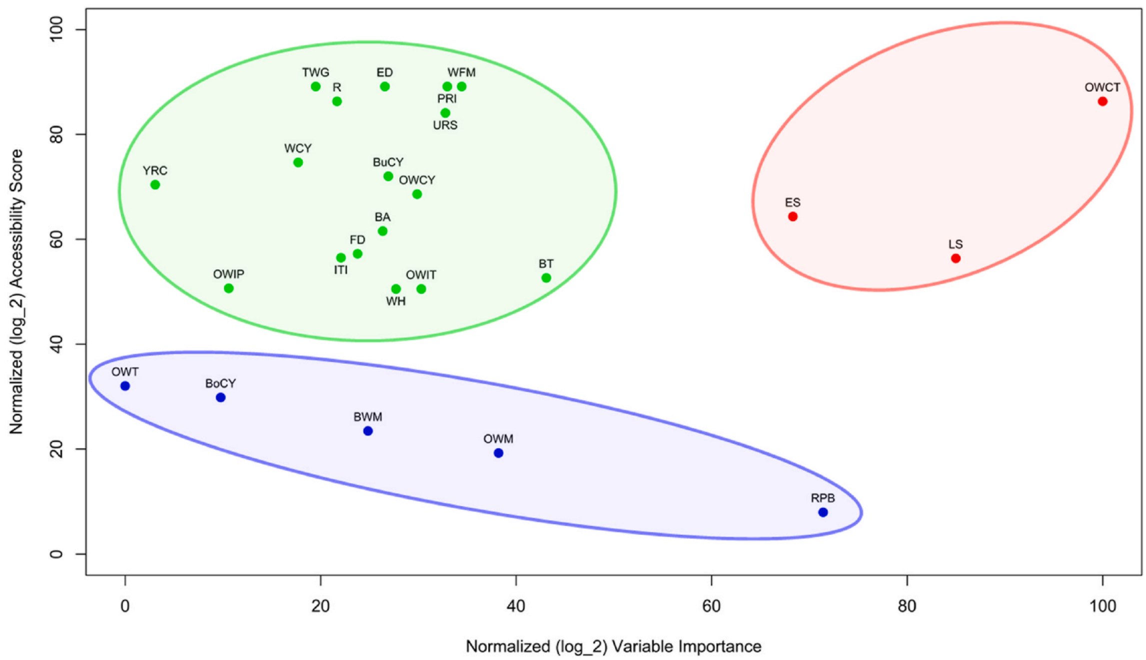

Fig. 7. The points indicate the normalized $\log_2$-variable importance and accessibility score for all available variables except the dependent variable in the training dataset.

Based on these results, we subsequently discuss general insights from the perspectives of VI and accessibility. First, regarding data collection, our study confirms related research from other domains that outline that more data (especially in terms of the number of variables) does not necessarily improve accuracy (Georganos et al., 2018; Zhang et al., 2018). When there are more variables, there is also the need for more data (records) to prevent overfitting. As outlined before, collecting data is not equally easy for all variables in this vein. To that end, analyzing input variables by accessibility in relation to their relative importance serves as a helpful tool for selecting the variables on which to build data-driven EQMs.

Second, turning to accessibility, there are three important findings. First, this study deliberately chose to measure accessibility by letting occupants voluntarily collect pre-qualified information influencing BEP. This helps draw a realistic picture of what data-driven EQMs could deliver under real-world conditions. Occupants provide the data they can and want to turn in, which allows deriving the number of valid responses per variable. The number of valid responses provides a straightforward, quantifiable measure. This allows comparing the various variables among each other in terms of accessibility. From the comparison, it is observable that the accessibility of variables relevant for energy quantification varies in orders of magnitude for occupants, e.g., the power rating of the boiler versus the window glazing. This finding suggests a practical implication: some variables cannot be obtained by simply investing a little more time by the occupant. Instead, missing knowledge and tooling must be considered additional confounding factors. Second, the accessibility score for occupants reveals that accessibility highly differs between variables. This is interesting as it allows discriminating variables to ensure high data quality. Third, the analysis has also raised concerns regarding common beliefs, e.g., that correctly identifying the material of the basement's windows is easy. To that end, the empirical analysis of accessibility has demonstrated its usefulness in uncovering non-obvious difficulties in data collection.

Third turning to the VI, there are two important findings. First, VI might be guessed a priori, e.g., based on previous studies. However, it can hardly be perfectly anticipated. This is due to at least three reasons. First, VI might differ among data-driven methods and certainly among regions. Second, VI is relative to other attributes taken into account. Third, VI is influenced by variable accuracy, i.e., the input data provided. A variable like living space might be highly important when the variable is accurate. However, it would be less critical when occupants could not specify the living space accurately. In contrast, controversial discussions in the literature about the lack of accuracy of the living space or the heated floor area are to be mentioned here (Platten et al., 2019). Particularly noteworthy is the application-oriented character of the method for the design of data-driven EQMs by considering possible sources of error already in the design phase.

