Do energy performance certificates allow reliable predictions of actual energy consumption and savings? Learning from the Swiss national database

Cozza, Stefano; Chambers, Jonathan; Deb, Chirag; Scartezzini, Jean-Louis; Schlüter, Arno; Patel, Martin

How to cite

COZZA, Stefano et al. Do energy performance certificates allow reliable predictions of actual energy consumption and savings? Learning from the Swiss national database. In: Energy and Buildings, 2020, vol. 224, n° 110235. doi: 10.1016/j.enbuild.2020.110235

This publication URL: https://archive-ouverte.unige.ch/unige:140496
Publication DOI: 10.1016/j.enbuild.2020.110235

Energy & Buildings 224 (2020) 110235

Contents lists available at ScienceDirect

Energy & Buildings

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

Do energy performance certificates allow reliable predictions of actual energy consumption and savings? Learning from the Swiss national database

Stefano Cozza [a,*], Jonathan Chambers [a], Chirag Deb [b], Jean-Louis Scartezzini [c], Arno Schlüter [b], Martin K. Patel [a]

[a] Energy Efficiency Group, Institute for Environmental Sciences and Department F.-A. Forel for Environmental and Aquatic Sciences – University of Geneva, Switzerland
[b] Architecture and Building Systems, Institute of Technology in Architecture – ETH Zurich, Switzerland
[c] Solar Energy and Building Physics Laboratory, Ecole polytechnique fédérale de Lausanne - EPFL, Switzerland

ARTICLE INFO

Article history:
Received 31 January 2020
Revised 4 June 2020
Accepted 10 June 2020
Available online 17 June 2020

Keywords:
Building retrofit
Actual energy consumption
Performance gap
Energy Savings Deficit

ABSTRACT

The thermal performance gap in buildings is defined as the difference between the theoretical and the actual energy consumption for heating, and is known to undermine energy retrofit strategies and policies. This study examines the performance gap in retrofitted buildings using the Swiss Cantonal Energy Certificate for Buildings (CECB) database, using a sample of 1172 buildings for which both theoretical and actual metered consumption were known. We found an average negative performance gap of –23% for pre-retrofit buildings (actual consumption smaller than calculated) and instead a good approximation of actual consumption with theoretical consumption after retrofitting (a positive gap of 2%). A regression analysis on the energy performance certificate input parameters characterizing the building led to the conclusion that these are poor predictors of actual consumption compared to the theoretical calculation: parameters such as the energy label and the thermal proprieties of the envelope (U-values) have minor explanatory power for the actual consumption despite explaining a high degree of change in the theoretical consumption. Analysis of the indicator Energy Savings Deficit (ESD) shows an overestimation (of 37%) of the achievable savings on the basis of the theoretical consumption, whereas the prediction of savings using measured consumption before retrofit resulted in a good agreement with the actual savings (3.6% overestimation). This implies that energy savings can be estimated rather accurately by comparing the actual current consumption with the expected theoretical consumption defined by the certificate after retrofit.

1. Introduction

1.1. Background

The thermal performance gap is defined as the difference between the measured and the calculated energy consumption for heating and domestic hot water of a building [1,2]. It has been observed in several countries that buildings with lower thermal performance tend to consume less energy than expected [3–6], while buildings of high thermal performance tend to consume more energy than expected [5,7,8].

A range of studies have been published in recent years focusing on the performance gap and its consequences for meeting the national energy saving goals in UK [9], France [10], Germany [11], Netherlands [12], Luxemburg [8], and Belgium [4]. Most of these studies are based on statistical analysis of large national databases, comparing different sources. Majcen et al. [5] analysed 340 000 performance certificates for Dutch residential buildings provided by the Agentschap-NL [13] combined with data from the CBS Statistics Netherlands and energy companies. Delghust et al. [4] used the Belgian national energy performance certificate database as well as surveys on occupation and user behaviour in households. In France, Cayre et al. [3] used the performance certificates provided by the CEREN national database for 923 residential buildings. All these studies are in agreement with regard to the finding that standard calculations based on the building model strongly overestimate space heating consumption in older housing

* Corresponding author at: Energy Efficiency Group, Institute for Environmental Sciences and Department F.-A. Forel for Environmental and Aquatic Sciences – University of Geneva, Switzerland.
E-mail address: stefano.cozza@unige.ch (S. Cozza).

https://doi.org/10.1016/j.enbuild.2020.110235
0378-7788/© 2020 The Authors. Published by Elsevier B.V.

Nomenclature

CECB	Swiss Cantonal Energy Certificate for Buildings
ERA	Energy reference area [m^2]
ESD	Energy savings deficit [%]
MDB	Multi-dwelling building
SDB	Single dwelling building
SIA	Swiss Society of Engineers and Architects
VIF	variance-inflation factor

as well as the related energy savings potential. This phenomenon is well known and has been described by some authors as the "rebound effect" [14,15] and "prebound effect" [11,16,17] to explain the performance gap [18,19]. The rebound effect occurs when energy-efficiency improvements reduce the marginal cost of energy services, thus encouraging the increased consumption of services, which may offset some or all of the reduction in energy consumption [17]. The prebound effect is defined as the under-consumption of energy services in old, inefficient dwellings prior to or in the absence of energy retrofit [11,16–18] also leading to lower than expected energy savings potential.

However, the use of the term rebound (and by extension prebound) can be misleading [11] as rebound is connected to a change (direct or indirect) of consumer behaviour [16,20]. Existing research indicates that a large share of the difference between theoretical and actual consumption has causes not linked to a change in building occupant behaviour. These include limitations of the building physics model as given by norms [21,22], the quality of the data collection used to implement the models [23,24], and basic physical effects [25]. Indeed, Love investigated how occupants change their heating behaviour in response to retrofit, and found that "most of the change in internal temperature occurred during unheated hours, and thus was a result of the change in the thermal efficiency of the building fabric, not of occupant behaviour" [26]. Therefore, in this work we exclusively use the terms performance gap before and after retrofit to refer to the difference between theoretical and actual consumption, in order to avoid any implicit assumptions on the causes of these differences.

Although the existence of the energy performance gap is widely acknowledged, it has been difficult to identify the key factors that influence it. According to recent literature, the main causes of this gap are inaccurate knowledge of input variables defining the building physics, uncertainty in weather variables, and inability to model occupant behaviour (because the way the inhabitants use the buildings is often different from standards and expectations) [25,27–29]. The three main inaccuracies that have been identified are 1) the expected indoor air temperature [30–35], 2) the U-values assumed for building façade elements [9,31,36–38], and 3) the ventilation air change rate [4,34,39–42]. Other factors that influence the building's performance calculation are the standard assumptions for the air tightness of the building envelope, the default assumptions concerning the space heating system (e.g. too high efficiency) [4], the thermal inertia of massive walls [43], and the parameters for modelling solar gains and the effect of shading [44]. In addition, problems may exist with the building itself such as malfunctioning of the heating system or its deficient execution (faulty installation or suboptimal design), or deficiencies in other building elements (deviation from design specs)[25]. This results in variable energy consumption over time and accordingly a variable performance gap [2,16].

The performance gap gives rise to uncertainty when property owners take decisions about retrofit investments [45,46]. When there is uncertainty in the calculation of the current dwelling energy consumption, this leads to uncertainty in energy saving potentials [10]. Notably, if less energy is consumed than calculated before retrofit, there is less potential for savings, with "implications for the economic viability of thermal retrofits [2]". As a consequence, another gap arises [47]: the gap between actual and theoretical energy savings after retrofit [2]. In this work, this is referred to as the "energy savings deficit" (ESD), as proposed by Galvin [16] (previously introduced by Haas and Biermayr as the unachieved energy conservation share or "rebound share" [7]). The ESD, which is defined as the shortfall in savings after an energy retrofit as a proportion of the expected savings, has been demonstrated in other studies [12,16,48]. All indicate that on average, the majority of thermal retrofits result in lower energy savings than calculated [47]. Furthermore, the problem is greater when energy performance certificates are used to calculate energy savings [49,50]. European studies using energy performance certificates on a large samples of buildings found that the ratio of actual to theoretical energy savings after retrofit ranged from 40% to 60% [48,51], with consequences for the energy reduction targets that can differ significantly from country to country [18].

Works have been published for several countries, including the Netherlands [47], France [3], Ireland [52], and UK [26], providing valuable insights about national energy policies on retrofit. This paper builds on these studies by contributing results for Switzerland. It is currently unclear whether and to what extent these findings can be applied to the Swiss building stock, as research so far has focused on a few case studies or small samples [36,37,53–55], since no large-scale dataset about the retrofitted buildings' actual energy consumption was previously available for Switzerland. It is important to clarify the relationship between the expected savings resulting from the improvement in the energy labels of the performance certificates, and the savings that are actually achieved. It is crucial for the success of the energy policies to know the true energy savings potential of the building stock when both the performance gap and the ESD are taking into account.

1.2. Contribution

This study, covering more than 1000 retrofitted buildings, is the most comprehensive so far conducted in Switzerland and it fills this data gap by analysing actual energy savings of dwellings at the stock level. This paper aims to advance understanding of the issues that arise in linking energy savings from retrofit and energy ratings based on the Swiss Cantonal Energy Certificate for Buildings (CECB) [56] with actual data on buildings' energy consumption. The CECB database, containing all residential buildings in Switzerland that have improved their energy rating, was used here for the first time to assess actual energy savings by comparing the actual consumption of buildings before and after retrofit.

We apply descriptive statistics and regression analysis to a large database of energy certificates in order to quantify the performance gap and to identify which buildings' parameters (e.g. U-values, HVAC systems) can be used to better assess actual consumption before retrofit and therefore actual energy savings after retrofitting.

An established definition of the energy savings deficit (ESD) [7,15,16] was used to investigate the effect of the performance gap on the difference between expected and achieved energy savings post-retrofit. Furthermore, we introduce a new distinction in the definition ESD to produce two different indicators, one based on theoretical values before retrofit and one based on actual values (see Section 2.3). The reason for this is to capture different perspectives on energy savings, showing on the one hand the systematic error that is made in estimating energy savings from theoretical values, and on the other hand demonstrating an alternative approach using actual values pre-retrofit.

2. Methods and datasets

2.1. Performance gap

The energy performance gap is defined in this work as the difference between the actual (billed) energy consumption and the theoretical energy consumption according to regulations assuming operating conditions as set out in the national standards [28]. To calculate it, Eq. 1 as proposed by Galvin [16] is applied by using the values for the actual and theoretical energy consumption given by the CECB, both expressed in kWh/(m^{2}y).

$$Performance\ gap\,[\%] = \frac{Actual\ consumption - Theoretical\ consumption}{Theoretical\ consumption} \times 100 \tag{1}$$

Trends in the distribution of performance gap values and energy consumption values as a function of the energy label were analysed.

2.2. Regression analysis

A regression analysis was performed to evaluate the main predictors of theoretical and actual energy consumption and to understand how much variation of the consumption can be explained using thermal performance parameters included in the CECB (described below in the Section 2.4). The analysis was carried out on the buildings pre-retrofit, in order to understand whether the existing certificates are reliable tools to determine the actual consumption and the actual energy savings to be obtained through retrofitting.

A preliminary analysis of the multicollinearity of the variables was performed to assess if there was a strong correlation between two or more variables in the regression model. There is no formal cut-off point for critical values of variance-inflation factors (VIF); in this paper a conservative value of 2.5 was used, as suggested in literature [57–59].

Afterwards, a regularization method was applied to select a subset of variables that explain most of the variability in the final energy consumption. The Lasso (Least Absolute Shrinkage and Selection Operator [60]) regularization method was used as it effectively performs variable selection by using a fitting procedure which aims to set some coefficients to zero, making the model sparse [59]. Lasso regression applies a 'L1-penalty' term, which is proportional to the sum of the absolute coefficients, with the effect of zeroing the coefficients with low explanatory power. Categorical variables (e.g. HVAC type) were converted to dummy variables using one-hot encoding [61]. After identifying which coefficients were set to zero using Lasso, a linear ordinary least squares regression analysis was repeated omitting those variables.

2.3. Energy savings deficit

The Energy Savings Deficit is the shortfall in energy savings related to energy retrofitting. It is expressed as percentage of expected energy savings. However, in their original formulation, Haas and Biermayr [7] did not specify whether the "expected savings" are based on the actual or on the theoretical consumption of a building before a retrofit. In this work we therefore compare two definitions for expected savings, namely "theoretical savings" and "anticipated savings". The "theoretical" savings are defined as the difference between theoretical consumption in kWh/(m^{2}y) before and after the energy retrofitting:

$$Theoretical\ savings = Theoretical\ consumption\ before - Theoretical\ consumption\ after \tag{2}$$

The "anticipated saving" in kWh/(m^{2}y) are defined as the savings that one could expect before a retrofit by comparing known current (actual) consumption (e.g. measured with an energy meter or energy bills) with the theoretical consumption given by the target energy label post-retrofit.

$$Anticipated\ savings = Actual\ consumption\ before - Theoretical\ consumption\ after \tag{3}$$

In these equations (Eqs. (2) and (3)), the consumption is always the final energy consumption for thermal use including heating and hot water, expressed in kWh/(m^{2}y). A summary of all the presented variables is reported in Fig. 1.

The savings defined in Fig. 1 are used to calculate two variants of the ESD. The first is the Energy Savings Deficit Regulatory (ESDr), so named because the calculated savings are obtained exclusively by using the energy performance evaluation of the building based on national regulations:

$$ESDr[\%] = \frac{Theoretical\ savings - Actual\ savings}{Theoretical\ savings} \times 100 \tag{4}$$

The second is the Energy Savings Deficit Anticipated (ESDa), which is based on the difference between the anticipated savings defined above and the actual savings:

$$ESDa[\%] = \frac{Anticipated\ savings - Actual\ savings}{Anticipated\ savings} \times 100 \tag{5}$$

The ESDr is a useful indicator for policy makers, as it gives a clear understanding of how far the theoretical savings are away from the real ones. An example is the energy retrofit that leads to an increase in the rating of the energy certificate of a building. In this case, the theoretical savings are calculated as a function of the number of label improvements gained through retrofit (e.g. Label improvement equal to 2 for an improvement of the energy class from C to A), and they can be used to set the targets for national or cantonal renovation strategies.

The ESDa, besides being useful for policy makers who may want to have a more reliable picture of energy savings in retrofitting, it is a useful indicator for the building owner who knows its own actual energy consumption before retrofit and can use this information to estimate the energy savings to be obtained from a retrofit. The building owner can consequently also assess the cost-effectiveness of the operation. The ESDa gives also a first indication of possible problems related to the functioning of the equipment and/or to user behaviour.

2.4. CECB dataset

The Cantonal Energy Certificate for Building (CECB) was introduced in 2009 [56] in Switzerland in the wake of the European Energy Performance of Buildings Directive [62]. It was designed to be used by certified experts to provide a quick and cost-effective way of estimating the energy performance of existing buildings, independently of the occupant's behaviour. Thanks to a simple energy scale, the efficiency of the building envelope, as well as the efficiency of final energy use can be visualized with an

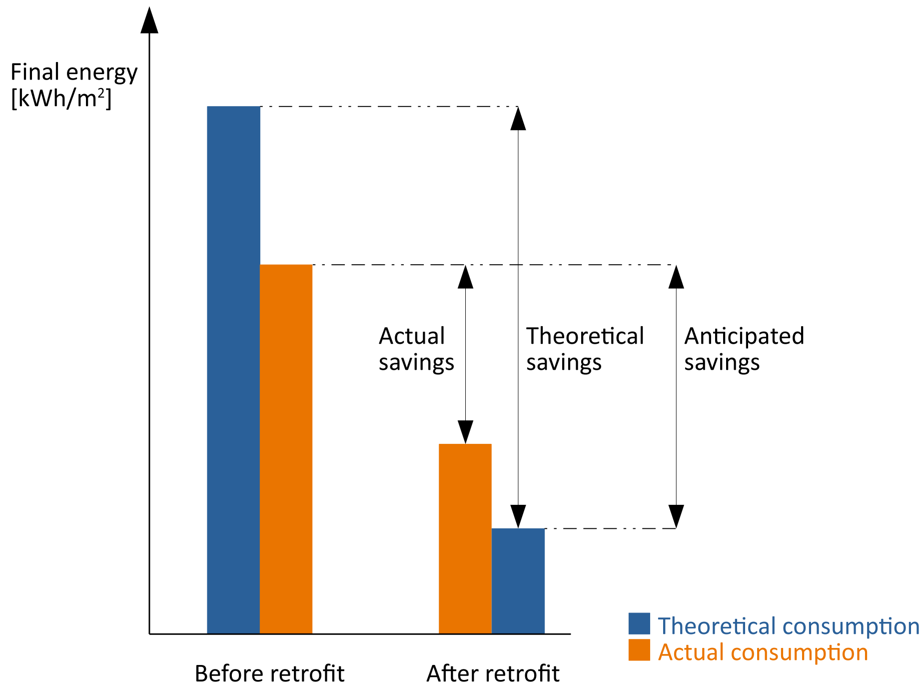

Fig. 1. Differences between theoretical and actual energy consumption, before and after energy retrofit.

energy label between A and G (very efficient to very inefficient) [63]. This rating is performed based on final energy weighted with factors [64] that reflect the desirability of low carbon energy sources (e.g. biomass has a low weighting), thereby favouring renewable energy sources and/or heat pumps. A side effect of this weighting is to cause a certain disconnection between the final energy consumption and the resulting label, as for example a relatively high consumption of a 'favoured' low carbon fuel such as biomass would still result in a good energy label.