4.2. *Accuracy*

After presenting and discussing the VI and accessibility, the focus lays next on the prediction accuracy of both EQMs. Doing so requires information on the BEP calculated by qualified auditors, the actual annual BEP, and the BEP predictions of the data-driven EQM (the ANN restricted to the important and easily accessible variables). The deviation between the actual BEP and the data-driven EQM's prediction gives the prediction accuracy of the data-driven EQM. In contrast, the deviation between the metered energy consumption and the calculated values by the qualified auditors gives the prediction accuracy of the engineering EQM. Based on these values, we can assess whether data-driven EQMs can potentially supersede engineering EQMs used in practice by qualified auditors. Thus, we exclude all variables from the lower-left area (blue cluster) in Fig. 7 – the variables where the complexity of deriving them exceeds their importance – and all variables, which are not available in the validation dataset. This is necessary, as otherwise, we would train the model on data unavailable for validation purposes.

We apply the trained data-driven EQM to our validation data and calculate the respective stratified CV. Furthermore, we also calculate the CV for the BEP depicted by the EPCs. As a result, the data-driven EQM yields a mean CV[4] of 41.1% compared to 54.1% from the engineering EQM (cf. Fig. 8 horizontal lines), which equates to comparative

[4] Note, that the mean is weighted by the number of buildings in the individual building classes.

Table 1
Variable names (for further details, refer to Table 4 in the Appendix).

Abbreviation	Variable	Cluster assignment		
		Low VI, high accessibility (green)	Low VI, low accessibility (blue)	High VI, high accessibility (red)
BA	Basement available	X		
BoCY	Boiler construction year		X	
BT	Building type	X		
BuCY	Building construction year	X		
BWM	Basement window material		X	
ED	Exterior design	X		
ES	Energy source			X
FD	Facade damage	X		
ITI	Interspace thermal insulation	X		
LS	Living space			X
MWF	Material of window frame	X		
OWCT	Outer wall construction type			X
OWCY	Outer wall construction year	X		
OWIP	Outer wall insulation placement	X		
OWIT	Outer wall insulation thickness	X		
OWM	Outer wall material		X	
OWT	Outer wall thickness		X	
PRI	Presence of roof insulation	X		
R	Roofing	X		
RPB	Rated power of the boiler in kW		X	
TWG	Type of window glazing	X		
URS	Use of roof space	X		
WCY	Window construction year	X		
WH	Water heating	X		
YRC	Year of the last roof covering	X		

advantage of 35% (after calculating the square root to ensure linear interpretability). This roughly translates to the error produced by the data-driven EQM being almost half the error produced by the engineering EQM. Thus, the data-driven EQM, using only data that is easily collectible by non-experts with high VI, exhibits significantly better results than the calculated EPCs conducted by qualified auditors.

To increase the transparency and validity of our analyses, the result are disaggregated and examined in different instantiations of construction years more closely. Table 2 and Fig. 8 depict the results of both

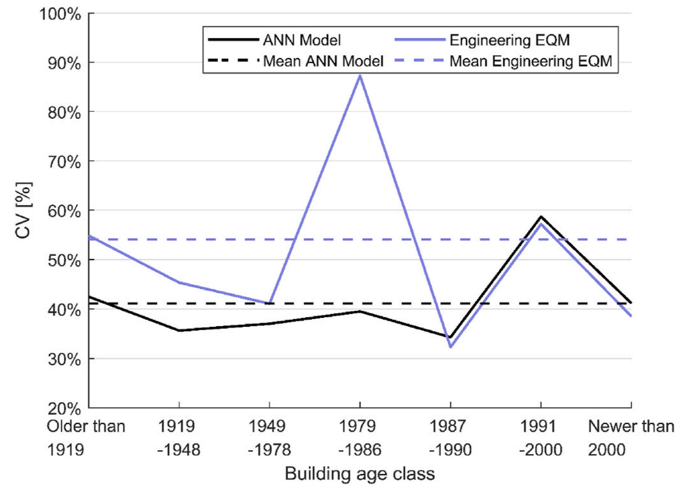

Fig. 8. Course of CV per buildings' construction year class for the data-driven and engineering EQMs.

EQMs regarding different classes of building construction years and visualize the magnitude of the CVs for both EQMs, respectively.

The vertical axis depicts the magnitude of the CV, and the horizontal axis depicts the different classes of building construction years taken from the census. The solid black line represents the data-driven EQM's performance, with the dashed black line indicating its mean performance value. The solid blue line depicts the results of the engineering EQM, with the dashed blue line visualizing its mean performance value regarding the CV. For the data-driven EQM, there is a general trend toward more accurate predictions for newer buildings up to 1990.