In a previous publication [65], the database, containing more than 50 000 buildings, was compared to the national building register, published by the Swiss Federal Statics Office [66], to check the representativeness of the CECB sample. It was found that despite some differences in representation between cantons, the CECB dataset was representative of the Swiss residential building stock [67,68].

The CECB provides two key energy indicators per certificate: the actual and the theoretical consumption of the building, with a breakdown of final energy into space heating and domestic hot water. The theoretical consumption represents the energy use of the building under standard conditions of use (Appendix A) and standard weather conditions (SIA 2028) [69]. The theoretical heat balance calculations follow the Swiss norm SIA 380/1, based on the static monthly balance in line with the European SN EN 13790. In contrast, the actual consumption is obtained from energy bills for the energy carriers used in the building, after climate correction following the methodology described in SIA 2031 [63]. Actual energy consumption is determined as the average of measured energy use over at least three consecutive years. The actual consumption is used in the CECB calculation tool to check the building model adopted for the theoretical consumption. This work considers only final energy consumption (in kWh/(m^{2}y)) for thermal use (space heating and domestic hot water).

The CECB dataset includes general building metadata (location, closest climate station, construction year), geometry (dimensions, energy reference area (ERA), orientation), envelope quality information (areas, U-values of each element, g-values and shade coefficients for windows) as well as the type of heating system (fossil, heat pump, solar thermal). These have been used as input in the regression analysis, to investigate their predictive power for the building's theoretical and actual energy consumption (the building parameters used in the analysis are reported in Table 3).

The CECB dataset did not include explicit information on retrofit occurrences. However, a large number of buildings had multiple CECB certificates. It was assumed that buildings with multiple certificates, where the certificates had different energy labels, and where the newer energy label was better than the old one, had been retrofitted. A subsample of the CECB dataset was produced which consists of residential buildings – single dwelling buildings (SDB) and multi-dwelling buildings (MDB) – that have a CECB certification both before and after retrofit and which included actual (billed) energy consumption for both. Of the 14076 buildings with a multiple CECB versions, 10178 included actual energy consumption values. However, for many of these the CECB label was not changing and we assumed that no retrofit had been performed. These cases were therefore discarded, leaving 2277 buildings in the sample. It was found that the CECB released after the retrofit often reported the identical actual energy consumption declared in the CECB before (in spite of changing labels), indicating that new actual consumption data were not recorded but directly copied from the previous certificate. These cases were deleted, leaving 1756 buildings. It was found that for some buildings the second energy label was worse than the first one (for no identified reason). Although these occurrences deserve further analysis (e.g. to determine whether a retrofit had been performed at all), this was determined to be beyond the scope of the current study and they were deleted from the sample, leaving 1575 buildings. Finally, the outliers in the energy consumption were removed. Outliers in this case are buildings with energy consumption values which are extremely different from other buildings and were hence considered to be erroneous data or highly atypical buildings. To identify them, the modified Z-scores method [70] was used with a score threshold of 3.5. This reduced the sample to the final size of 1172 residential buildings (MDB + SDB) with certification before and after the retrofit. The implications of this data cleaning process are discussed in Section 3.4.

3. Results and discussion

3.1. Quantification of the performance gap in retrofitted buildings

Fig. 2 presents the sample of 1172 buildings before and after the retrofit, showing that before retrofit a large majority of buildings have a label equal to or lower than D-label, while after retrofit they become equal or higher than D-label (for more details on the depth of the retrofit please see Appendix B). It must be highlighted that this sample, being a relatively small subsample of the CECB database, cannot be guaranteed to be representative of the building stock as a whole. However, it constitutes an important case study on the characteristics of retrofitted buildings.

Table 1 reports the final energy consumption per square meter for the total subsample. This highlights the significant improvement achieved at the level of the individual building. The median value of actual consumption per square meter before energy retrofit according to the CECB subsample equals 146 kWh/(m^{2}y) which is very close to the reported total national average for space heating (with climate correction) and domestic hot water according to the Odyssee database (144 kWh/(m^{2}y) [71]). The median value of actual consumption after energy retrofit is 59 kWh/(m^{2}y) which is equal to the target indicated in the Swiss norms for retrofitted buildings (60 kWh/(m^{2}y) [72]). The fact that the median building in our sample does meet the target implies that it is possible and realistic to achieve the set objectives.

Fig. 3 shows the theoretical and actual final energy consumption per meter square per energy label before and after retrofit. As can be seen, the actual consumption in F and G rated buildings before retrofit is much less than the theoretical one. Furthermore these buildings are the most affected by retrofit, as shown in Fig. 2.

It is important to note the minimum and maximum values in Fig. 3, which show that the actual final energy consumption of some buildings is actually very different than expected from their label, resulting in a strong overlap in energy consumption across the ratings (the reason was explained in the first paragraph of Section 2.4). This means that using only the energy rating as proxy for actual energy consumption entails the risk of incorrect assessment. Such a rough approach should not be used to evaluate the actual performance of a building prior to retrofit.

The large size of the boxes in Fig. 3 and their overlap across labels raises questions about the significance of the difference between actual and theoretical consumption. Therefore, these observations have been tested using a Wilcoxon signed-rank test [73]. This non-parametric test is used to compare two sets of scores from the same sample, in this case buildings. Moreover, given the different sample size of each energy label, the effect size (i.e. a standardized measure of the magnitude of an observed effect [61]) was also calculated. As Table 2 shows, the test yields a medium to large effect size for buildings labelled E, F, and G indicating that these findings are statistically significant.

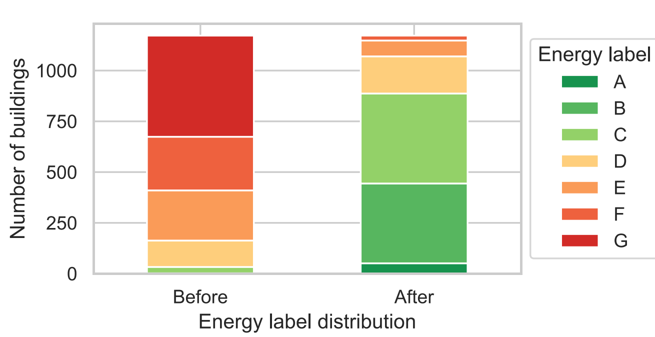

Fig. 2. Energy label distribution before and after retrofit.

Table 1
Final energy consumptions of the CECB sub-sample, before and after retrofit (1172 buildings).

[kWh/(m^{2}y)]	Theoretical consumption		Actual consumption	
	Before	After	Before	After
Median	199	53.2	146	59.3
Minimum	15.6	6.39	14.7	4.24
Maximum	478	264	321	226
Mean	216	65.8	153	66.3
Std.	95.2	49.0	60.8	43.5

Table 2 also reports the performance gap calculated with Equation 1 (a negative gap implies that the actual consumption is smaller than the theoretical consumption). The results show that buildings with a low thermal performance before retrofit have a negative performance gap. Vice versa, buildings with higher performance (B and C) after retrofit consume marginally more than predicted (positive performance gap). This result supports previous finding in the literature [4–6,10].

The performance gap in retrofitted buildings does not substantially differ from that of un-retrofitted buildings (i.e. the gap for D-label before retrofit is equal to -5.8% and after is equal to -4.1%), however the median performance gap across all labels changes from -23% before retrofit to +2.1% afterwards – this implies that after retrofitting, theoretical consumption is a good proxy for actual consumption. Therefore, the consequence of retrofitting buildings with a negative gap (E, F, and G before) to buildings with a somewhat positive gap (B and C after) is that lower savings will be achieved than expected.

Finally, care should be taken when considering the percentage values for the performance gap for different energy labels shown in Table 2, as one percent point represents very different energy consumption values in terms of kWh/(m^{2}y) for the low rating compared to the high rating buildings. For example, prior to retrofit, a gap of -23% in label F corresponds to a difference between actual and theoretical energy consumption of -52 kWh/(m^{2}y) (Fig. 3a), while in label B a similar gap (with inverse sign) of +17% corresponds to an absolute difference of only +6 kWh/(m^{2}y) (Fig. 3a).

3.2. Analysis of the performance gap

3.2.1. Regression analysis of the energy consumptions

This section presents the results of the regression analysis. The goal of the regression analysis is to compare how much of the variation of the theoretical and actual energy consumption can be explained by the technical variables of the building that are recorded in its CECB certificate. This gives a measure of the extent to which the variables collected are predictors of the actual consumption. When there is significant deviation between the theoretical and actual values, this can indicate that the coefficients of the variables collected are incorrect or that other variables that are not captured in the certificate (such as user behaviour) have a large influence.