There is one pronounced spike in the engineering EQMs' performance for the class of building construction year from 1979 to 1986. However, this spike might appear more pronounced due to a quadratic operation for the calculation of the CV and may be caused by outliers distorting the result. Note that this demonstrates that even qualified auditors performing on-site visits cannot always adequately assess the energy performance. However, this is one of the premises why on-site visits by qualified auditors are enforced. The data-driven EQM produces better results than the engineering EQM across almost all classes of building construction years. This is underlined by similar results arising for other performance evaluation measures. Moreover, the data-driven EQM exhibits a significantly smaller spread in the results than the engineering EQM. The newest classes of building construction years exhibit similar accuracy for the engineering EQM and the data-driven EQM.

Comparing the results of the data-driven EQM with the preselected variables to the engineering EQM, a general tendency towards more exact predictions for newer buildings erected before 1990 can be observed. This pattern might result from lower input data quality. Since we use the same model for all classes of building construction years, the only significant difference between the older and newer buildings is the quality of the input data gathered by qualified auditors or occupants. For instance, in the case of an older building, it is a challenge to identify the materials or construction methods used for the building and its components with high accuracy (Claesson, 2011; Foucquier et al., 2013), which leads to lower data quality and thus to less accurate predictions. The engineering EQM achieves slightly higher accuracy than the data-driven EQM for the newer classes of building construction years. However, as Fig. 6 depicts, we dispose of three times fewer buildings in the training dataset for the newest class of building construction year than required to resemble the German distribution. Although trying to account for that by post-stratification, the data-driven EQM could not fully learn underlying dependency structures for the newest class of building construction year. This is further aggravated by the training

Table 2
CV per buildings' construction year for both data-driven and engineering EQM.

Class of building construction year	<1919	1919–1948	1949–1978	1979–1986	1987–1990	1991–2000	>2000
data-driven EQM	42.5%	35.6%	37.0%	39.5%	34.3%	58.7%	41.1%
engineering EQM	54.9%	45.4%	41.1%	87.3%	32.3%	57.2%	38.5%

dataset being collected between 2007 and 2014. Hence, no buildings constructed from 2015 onwards are included in the training dataset. In fact, with the correct training dataset, the most accurate results for this class of building construction year due to more precise construction methods and data quality are expected.

Predictions performed by qualified auditors do not improve accuracy (or only seldom and slightly) for buildings with the same characteristics as in this study can be concluded. Our empirical results illustrate that, instead, the opposite is true, namely that the data-driven EQMs on data collected by occupants can reliably achieve equal or more accurate results when assessing German residential single- and two-family houses - even with non-expert knowledge.

4.3. Policy and practical implications

The results presented could influence policy initiatives and regulatory frameworks in the area of EPCs, as reflected in the following implications relevant to policymakers. Beyond that, implications for practice can be derived from our findings.

First, this study concludes that when designing an EPC based on a data-driven EQM, its beneficial to consider and measure VI and accessibility simultaneously. For that purpose, the presented method of this study may serve as a starting point. However, we suggest that the variables might differ in number and combination depending on the region, building type, energy use form, and the chosen data-driven method. That is why we highly recommend policymakers and practitioners conduct a similar study for new settings to obtain accurate results. Clearly, that comes at a cost. However, data-driven EPCs could minimize the costs of BEP benchmarking. Assuming a conservative € 300 cost per EPC (Grabolle, 2015) and an equally conservative 200,000 issued EPC per year in Germany, this results in € 60 million in costs. To that end, the presented method allows for scaling BEP predictions because of its low-effort basis compared to qualified auditor-based formal EPCs. Consequently, more regular updates of comparable BEP benchmarking might inform local policymakers on energy efficiency progress.

Second, this study provides evidence that occupants can collect some set of attributes accurately enough (occupants collected all training data in this study) such that an established and well-documented data-driven method (Buratti et al., 2014) delivers more accurate predictions than engineering EQMs. This finding can be re-produced following our method for other regions, other building types, and alternative forms of energy use. To that end, we recommend considering a certification process for data-driven EQMs similar to the standards and norms procedures for engineering EQMs. These new procedures might require testing against data collected by or on behalf of a third-party organization that ensures that data collection procedures are similar for both training data and EQM certification. It is crucial to involve auditors closely as key stakeholders in the validation and evaluation process to develop practical, applicable, and acceptable solutions. The introduction of such methods and procedures clearly demands revising education and training material for qualified auditors to leverage the potential of data-driven EQMs in issuing EPCs entirely. Considering accessibility as an influencing factor for EQMs' accuracy is a novel concept in the policy-related debate on benchmarking BEP. Previously, the focus rested on the qualified auditor only. While it might seem apparent that not all data is equally accessible to occupants, this study raises some concerns about how well-qualified auditors collect their required data. In this regard, it is advised to have education and training material reflect the accessibility of variables, especially concerning their importance. In addition, the standard values that auditors apply for their engineering EQM may be reconsidered and revised where it seems appropriate.

Third, the presented method actively involves occupants in the BEP benchmarking process and thereby opens the process to user groups previously only – at best – having a passive role. However, awareness of energy efficiency and the vast potential of the building sector is a crucial first step to fostering acceptance of energy efficiency and related measures. Active involvement supports increasing awareness. Thus, policymakers are recommended to consider more intensively how to involve occupants in issuing EPCs to increase acceptance of – and thus exploit – the existent energy efficiency potential in the residential building stock. The same is true for practitioners and auditors, as greater adoption could also lead to greater confidence and, thus higher investment in energy efficiency and, consequently, higher turnover.

Last, when occupants collect data, checks and controls might become more relevant than today if benchmarking results should not only inform the occupants themselves. A digital process for validation is likely to apply here. Given that banks already today assign mortgage loans for some types of buildings without on-site visits by a real estate assessor, this study supports the hypothesis that this can be done for energy performance benchmarking as well. To that end, looking into digital validation processes more, e.g., through remote sensing, is recommended. Professional organizations in Europe and the US already perform such services in banking and insurance (Betterview, 2021; credium GmbH, 2021). New providers and business models could develop for such services in the digital energy ecosystem.

5. Conclusion

Energy retrofitting and energy performance certificates (EPC) are pivotal for the European Commission's aim of reducing net greenhouse gas emissions across sectors by at least 55% from 1990 levels by 2030 (European Commission, 2021). However, research frequently questioned the accuracy of official EPCs for building energy performance (BEP) benchmarking despite being carried out by qualified auditors. In contrast, previous research has demonstrated that data-driven energy quantification methods (EQMs) can learn from non-physical and input variables that are potentially simpler to collect, while evidence for data-driven EQMs' accuracy is growing. This study examined whether these traits of data-driven EQMs can be utilized to allow building occupants to collect input data reliably. A method that allows identifying the variables of importance, measuring the ability to have them collected by occupants (accessibility), and eventually building a data-driven EQM based on the identified variables is presented. The resultant data-driven EQM is the basis for comparing its predictions with official EPC data. This study reports a 35% improvement in accuracy by the data-driven EQM applied to data collected by occupants. As a central contribution, a stepwise method to design data-driven EPCs is proposed, design recommendations are outlined, and implications for policy and practice derived.

Naturally, as with any research endeavor, we point to the study's limitations and boundaries of generalizability of the results. This study focused on single- and two-family houses, occupants as data collectors, and Germany as geography. Therefore, it is relevant to note that accessibility and variable importance can vary over all these dimensions. For example, it is reasonable to assume that the accessibility score for qualified auditors would be lower for some or perhaps many attributes. Accessibility can vary even depending on the class of building construction year. Particularly for older buildings, information tends to be scarcer and less accessible. Studies with other groups of data collectors (e.g., qualified auditors), building types, and other regions require new empirical analyses. Different results are expected due to the varied building construction forms, revisions occur more frequently and diligently, or auditors are comparably paid more or fined more for poor quality. To make this study robust, we have intentionally selected two different energy auditing companies and asked for their EPC data. Different auditors collected the data. Nonetheless, not all risks can be ruled out both auditing companies have been operating below average regarding the accuracy of EPCs. Therefore, further validation for German single- and two-family houses is beneficial and welcome.

Regarding future research, we outline some focal points. First, applying the presented EQM to data collected by qualified auditors instead of occupants can be highly insightful. Second, it is interesting to further identify potentials for reducing time and effort and, thus the costs involved with on-site visits. Third, as stated before, we see merit in applying this method in other geographies and for different building types. It will be of interest to research to identify common and distinguishing variables of importance and of the (high/low) accessibility as well as the analysis of variables describing the operation, use, and occupancy (especially for non-residential buildings). Fourth, future research might extend our findings to hybrid EQMs combining aspects of engineering and data-driven EQMs. Fifth, researchers could analyze how other calculation approaches for the accessibility score or methods to quantify variable importance might influence the variables (al)location to different clusters.

This study provides a practical method and first evidence for designing data-driven and accurate EPCs for benchmarking BEP to support environmentally friendly energy policymaking.