The regression model uses as independent variables the parameters from the CECB that are the basis of the theoretical consumption calculation (see Section 2.4), and as dependent variables the theoretical and the actual energy consumption of the building. Multicollinearity among the variables was not found to be an issue, as the VIF of all the variables used is below the threshold value of 2.5 (the complete VIF results are reported in Appendix C, Table C.1.)

The Lasso regression selected a subset of 18 variables; eliminating 8 variables (setting their coefficients to zero). For details, see Appendix D. The ordinary least squares model was run omitting these eight variables. In Table 3, results of the regression are presented, with the variables set to zero (i.e. not used in the second analysis) and the coefficient of the selected variables.

The final model for the theoretical consumption developed using the 18 variables resulted in a coefficient of variation R^2 of 70%, implying that as expected the variation is well explained by the CECB variables. However, for the actual consumption only 40% of response variation was explained. This implies that only a fraction of the building's actual consumption can be predicted based on the parameters in the certificate, and therefore most of the variation in actual consumption can be attributed either to parameters that are not listed among the variables of the energy certificate (e.g. internal temperature, ventilation rate), or the values of the parameter listed are inaccurate. Even though the regression model is obtained on a limited sample (1172 buildings), the adjusted R^2 values (69.9% and 39.7%) are very similar to the observed values of R^2 indicating that the cross-validity of the model (i.e. how well the model generalizes) is very good [61]. More details on the cross-validation of the model can be found in Appendix D.

The *B*-values indicate the relationship between the energy consumption and the independent variables, and they are a function of the unit in which the independent variable is measured and are not intercomparable. The β-values are reported in terms of standard deviation and are therefore intercomparable.

The comparison of the β-values of the same independent variable (e.g. U-value walls) for the theoretical and actual consumption allows us to understand whether the variable has the same predictive power for both types of consumption. We would expect that if the independent variables as reported are a good representation of reality, they would have similar predictive power for both theoretical and actual consumption. If this is not the case – i.e. the β-value of a variable for actual consumption is smaller than the one for theoretical consumption – the variable has a smaller predictive power for the actual consumption than the theoretical consumption. Whenever a variable has different β-values for actual and theoretical consumption, that variable contributes to increasing the gap between the final actual and theoretical consumption.

As expected, the energy label is a very strong predictor for energy consumption (high B-value). The label is properly accounted for in the regression analysis (i.e. the higher the label the lower the energy consumption), except for label B that is very poorly represented in the sub-sample of the building before retrofit. However, the β-values for the actual consumption are smaller than those for the theoretical (in bold in Table 3), indicating that using only the energy label as predictor of actual energy consumption is subject to high uncertainty. More importantly, label E was not selected by Lasso, indicating the low predictive power of this variable for the final energy consumption. This means that a building with label E before retrofit has a very large range of actual con-

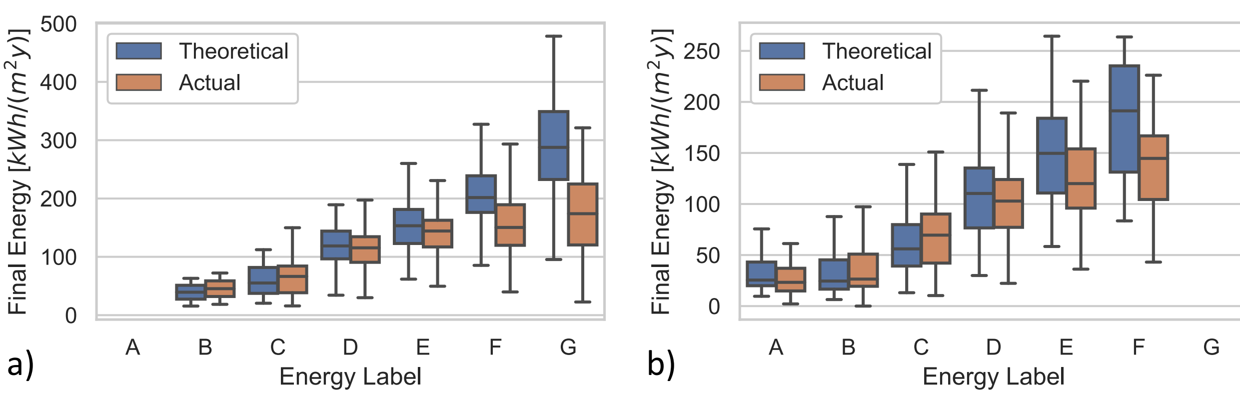

Fig. 3. Theoretical and actual consumption per energy label before (a) and after retrofit (b). The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the edges of box to indicate the variability outside the interquartile range (IQR = upper – lower). The position of the whiskers (i.e. minimum and maximum) is set to 1.5 * IQR from the edges of the box. Outliers are not reported (this also applies to the other box-plots presented in this manuscript).

Table 2
Performance gap and Wilcoxon signed-rank test (z) per energy label, before and after retrofit.

Energy Label	Retrofit	Sample size	Performance Gap [%]	*z*	Effect size
A	Before	–	–	–	–
	After	51	−7.21	−2.58**	−0.255
B	Before	2	16.8	1.34*	0.671
	After	393	9.86	5.98***	0.213
C	Before	31	−0.29	−0.02*	−0.003
	After	443	5.95	5.47***	0.184
D	Before	130	−5.83	−1.54**	−0.096
	After	183	−4.08	−2.21**	−0.115
E	Before	247	−9.21	−5.06***	−0.228
	After	78	−16.2	−4.69***	−0.375
F	Before	264	−23.4	−11.23***	−0.491
	After	24	−22.4	−3.69***	−0.532
G	Before	498	−37.3	−18.6***	−0.591
	After	–	–	–	–
All	**Before**	**1172**	**−22.9**	**−22.7*** **	**−0.469**
	After	**1172**	**2.11**	**2.07** **	**0.043**

*p < 0.8, **p < 0.05, ***p < 0.001.

Table 3
Results of the regression analysis for the determinants of theoretical and actual pre-retrofit energy consumption using the parameters recorded in the CECB as independent variables. The most significant results (i.e. the largest differences between β-values for the same variable) have been highlighted in bold.

Independent variables		Final energy consumption [kWh/m²]					
		Theoretical [R^2 = 70.4%]			Actual [R^2 = 40.5%]		
	Dummies	B	β	Sign.	B	β	Sign.
Constant		1082.89		0	888.526		0
Building type	SDB – MDB	2.397	0.012	0.564	−15.957	−0.129	0.000
Construction year	Ratio Variable	0.015	0.010	0.586	0.015	0.015	0.555
ERA	Ratio Variable	–	–	–	–	–	–
Envelope factor	Ratio Variable	81.788	0.404	0.000	42.036	0.326	0.000
Energy Label [dummy variable is G label]	B	−22.203	−0.020	0.254	−28.054	−0.040	0.110
	C	−44.668	−0.101	0.000	−47.763	−0.170	0.000
	D	−32.844	−**0.217**	0.000	−17.820	−**0.100**	0.000
	E	–	–	–	–	–	–
	F	−15.993	−**0.169**	0.000	−3.282	−**0.022**	0.371
Mechanical ventilation	Yes – no	–	–	–	–	–	–
Heating System [dummy variable is Oil boiler]	DH*	–	–	–	–	–	–
	Electro	−49.588	−0.289	0.000	−49.897	−0.298	0.000
	Gas	–	–	–	–	–	–
	HP*	−117.094	−0.205	0.000	−69.406	−0.191	0.000
	Wood	25.148	0.057	0.002	1.928	0.007	0.789
Heating System construction year	Ratio Variable	−0.557	−0.109	0.000	−0.425	−0.130	0.000
U-value ground	Ratio Variable	–	–	–	–	–	–
U-value roof/ceiling	Ratio Variable	37.016	**0.246**	0.000	11.343	**0.118**	0.000
U-value external walls	Ratio Variable	80.096	**0.355**	0.000	18.317	**0.127**	0.000
U-value windows	Ratio Variable	10.571	0.098	0.000	7.875	0.112	0.000
Construction Type [dummy variable is Heavy construction]	Medium	–	–	–	–	–	–
	Light	15.198	0.042	0.015	−14.629	−0.064	0.010
	Very light	–	–	–	–	–	–

* DH: district heating; HP: heat pump.

sumptions. This result is in line with previous findings in the literature that clearly showed that using building's energy label before retrofit to predict energy consumption tends to overestimate the actual consumption [2,19,50].

It is interesting to note that buildings using heat pumps or an electric heater to provide space heating are expected to have lower energy consumption (negative B-values) in both actual and theoretical terms when oil-fired boiler is used as the reference dummy variable. However, their β-values are very similar, meaning that they have similar weight in predicting both actual and theoretical consumption, and therefore the type of heating system does not seem to contribute to the performance gap.

Other strong predictors for the theoretical energy consumption are the U-values of the envelope (roof and external walls). In this case too, the β-values for the actual consumption are much smaller than those for the theoretical. A weak correlation with the actual consumption suggests that the U-values reported in the CECB (usually standard U-values, function of the building type, the construction period, and the general condition of the façade) prior to retrofit have a high level of uncertainty. This finding that incorrect standard insulation values used in the energy certificates contribute to the performance gap supports previous findings in Europe [12,18]. This also supports findings for Switzerland where measurement studies on wall U-values confirmed that standard values used for certificates exceeded the measured ones by 10% to 200% [36,55].

Conversely, the predictive power of the windows U-value is very similar for both actual and theoretical consumption, which suggests that this variable as reported in the CECB is more reliable. This could be a reflection of the higher level of detail that is requested to describe the windows in the CECB, including glazing and frame types, dimensions, facade proportions and general condition. This could finally suggest that a more detailed description of the building element would allow to choose more appropriate standard values which would then result in a better agreement of the calculated U-value with the actual U-value.

For completeness, we replicated the same regression analysis on the buildings post-retrofit, using the variables of the certificates issued after the retrofitting. The result again demonstrated a higher degree of predictive power of the CECB variables for the theoretical consumption compared to the actual consumption. However, in this case there was a smaller distance between the R^2 values. These findings were consistent with the findings on the performance gap post-retrofit presented in Section 3.1, and are presented in Appendix C, Table C.2.

3.2.2. Regression analysis of the performance gap

Additionally, a regression analysis was performed with the performance gap as dependent variable, in order to understand if the variables reported in the CECB were predictors of the gap itself. However, this analysis did not yield statistically significant results. Only 27% of the variation of the performance gap was explained by the parameters in the certificate. This is because the actual consumption is not strongly correlated with the CECB parameters, and therefore the difference between actual and theoretical consumption cannot be correlated with the CECB parameters.

Considering the single independent variables, none of them, taken alone, have sufficient predictive power to accurately predict the gap (e.g. it is not possible to indicate that buildings with gas heating system have a larger pre-retrofit performance gap than buildings equipped with a heat pump). The summary of the complete results is presented in Appendix C, Table C.1.

This result supports previous finding in the literature [2,25,47,74] indicating that the performance gap (as well as the energy performance of the building itself) is the result of complex interactions between many factors. Nevertheless, these results indicated that the correct definition of the U-values of the envelope can have a large influence on the performance gap, but it has not been possible to reliably quantify the magnitude of the impacts of each variable due to their deep intercorrelation and due to the fact that many variables were not included in the certificate (e.g. user behaviour).

3.3. Consequences for the performance gap

Finally, the consequences of incorrect interpretation and misuse of certificates when retrofitting have been investigated. Further calculations have been performed to explore the difference between theoretical savings (the difference between the two theoretical energy consumptions), anticipated savings (the difference between current real energy consumption and expected energy consumption of the retrofitted building), and actual savings (difference between the measured energy consumption before and after retrofit), using Eqs. (2) and (3) as reported in Fig. 1.

As already proven in previous studies [2,7], the clearest relationships were found when considering only the depth of the retrofit (Label improvement, as presented in Fig. 4), rather than the energy label either before or after the retrofit.

Fig. 4 shows that the energy savings in absolute values, expressed in kWh/(m^2y), are increasing with the depth of the renovation. It is interesting to note that the theoretical savings grow steadily (as expected) as the Label improvement increases, and much more strongly than the actual savings. Actual savings are up to around half of the theoretical ones for Label improvements 3 to 6. On the contrary, anticipated and actual values are always very close, implying that anticipated savings are a good proxy for the actual savings.

The boxplot in Fig. 4 shows that as expected no single theoretical value is negative (i.e. the retrofits considered always aimed to save energy when improving the energy rating, due to the filtering operation explained in Section 2.4). Nevertheless, actual savings show some negative values, indicating that some buildings consume more final energy per square meter after the retrofit, even if their energy rating is improving. While this calls for further investigation it is beyond the scope of this paper.

The fact that anticipated savings are good predictors of actual savings is even more evident when comparing the distribution of values for each building within the sub-sample. As shown by the slope of the regression lines in Fig. 5, the anticipated savings correlate much more with the actual savings ($r = 0.92$, Fig. 5b) than with the theoretical savings do ($r = 0.64$, Fig. 5a).

All three types of savings are used to calculate the Energy Savings Deficit regulatory (ESDr, Eq. (4)) and the Energy Savings Deficit anticipated (ESDa, Eq. (5)), resulting in a median ESDr of 37.3% and an ESDa of 3.60%. Based on the theoretical values, only 62.7% (100% – 37.3%) of the expected savings are actually achieved. Instead, when using the anticipated values, 96.4% (100% – 3.60%) of the expected savings are achieved.

This finding on the one hand is encouraging (i.e. the small ESDa) because it implies that after retrofit the actual consumption obtained is in line with the theoretical value, and as shown in Table 1, the median actual consumption after retrofit is in accordance with the target indicated in the Swiss norms for retrofitted buildings (60 kWh/(m^2y) [72]). On the other hand, such a high difference before retrofit between the theoretical and actual consumption (i.e. the large ESDr) shows that the theoretical savings available from the building stock cannot be considered accurate with the current calculation models. This raises concerns about setting targets on the theoretical values. The Swiss Energy Strategy 2050 [75] includes a target for residential buildings to reduce their final energy consumption for space heating and domestic hot water from 58.8 TWh/y to 20.8 TWh/y, implying a reduction of 65% by 2050 compared to current levels [76]. This reduction is expected to be achieved also through building retrofit, with the efficiency target based on energy labels. However, due to the large ESDr the energy saved would be significantly smaller than expected. Therefore, the current target for retrofit needs to be revised. To illustrate: the current theoretical consumption of pre-retrofit buildings is 199 kWh/(m^2y) (Table 1). A reduction of 65% gives a theoretical consumption of 70 kWh/(m^2y). On this basis, Swiss norms set a retrofit target performance of 60 kWh/(m^2y) [72]. However, the actual consumption before retrofit is 146 kWh/(m^2y) (Table 1) so to achieve the same 65% saving requires a post-retrofit performance of 50 kWh/(m^2y), below the current requirement. This example highlights the problem of using theoretical values.

Finally, both ESDr and ESDa were calculated as a function of the energy labels steps improved through retrofit (Label improvement). In Fig. 6 the expected savings are set equal to 100%. The ESDr and the ESDa represent the share of the achieved savings.

Fig. 6a shows that the ESDr increases with the Label improvement, meaning that the achieved energy savings obtained through a deep retrofit (Label improvement 5 and 6) are about half of the theoretical savings, as also observed in other studies [12,49,77,78]. The Label improvement (reflecting the number of energy savings measures) increases the gap between achieved and theoretical savings. A similar observation has been made for the Netherlands [48].

Fig. 6b shows the opposite trend for ESDa. The ESDa indeed decreases with the increase of the Label improvement, until it becomes even negative for Label improvement equal to 6 (which means that more energy is saved than anticipated). Therefore, deep retrofits produce lower energy savings than expected with theoretical calculation (Fig. 6a), but nevertheless result in higher actual energy saving (measured in absolute terms) than shallow retrofit measures (Fig. 6b). It has been suggested that deep retrofits can guarantee both thermal comfort and low energy use, whereas

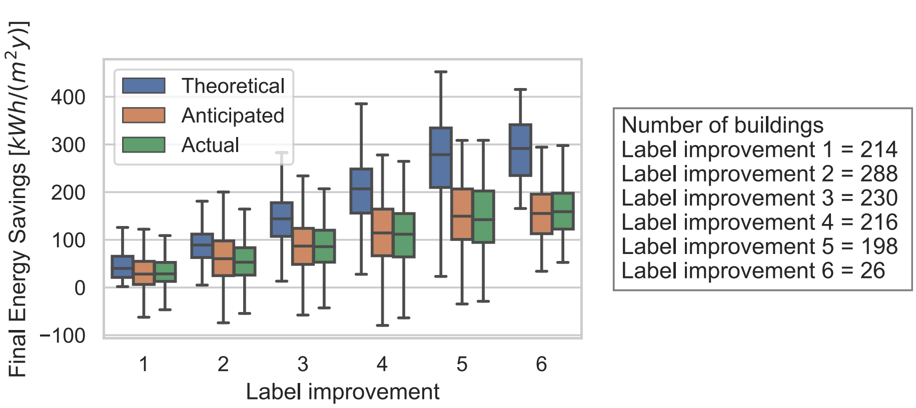

Fig. 4. Distributions of energy savings according to the different calculation methods (tabulated values in Table E.1).