CRediT authorship contribution statement

Lars Wederhake: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Project administration. **Simon Wenninger:** Software, Validation, Conceptualization, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Christian Wiethe:** Software, Validation, Conceptualization, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Gilbert Fridgen:** Supervision, Funding acquisition. **Dominic Stirnweiß:** Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgment

The authors gratefully acknowledge the financial support of the Bavarian state ministry of economics and media, energy, and technology (grant number IUK-1606-0002//IUK491/001) as well as the Ministry of the Environment, Climate Protection and the Energy Sector Baden-Wuerttemberg (grant number BWT17004). The authors are responsible for the contents of this publication. This research was funded in part by the Luxembourg National Research Fund (FNR) and PayPal, PEARL grant reference 13342933/Gilbert Fridgen.

Appendix

Table 3
Overview of input variables and associated values for our data-driven EQM.

Attributes	Values	Abbreviation
Miscellaneous		
Basement available	Yes, no	BA
Building construction year	Numeric value (year)	BuCY
Building type	Detached, attached	BT
Living space	Numeric value (m^2)	LS
Water heating	Central, central during winter, separate, no water heating	WH
Wall		
Outer wall thickness	Numeric value (cm)	OWT
Outer wall construction year	Numeric value (year)	OWCY
Outer wall insulation thickness	Numeric value (cm)	OWIT
Outer wall material	Yes, no	OWM
Outer wall insulation placement	Inside, outside, none	OWIP
Outer wall construction type	Solid single shell, double-shell air-layer masonry, prefabricated, half-timbered	OWCT
Interspace thermal insulation	Numeric value (cm)	ITI
Exterior wall design	Not specified, plastered. with cladding, exposed masonry	ED
Facade damage	None/unknown, few, many	FD
Heating system		
Type of energy source	Oil, gas, district heat, heat pump	ES
Boiler construction year	Numeric value (year)	BoCY
Rated power of the boiler	Numeric value (kW)	RPB
Windows		
Window frame material	Wood, plastic, aluminum	WFM
Type of window-glazing	Single, double old, double modern, triple-glazing thermal insulated	TWG
Window construction year	Numeric value (year)	WCY
Basement window material	Wood, plastic, aluminum	BWM
Roof		
Presence of roof insulation	Unknown/none, partial, full	PRI
Year of the last roof covering	Numeric value (year)	YRC
Roofing	Good as new, still passable, in need of modernization	R

(*continued on next page*)

Table 3 (*continued*)

Attributes	Values	Abbreviation
Use of roof space	Expanded, expandable, not expandable, flat roof	URS
Final energy performance		
Weather effects adjusted final energy performance	Numeric value (kWh/(m^{2}a))	BEP

Table 4
Survey questionnaire (translated to English)

Nr.	Question
1	Please enter the zip code of your residential building.
2	Please enter the living space of your residential building. Note: The living area corresponds to the sum of the areas of all habitable floors of your residential building excl. basement and attic.
3	Please indicate the time period in which your residential building was constructed.
4	Please indicate the type of your residential building.
5	Is the roof of your residential building insulated?
6	Please indicate the roof shape of your residential building.
7	Please indicate the year in which your residential roof insulation was installed.
8	Please specify the insulation thickness of the roof insulation of your residential building. Note: Indication in cm
9	Is the facade (outer wall) of your residential building insulated?
10	Please specify the insulation thickness of the facade insulation of your residential building. Note: Indication in cm
11	Please specify the material of the facade insulation of your residential building.
12	Please indicate the year in which the facade insulation of your residential building was installed.
13	Please indicate the year in which the windows of your residential building were installed. Note: If windows of your residential building were installed at different times, please consider only the year of installation of those windows with the largest surface area.
14	Please enter the exact U-Value of those windows that account for the largest share of your residential building in terms of area. Note: If you do not know the exact U-Value of the windows, please leave the field blank.
15	Please indicate the most common window frame material used in your residential building.
16	Please indicate the most common type of glazing used in the windows of your residential building.
17	Is the basement of your residential building heated?
18	Is the basement wall of your residential building insulated?
19	Is the basement ceiling of your residential building insulated?
20	Please indicate in which year the insulation of the basement wall of your residential building was installed.
21	Please indicate in which year the insulation of the basement ceiling of your residential building was installed.
22	Please specify the insulation thickness of the basement walls of your residential building. Note: Indication in cm
23	Please specify the insulation thickness of the basement ceiling of your residential building. Note: Indication in cm
24	Please specify the insulation material of the basement wall insulation of your residential building.
25	Please specify the insulation material of the basement ceiling insulation of your residential building.
26	Please indicate in which year the heating system of your residential building was installed.
27	Please indicate the type of heating system used in your residential building.
28	Please indicate the type of boiler of the heating system of your residential building.
29	Do you use solar thermal on your residential building in addition to your heating?
30	Please indicate the energy source you use for your residential building. Note: 1) If you use more than one energy carrier, please select the option “Mixed”. 2) The units in brackets indicate in which unit the energy consumption per energy carrier is to be indicated.
31	Please indicate the consumption of the energy carrier(s) indicated above for the last three years. Note: 1) Please use the energy carrier specific units of measure to indicate consumption. See previous question. 2) If you use multiple energy carriers in parallel, please enter your total energy consumption in the text box and select “kWh” as the unit in the previous question. 3) Please provide consumption for at least one of the three years. Ideally, enter your consumption for all three years.
32	Please indicate the year of construction of your residential building.
33	Do you have an energy certificate for your residential building?
34	What type of energy certificate do you have for your residential building?
35	Please indicate the energy demand entered on your energy demand certificate: Note: The energy demand value has the unit kWh/m^2 a
36	Please indicate the energy consumption indicated on your energy consumption certificate: Note: The energy consumption value has the unit kWh/m^2 a