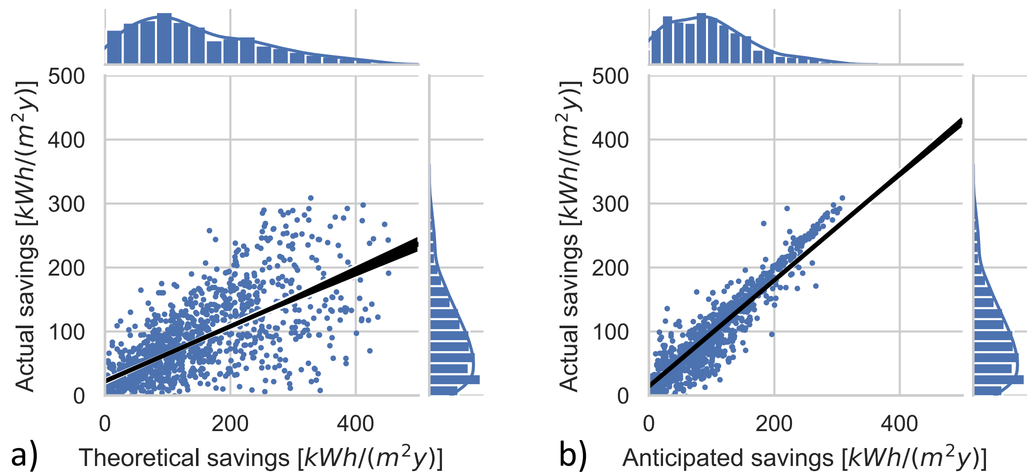

Fig. 5. Actual vs. Theoretical savings (a) and Actual vs. Anticipated savings (b), with linear regression trend line marked in black (each dot represents a building).

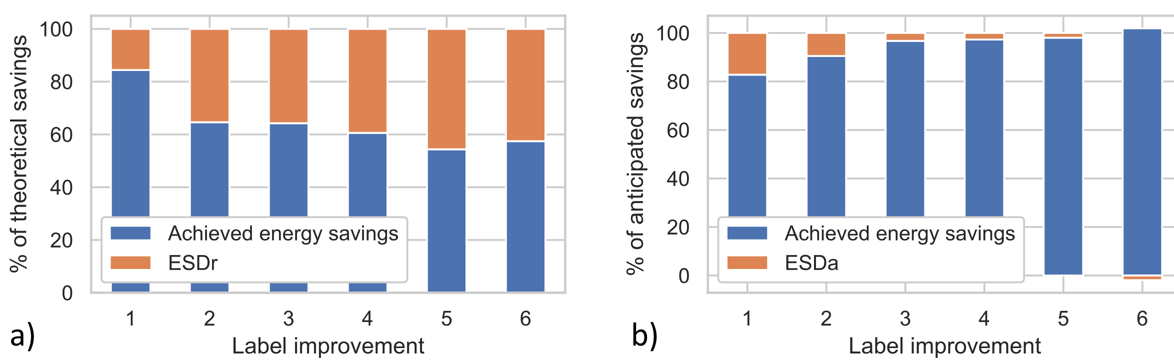

Fig. 6. Fraction of achieved theoretical savings (a) and anticipated savings (b) as a function of the Label improvement.

shallow retrofit can only provide one or the other [26]. This result is similar to the findings in other European countries as the UK [26], the Netherlands [47], and Ireland [52].

These results suggest that the ESDr is not a useful indicator to evaluate the success of an energy retrofit, as most of the shortfall in energy is due to an over-estimation of the consumption of the building before the retrofit, as previously pointed out by other authors [2,11,19,49].

Conversely, the ESDa is a better indicator for judging the success of a retrofit as the model error is small in absolute terms for the post-retrofitted buildings. This has the benefit that significant deviations between anticipated and actual savings could be used to identify system or operational failures (i.e. if the building does not perform as anticipated, this is most likely because of a practical issue rather than an error in the calculation). Finally, the small value of ESDa suggests that the overall quality of the energy retrofitting performed within this subsample is high, indicating that on-site workmanship does not seem to be a major problem in Switzerland.

3.4. Limitations

It is important to note the minimum and maximum values in Fig. 3, which show that the energy consumption of some buildings is actually very different than expected. This spread of values also demonstrates that there is a strong overlap in final energy consumption across the ratings. This means that using only the energy rating based on theoretical calculation as predictor of actual energy consumption is subject to high uncertainty. The finding for the buildings after retrofit, according to which the performance gap is only 2.1% (i.e. actual consumption slightly larger than calculated), is reassuring but it is in contrast to some previous studies relying on case studies in Switzerland [23,24]. Our finding is based on a sample from the CECB and data cleaning may have biased the sample towards better performance. Indeed, it was decided to remove all buildings that showed a deterioration of the energy label after retrofit. Moreover, although the removal of outliers (i.e. buildings with energy consumption values which are extremely different from other buildings) should have removed both cases with above-average consumption and cases with below-average consumption, further research would be required to understand whether the buildings included in the original CECB database perform better than average in Switzerland.

Another source of uncertainty related to the CECB database is the acquisition of the actual energy data. This consumption value should be determined as an average over three years, thereby correcting for heating degree days relative to the standard year which should eliminate the effect of particularly extreme conditions (particularly cold or hot years). It is important to note that the accuracy of the actual consumption data and any cross-checking are entirely under the CECB Expert's responsibility, neither the online tool [79] nor any authorities can check them. However, these internal quality cross-checks alone do not guarantee enough quality of overall results. On a cantonal level, the authorities handling the CECB reports in the context of subsidy requests can reject a report considered not reliable or insufficient. A quality check program has been initiated in 2017 on a nationwide scale, not only to evaluate the quality of the certificates generated, but also to help the CECB Experts to improve their skills and to optimize the online tool itself.

It should be kept in mind that the recommendations provided in this paper are based on a relatively small sample of the CECB database, for which before and after retrofit labels were available. This gives an incomplete picture of the renovation, notably it is not known what the target efficiency of the renovation was, only the actual outcome. We do not have information as to whether the observed outcome was the intended one.

4. Conclusions

To analyse the performance gap, a subsample of 1172 buildings in the Swiss CECB database was produced consisting of residential buildings with energy certification both before and after retrofit as well as actual (billed) energy consumption for space heating and domestic hot water for both. It was found that buildings with poor thermal performance and low energy rating (E, F, and G) before retrofit present a negative energy performance gap (actual consumption is smaller than calculated, respectively −9.2, –23, and −37%). A median negative performance gap of –23% before retrofit and a good approximation of actual consumption with calculated consumption after retrofit (a positive gap of 2%) was found. The absolute value of performance gap was small for highly efficient buildings even in cases where percentage values were large. This implies that there is no counter-indication to high-efficiency retrofits – because despite a slightly positive gap, the energy savings compared to the pre-refit condition are still substantial.

A regression analysis was performed to establish which of the building's parameters present in the CECB were the main determinants of the identified gap. It was found that using the values provided in the energy certificate a much higher degree of variation in the theoretical consumption was explained by the regression model (70%) when compared with the actual consumption (40%). It was found that variables which, as expected, are strong indicators of theoretical consumption such as the energy label and the thermal proprieties of the envelope (U-value of the roof and external walls) are less reliable in characterizing actual consumption. This result could be partially a reflection of the low level of detail that is requested to describe the external walls in the CECB, where mostly standard values are used, contrary to the higher level of accuracy applied for windows, resulting in a better prediction of the actual consumption. Therefore, a more detailed description of the building element would allow to choose more appropriate standard values which would then result in a better agreement of the theoretical with the actual value.

Analysis of the Energy Savings Deficit found that using ESDr (using the change in theoretical consumption for predicting energy savings) overestimated the savings by 37% compared to reality. This is important for energy policy as the expected reduction in energy consumption from energy retrofit could be a fraction of what was planned. This result suggests once again [48,50,52] the need to revise the calculation method applied for energy performance certificates, or at least to adapt the standard values used as part of it in order to more accurately estimate the consumption at the building stock scale and the energy savings potential. Moreover, the results seem to indicate that such standard values have been in use in order to falsely inflate the theoretical energy savings potential. The current grant scheme provides uniform subsidies for every extra label level (Label improvement) gained regardless of the starting point [65]. Based on our findings, the existing approaches to subsidising CECB energy label improvements should be reviewed. A possible solution could be to request an independent expert not involved in the retrofitting activities to prepare the certificate before retrofit or the new certificate after retrofit, as similarly suggested for Ireland [52].