References

Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the Energy Performance of Buildings.

Al-Sanea, S.A., Zedan, M.F., Al-Hussain, S.N., 2012. Effect of thermal mass on performance of insulated building walls and the concept of energy savings potential. Appl. Energy 89, 430–442.

Ali, U., Haris Shamsi, M., Hoare, C., Alshehri, F., Mangina, E., O’Donnell, J., 2020a. Application of intelligent algorithms for residential building energy performance rating prediction. In: Corrado, V., Fabrizio, E., Gasparella, A., Patuzzi, F. (Eds.), Proceedings of Building Simulation 2019: 16th Conference of IBPSA. IBPSA, pp. 3177–3184.

Ali, U., Shamsi, M.H., Bohacek, M., Purcell, K., Hoare, C., Mangina, E., O’Donnell, J., 2020b. A data-driven approach for multi-scale GIS-based building energy modeling for analysis, planning and support decision making. Appl. Energy 279, 115834.

Amasyali, K., El-Gohary, N.M., 2018. A review of data-driven building energy consumption prediction studies. Renew. Sustain. Energy Rev. 81, 1192–1205.

Amirkhani, S., Bahadori-Jahromi, A., Mylona, A., Godfrey, P., Cook, D., 2020. Impact of adding comfort cooling systems on the energy consumption and EPC rating of an existing UK hotel. Sustainability 12, 2950.

Arcipowska, A., Anagnostopoulos, F., Mariottini, F., Kunkel, S., 2014. Enery Performance Certificates across the EU. A Mapping of National Approaches. https://bpie.eu/wp-content/uploads/2015/10/Energy-Performance-Certificates-EPC-across-the-EU.-A-mapping-of-national-approaches-2014.pdf. (Accessed 5 August 2019).

Asdrubali, F., D’Alessandro, F., Schiavoni, S., 2015. A review of unconventional sustainable building insulation materials. Sustain. Mater. Technol. 4, 1–17.

Balaras, C.A., Dascalaki, E.G., Droutsa, K.G., Kontoyiannidis, S., 2016. Empirical assessment of calculated and actual heating energy use in Hellenic residential buildings. Appl. Energy 164, 115–132.

Ballarini, I., Corgnati, S.P., Corrado, V., 2014. Use of reference buildings to assess the energy saving potentials of the residential building stock: the experience of TABULA project. Energy Pol. 68, 273–284.

Betterview, 2021. Betterview - understand and manage property risk. https://www.betterview.com/. (Accessed 23 May 2021).

Bowley, A.L., 1925. Measurement of the Precision Attained in Sampling. Cambridge University Press.

Breiman, L., 2001. Random Forests. Kluwer Academic Publishers, Boston, Mass.

Buratti, C., Barbanera, M., Palladino, D., 2014. An original tool for checking energy performance and certification of buildings by means of Artificial Neural Networks. Appl. Energy 120, 125–132.