Estimating energy savings using the ESDa gives a much smaller difference compared to the actual savings (ESDa 3.6%), implying that a reasonably realistic assessment of real energy savings can be achieved by comparing the actual current consumption with the expected theoretical consumption defined by the certificate after retrofit. This demonstrates that the use of ESDa does not entail any significant change in the building model or in the entire certification process, but instead it requires the use of actual values. This helps minimise the impact of the performance gap and highlights the importance of monitoring energy consumption. The ESDa can be used as an indicator of renovation quality: since the expected uncertainty is smaller for this indicator, significant deviations would indicate a problem in the execution of the construction.

Overall, this paper has shown once more that the certification method currently in use cannot be considered accurate if compared to actual consumptions. Therefore, care is necessary when using the CECB, which have been "designed around standard operating conditions to assist with comparisons between buildings [49] " as a tool to predict actual energy savings across the building stock. More realistic standard values should be used, that would result in a more accurate estimation of energy consumption, or alternatively methods based on actual energy data should be developed.

While the median performance gap observed was smaller than those observed in case studies within Switzerland [36,53], within our sample a number of cases were observed with large deviations from the median, which is consistent with these case studies. We were not able to establish the extent to which our sample of retrofitted buildings was representative of the national building stock (it is nevertheless the largest sample studied so far in Switzerland), or whether the database used includes a systematic bias which reduces the median gap. Further research is needed on this aspect.

CRediT authorship contribution statement

Stefano Cozza: Conceptualization, Methodology, Software, Writing - original draft. **Jonathan Chambers:** Methodology, Software, Validation, Data curation, Writing - review & editing. **Chirag Deb:** Writing - review & editing. **Jean-Louis Scartezzini:** Supervision. **Arno Schlüter:** Supervision. **Martin K. Patel:** Validation, Writing - review & editing, Funding acquisition, Supervision.

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.

Acknowledgement

This research is financially supported by the Swiss Innovation Agency Innosuisse and is part of the Swiss Competence Center for Energy Research on Future Energy Efficient Buildings & Districts (SCCER FEEB&D). It builds on work done in the GAPxPLORE project ("Energy performance gap in existing, new and renovated buildings – Learning from large-scale datasets") funded by the Swiss Federal Office of Energy. The authors would also like to thank Sylvain Sardy from University of Geneva for his valuable inputs on the regression analysis.

Appendix A

The standard condition of use of the building for the theoretical calculation are given in the Swiss norms SIA 380/1 and SIA 2024, based on the EU norm EN ISO 13790, and summarized in Table A.1.

Table A1
Standard conditions of use.

Operational parameters	Unit	Standard conditions
Indoor temperature	°C	20
Surface per person	m^2/p	40
Thermal gain per person	W/p	70
Metabolic activity	met	1.2
Days of use per year	d/y	365
Occupied hours	h/d	12
Appliance thermal gain	W/m^2	8
Appliance use hours	h/d	6.1
Lighting thermal gain	W/m^2	2.7
Lighting use hours	h/d	7
Air change rate	$m^3/(m^2*h)$	0.7
Hot water demand per person	l/(d*p)	35

Appendix B

The sample size for each label improvement, together with the energy label after retrofit are reported in Table B.1. After a low-impact retrofit (Label improvement 1, e.g. from F to E) the energy label achieved is quite heterogeneous (35% of the D-rated buildings go to C, 21% of the E-rated buildings to D, and 19% of the F-rated buildings to E), showing that minor improvements operations (e.g. replacing of the windows or of the heating system) are executed regardless of the initial stage of the building

Table B1
Sample size and breakdown for different Label improvement.

Label improvement	Total buildings	Energy label after retrofit	Buildings no.	Buildings %
1	214	A	2	0.9
		B	27	12.6
		C	75	35.0
		D	46	21.5
		E	40	18.7
		F	24	11.2
2	288	A	4	1.4
		B	44	15.3
		C	139	48.3
		D	63	21.9
		E	38	13.2
3	230	A	11	4.8
		B	58	25.2
		C	87	37.8
		D	74	32.2
4	216	A	4	1.9
		B	70	32.4
		C	142	65.7
5	198	A	4	2.0
		B	194	98.0
6	26	A	26	100.0

Appendix C

Tables C.1 and C.2.

Table C1
Results of the regression analysis for the performance gap using the parameters recorded in the CECB as independent variables. The VIF value for each variable is also reported in the last column.

Independent variables	Dummies	Performance gap [R^2 = 26.9%]			
		B	β	Sign.	VIF
Constant		244.162		0.089	
Building type	SDB – MDB	−7.258	−0.080	0.017	1.841
Construction year	Ratio Variable	0.015	0.019	0.485	1.182
ERA	Ratio Variable	–	–	–	1.462
Envelope factor	Ratio Variable	−12.112	−0.128	0.000	2.149
Energy Label [dummy variable is G label]	B	98.899	0.190	0.000	1.320
	C	14.841	0.072	0.015	2.060
	D	10.036	0.077	0.011	2.416
	E	–	–	–	2.236
	F	−1.238	−0.011	0.678	1.640
Mechanical ventilation	Yes – no	–	–	–	1.213
Heating System [dummy variable is Oil boiler]	DH*	–	–	–	1.221
	Electro	−12.550	−0.102	0.000	1.346
	Gas	–	–	–	1.178
	HP*	45.098	0.169	0.000	1.286
	Wood	−7.363	−0.036	0.209	1.234
Heating System construction year	Ratio Variable	−0.114	−0.048	0.103	1.445
U-value ground	Ratio Variable	–	–	–	1.325
U-value roof/ceiling	Ratio Variable	−7.080	−0.101	0.001	1.407
U-value external walls	Ratio Variable	−24.797	−0.235	0.000	1.754
U-value windows	Ratio Variable	−0.351	−0.007	0.809	1.215
Construction Type [dummy variable is Heavy construction]	Medium	–	–	–	1.115
	Light	−11.929	−0.071	0.009	1.194
	Very light	–	–	–	1.058

* DH: district heating; HP: heat pump.

Table C2
Results of the regression analysis for the determinants of theoretical and actual energy consumption, and performance gap post-retrofit using the parameters recorded in the CECB as independent variables.

Independent variables	Dummies	Theoretical [R^2 = 55.8%]			Actual [R^2 = 35.5%]			Performance gap [R^2 = 13.4%]		
		B	β	Sign.	B	β	Sign.	B	β	Sign.
Constant		629.685		0.000	603.457		0.000	29.918		0.836
Building type	SDB – MDB	−7.989	−0.070	0.005	−14.385	−0.146	0.000	−18.764	−0.160	0.000
Construction year	Ratio Variable	0.040	0.042	0.037	0.038	0.046	0.054	0.014	0.014	0.617
ERA	Ratio Variable	–	–	–	–	–	–	–	–	–
Envelope factor	Ratio Variable	22.417	0.187	0.000	14.960	0.145	0.000	−11.639	−0.095	0.006
Energy Label [dummy variable is A label]	B	−3.534	−0.005	0.799	8.684	0.015	0.542	−8.836	−0.013	0.654
	C	2.261	0.008	0.699	4.605	0.019	0.442	−5.722	−0.020	0.490
	D	1.595	0.010	0.638	0.804	0.006	0.817	−3.407	−0.020	0.479
	E	–	–	–	–	–	–	–	–	–
	F	–	–	–	–	–	–	–	–	–
Mechanical ventilation	Yes - no	–	–	–	–	–	–	–	–	–
Heating System [dummy variable is Oil boiler]	DH*	−11.430	−0.025	0.224	−13.907	−0.035	0.163	16.235	0.021	0.051
	Electro	–	–	–	–	–	–	–	–	–
	Gas	–	–	–	–	–	–	–	–	–
	HP*	−14.730	−0.041	0.054	−8.368	−0.027	0.285	13.914	0.038	0.199
	Wood	4.357	0.020	0.333	−0.481	−0.003	0.917	−13.206	−0.059	0.039
Heating System construction year	Ratio Variable	−0.370	−0.161	0.000	−0.331	−0.168	0.000	0.022	0.009	0.733
U-value ground	Ratio Variable	–	–	–	–	–	–	–	–	–
U-value roof/ceiling	Ratio Variable	42.750	0.196	0.000	29.074	0.155	0.000	−11.141	−0.050	0.114
U-value external walls	Ratio Variable	82.784	0.452	0.000	51.606	0.329	0.000	−38.151	−0.203	0.000
U-value windows	Ratio Variable	25.696	0.244	0.000	17.720	0.196	0.000	−12.849	−0.119	0.000
Construction Type [dummy variable is Heavy construction]	Medium	–	–	–	–	–	–	–	–	–
	Light	−0.785	−0.004	0.851	−2.984	−0.017	0.485	−7.512	−0.036	0.204
	Very light	–	–	–	–	–	–	–	–	–

* DH: district heating; HP: heat pump.