Calì, D., Osterhage, T., Streblow, R., Müller, D., 2016. Energy performance gap in refurbished German dwellings: lesson learned from a field test. Energy Build. 127, 1146–1158.
Casals, X.G., 2006. Analysis of building energy regulation and certification in Europe: their role, limitations and differences. Energy Build. 38, 381–392.
Chae, Y.T., Horesh, R., Hwang, Y., Lee, Y.M., 2016. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy Build. 111, 184–194.
Chapman, J., 1991. Data accuracy and model reliability. BEPAC Conf. 10–19. Canterbury, UK.
Claesson, J., 2011. CERBOF Projekt No. 72: Utfall Och Metodutvärdering Av Energideklaration Av Byggnader. https://www.researchgate.net/publication/237005861_CERBOF_Projekt_no_72_Utfall_och_metodutvardering_av_energideklaration_av_byggnader. (Accessed 4 January 2022).
Cozza, S., Chambers, J., Deb, C., Scartezzini, J.-L., Schlüter, A., Patel, M.K., 2020a. Do energy performance certificates allow reliable predictions of actual energy consumption and savings? Learning from the Swiss national database. Energy Build. 224, 110235.
Cozza, S., Chambers, J., Patel, M.K., 2020b. Measuring the thermal energy performance gap of labelled residential buildings in Switzerland. Energy Pol. 137, 111085.
Crawley, J., Biddulph, P., Northrop, P.J., Wingfield, J., Oreszczyn, T., Elwell, C., 2019. Quantifying the measurement error on England and wales EPC ratings. Energies 12, 3523.
credium GmbH, 2021. credium - Building Data Insights. https://www.credium.de/. (Accessed 23 May 2021).
Droutsa, K.G., Kontoyiannidis, S., Dascalaki, E.G., Balaras, C.A., 2016. Mapping the energy performance of hellenic residential buildings from EPC (energy performance certificate) data. Energy 98, 284–295.
Enshen, L., Zixuan, Z., Xiaofei, M., 2005. Are the energy conservation rates (RVRs) approximate in different cities for the same building with the same outer-wall thermal insulation measures? Build. Environ. 40, 537–544.
European Commission, 2021. MAKING OUR HOMES AND BUILDINGS FIT FOR A GREENER FUTURE. https://ec.europa.eu/commission/presscorner/api/files/attachment/869476/Buildings_Factsheet_EN_final.pdf.pdf. (Accessed 27 July 2021).
European Parliament and of the Council, 2018. Directive (EU) 2018/2002 of the European Parliament and of the Council of 11 December 2018 Amending Directive 2010/31/EU on the Energy Performance of Buildings and Directive 2012/27/EU on Energy Efficiency.
Fabbri, K., Marinosci, C., 2018. EPBD independent control system for energy performance certification: the Emilia-Romagna Region (Italy) pioneering experience. Energy 165, 563–576.
Fan, C., Xiao, F., Wang, S., 2014. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Appl. Energy 127, 1–10.
Federal Ministry of Justice Germany, 2020. Gesetz zur Einsparung von Energie und zur Nutzung erneuerbarer Energien zur Wärme- und Kälteerzeugung in Gebäuden* (Gebäudeenergiegesetz - GEG). https://www.gesetze-im-internet.de/geg/BJNR172810020.html. (Accessed 19 October 2022).
Federal Statistical Office of Germany, 2011. Ergebnisse des Zensus 2011. Gebäude und Wohnungen sowie Wohnverhältnisse der Haushalte. https://ergebnisse.zensus2011.de/auswertungsdb/download?pdf=00&tableId=1&locale=DE&gmdblt=1. (Accessed 5 September 2019).
Federal Statistical Office of Germany, 2018. Gebäude und Wohnungen. Bestand an Wohnungen und Wohngebäuden - Bauabgang von Wohnungen und Wohngebäuden - Lange Reihen ab 1969 - 2017. https://www.destatis.de/EN/Press/2019/07/PE19_285_31231.html. (Accessed 17 March 2020).
Foucquier, A., Robert, S., Suard, F., Stéphan, L., Jay, A., 2013. State of the art in building modelling and energy performances prediction. Rev. Renew. Sustain. Energy Rev. 23, 272–288.
Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Kalogirou, S., Wolff, E., 2018. Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. GIScience Remote Sens. 55, 221–242.
German Energy Agency, 2016. dena-Gebäudereport. Statistiken und Analysen zur Energieeffizienz im Gebäudebestand. https://www.dena.de/fileadmin/user_upload/8162_dena-Gebaeudereport.pdf. (Accessed 19 October 2022).
German Energy Agency, 2018. Dena Concise Building Report. Energy Efficiency in the Building Stock – Statistics and Analyses.
German Federal Ministry for Economic Affairs and Energy (BMWi), 2018. Energieeffizienz in Zahlen. Entwicklungen und Trends in Deutschland 2018.
Grabolle, A., 2015. Energieausweis: Die Kosten auf einen Blick. https://www.co2online.de/modernisieren-und-bauen/energieausweis/energieausweis-kosten/. (Accessed 18 June 2019). Accessed.
Hardy, A., Glew, D., 2019. An analysis of errors in the Energy Performance certificate database. Energy Pol. 129, 1168–1178.
Iribar, E., Sellens, I., Angulo, L., Hidalgo, J.M., Sala, J.M., 2021. Nonconformities, deviation and improvements in the quality control of energy performance certificates in the Basque country. Sustain. Cities Soc. 75, 103286.
Johansson, P., Wahlgren, P., Dalenbäck, J.-O., 2016. Sweden| Differences between Measured and Calculated Energy Use in EPCs versus Building Permits. New Field Study/2016. http://qualicheck-platform.eu/wp-content/uploads/2017/02/QUALICHeCK-Filed-study-Sweden.pdf. (Accessed 9 May 2020).
Kaymakci, C., Wenninger, S., Sauer, A., 2021. A holistic framework for AI systems in industrial applications. 16. Internationale Tagung Wirtschaftsinformatik 2021.
Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H., 2018. Feature selection. ACM Comput. Surv. 50, 1–45.
Li, Y., Kubicki, S., Guerriero, A., Rezgui, Y., 2019. Review of building energy performance certification schemes towards future improvement. Renew. Sustain. Energy Rev. 113, 109244.
Likas, A., Vlassis, N., J. Verbeek, J., 2003. The global k-means clustering algorithm. Pattern Recogn. 36, 451–461.
Pasichnyi, O., Wallin, J., Levihn, F., Shahrokni, H., Kordas, O., 2019. Energy performance certificates — new opportunities for data-enabled urban energy policy instruments? Energy Pol. 127, 486–499.
Platten, J. von, Holmberg, C., Mangold, M., Johansson, T., Mjörnell, K., 2019. The renewing of Energy Performance Certificates—reaching comparability between decade-apart energy records. Appl. Energy 255, 113902.
Poel, B., van den Brink, L., 2009. Approaches and Possible Bottlenecks for Compliance and Control of EPBD Regula-Tions. Information Paper P178 of ASIEPI Project, European Commission.
Poel, B., van Cruchten, G., Balaras, C.A., 2007. Energy performance assessment of existing dwellings. Energy Build. 39, 393–403.
Reilly, A., Kinnane, O., 2017. The impact of thermal mass on building energy consumption. Appl. Energy 198, 108–121.
Seyedzadeh, S., Pour Rahimian, F., Oliver, S., Rodriguez, S., Glesk, I., 2020. Machine learning modelling for predicting non-domestic buildings energy performance: a model to support deep energy retrofit decision-making. Appl. Energy 279, 115908.
Strong, D.M., Lee, Y.W., Wang, R.Y., 1997. Data quality in context. Commun. ACM 40, 103–110.
Veiga, R.K., Veloso, A.C., Melo, A.P., Lamberts, R., 2021. Application of machine learning to estimate building energy use intensities. Energy Build. 249, 111219.
Walter, T., Sohn, M.D., 2016. A regression-based approach to estimating retrofit savings using the Building Performance Database. Appl. Energy 179, 996–1005.
Walter, T., Price, P.N., Sohn, M.D., 2014. Uncertainty estimation improves energy measurement and verification procedures. Appl. Energy 130, 230–236.
Wederhake, L., Wenninger, S., Wiethe, C., Fridgen, G., 2022. On the surplus accuracy of data-driven energy quantification methods in the residential sector. Energy Inf.
Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M., Zhao, X., 2018. A review of data-driven approaches for prediction and classification of building energy consumption. Renew. Sustain. Energy Rev. 82, 1027–1047.
Wenninger, S., Wiethe, C., 2021. Benchmarking Energy Quantification Methods to Predict Heating Energy Performance of Residential Buildings in Germany. Business & Information Systems Engineering.
Wenninger, S., Kaymakci, C., Wiethe, C., 2022a. Explainable long-term building energy consumption prediction using QLattice. Appl. Energy 308, 118300.
Wenninger, S., Kaymakci, C., Wiethe, C., Römmelt, J., Baur, L., Häckel, B., Sauer, A., 2022b. How Sustainable Is Machine Learning in Energy Applications? – the Sustainable Machine Learning Balance Sheet. 17th International Conference on Wirtschaftsinformatik. Nürnberg, Germany.
Wetterdienst, Deutscher, 2020. Wetter und Klima. CDC (Climate Data Center). https://www.dwd.de/EN/climate_environment/cdc/cdc_node.html. (Accessed 12 April 2020).
Yuan, P., Duanmu, L., Wang, Z., 2019. Coal consumption prediction model of space heating with feature selection for rural residences in severe cold area in China. Sustain. Cities Soc. 50, 101643.
Zhang, C., Cao, L., Romagnoli, A., 2018. On the feature engineering of building energy data mining. Sustain. Cities Soc. 39, 508–518.
Zhang, R., Indulska, M., Sadiq, S., 2019. Discovering data quality problems. Bus. Inf. Syst. Eng. 61, 575–593.
Zhao, H., Magoulès, F., 2012a. A review on the prediction of building energy consumption. Renew. Sustain. Energy Rev. 16, 3586–3592.
Zhao, H., Magoulès, F., 2012b. Feature selection for predicting building energy consumption based on statistical learning method. J. Algorithm Comput. Technol. 6, 59–77.
Zou, P.X., Xu, X., Sanjayan, J., Wang, J., 2018. Review of 10 years research on building energy performance gap: life-cycle and stakeholder perspectives. Energy Build. 178, 165–181.