Appendix D

Lasso aims at low prediction error while disregarding insignificant variables by controlling a regularization parameter $\lambda > 0$. The goal of cross-validation is to select the λ with best predictive error. To that aim the data is repeatedly partitioned into training and validation data. The model is fitted to the training data and the validation data is used to estimate the prediction error. This allows the identification of the values of λ that optimize the predictive performance (i.e., minimize the estimated squared prediction error). The plot demonstrates the prediction error as a function of log(λ) for the Lasso model using the full data with interaction terms. The λ_{min} value that minimizes the absolute prediction error is indicated by the grey line on the left. This error has a certain standard error (depicted by the light grey whiskers to every red point). The grey line on the right represents the value that is a bit less prone to overfitting, λ_{1se}, obtained extending the height of the upper whisker of the λ_{min} to the right until the last point which is still below that imaginary line. The top numbering of the plot indicates the number of variables the model is using, going from all variables (left) to more sparse models (right) [80]. This plot helps choosing the best λ. The tuning parameter λ controls the overall strength of the L1 penalty used in Lasso method (Fig. D1).

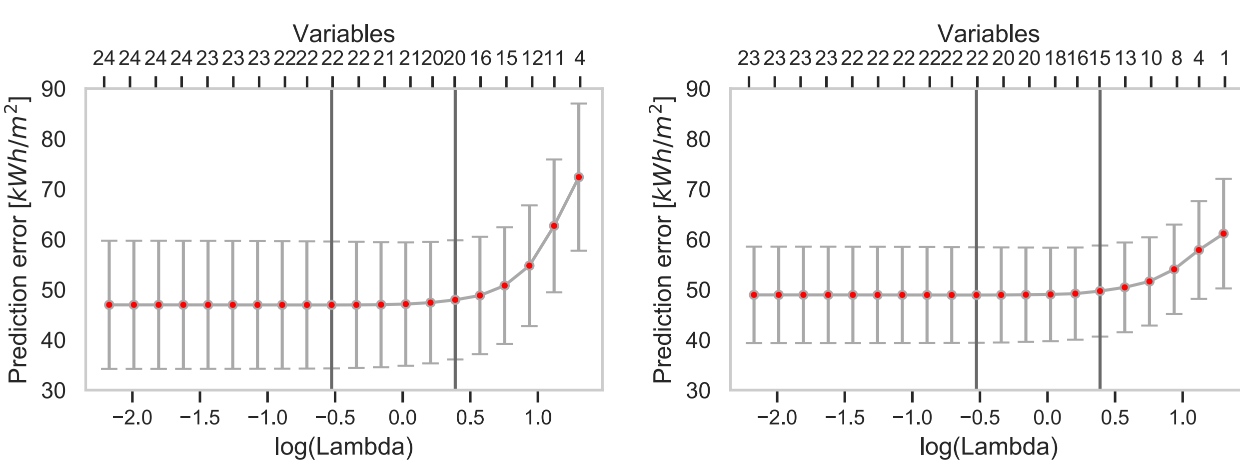

Fig. D1. Explanatory plots for cross-validate errors for the theoretical (a) and actual (b) model.

Appendix E

Tabulated values for the box-plots presented in Fig. 4. Table E.1.

Table E1
Theoretical, Anticipated and Actual savings [kWh/(m^2y)] per Label improvement.

Label improvement	No. Buildings	Savings	Min.	25%	50%	75%	Max.
1	214	Theoretical	1.91	21.2	40.1	65.5	313
		Anticipated	−141	6.4	28.2	54.9	191
		Actual	−104	12.8	28.6	52.6	175
2	288	Theoretical	5.08	62.9	89.1	112	337
		Anticipated	−90.3	24.8	60.4	97.8	200
		Actual	−62.1	26.0	53.0	83.5	192
3	230	Theoretical	13.3	107	144	178	340
		Anticipated	−98.2	48.5	86.9	124	261
		Actual	−42.8	53.1	85.6	120	207
4	216	Theoretical	4.69	156	207	248	435
		Anticipated	−79.5	66.6	114	164	278
		Actual	−63.7	64.2	112	155	265
5	198	Theoretical	23.3	210	279	335	452
		Anticipated	−34.6	101	150	207	309
		Actual	−29.0	94.6	142	202	309
6	26	Theoretical	166	235	292	341	415
		Anticipated	34.0	113	155	196	294
		Actual	52.6	123	159	198	298

References

[1] P. de Wilde, The gap between predicted and measured energy performance of buildings: a framework for investigation, Autom. Constr. 41 (2014) 40–49, https://doi.org/10.1016/j.autcon.2014.02.009.
[2] M. Sunikka-Blank, R. Galvin, Introducing the prebound effect: the gap between performance and actual energy consumption, Build. Res. Inf. 40 (2012) 260–273, https://doi.org/10.1080/09613218.2012.690952 To.
[3] E. Cayre, B. Allibe, M.-H. Laurent, D. Osso, There are people in the house! How the results of purely technical analysis of residential energy consumption are misleading for energy policies, in: ECEEE 2011 Summer Study – Energy Effic. First Found, A Low-Carbon Soc., 2011.
[4] M. Delghust, W. Roelens, T. Tanghe, Y. De Weerdt, A. Janssens, Regulatory energy calculations versus real energy use in high-performance houses, Build. Res. Inf. 43 (2015) 675–690, https://doi.org/10.1080/09613218.2015.1033874.
[5] D. Majcen, L. Itard, H. Visscher, Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications, Energy Policy. 54 (2013) 125–136, https://doi.org/10.1016/j.enpol.2012.11.008.
[6] T. Sharpe, D. Shearer, Adapting the Scottish tenement to 21st century standards: an evaluation of the performance enhancement of a 19th century "Category B" listed tenement block in Edinburgh, J. Cult. Herit. Manage. Sustain. Dev. 3 (2013) 55–67, https://doi.org/10.1108/20441261311317400.
[7] R. Haas, P. Biermayr, The rebound effect for space heating empirical evidence from Austria, Energy Policy 28 (2000) 403–410, https://doi.org/10.1016/S0301-4215(00)00023-9.
[8] A. Merzkirch, T. Hoos, S. Maas, F. Scholzen, D. Waldmann, Wie genau sind unsere Energiepässe?, Bauphysik (2014) 40–43, https://doi.org/10.1002/bapi.201410007.
[9] C. Loucari, J. Taylor, R. Raslan, E. Oikonomou, A. Mavrogianni, Retrofit solutions for solid wall dwellings in England: the impact of uncertainty upon the energy performance gap, Build. Serv. Eng. Res. Technol. 37 (2016) 614–634, https://doi.org/10.1177/0143624416647758.
[10] M. Raynaud, Evaluation ex-post de l'efficacité de solutions de rénovation énergétique en résidentiel, (PhD. Thesis – ParisTech), Paris Institute of Technology, 2014. http://tel.archives-ouvertes.fr/pastel-00973727/.
[11] D. Calì, T. Osterhage, R. Streblow, D. Müller, Energy performance gap in refurbished German dwellings: lesson learned from a field test, Energy Build. 127 (2016) 1146–1158, https://doi.org/10.1016/j.enbuild.2016.05.020.
[12] D. Majcen, L. Itard, H. Visscher, Actual heating energy savings in thermally renovated Dutch dwellings, Energy Policy 97 (2016) 82–92, https://doi.org/10.1016/j.enpol.2016.07.015.
[13] Netherlands Enterprise Agency, https://english.rvo.nl/ (2020)
[14] P.H.G. Berkhout, J.C. Muskens, J.W. Velthuijsen, Defining the rebound effect, Energy Policy 28 (2000) 425–432.
[15] A. Druckman, M. Chitnis, S. Sorrell, T. Jackson, Missing carbon reductions? Exploring rebound and backfire effects in UK households, Energy Policy 49 (2011) 778, https://doi.org/10.1016/j.enpol.2012.06.045.
[16] R. Galvin, Making the 'rebound effect' more useful for performance evaluation of thermal retrofits of existing homes: defining the 'energy savings deficit' and the 'energy performance gap', Energy Build. 69 (2013) 515–524, https://doi.org/10.1016/j.enbuild.2013.11.004.
[17] S. Sorrell, J. Dimitropoulos, M. Sommerville, Empirical estimates of the direct rebound effect: a review, Energy Policy. 37 (2009) 1356–1371, https://doi.org/10.1016/j.enpol.2008.11.026.
[18] R. Galvin, M. Sunikka-Blank, Quantification of (p)rebound effects in retrofit policies – why does it matter?, Energy. (2016), https://doi.org/10.1016/j.energy.2015.12.034.
[19] M. Brøgger, P. Bacher, H. Madsen, K.B. Wittchen, Estimating the influence of rebound effects on the energy-saving potential in building stocks, Energy Build. 181 (2018) 62–74, https://doi.org/10.1016/j.enbuild.2018.10.006.
[20] M. Chitnis, S. Sorrell, Living up to expectations: estimating direct and indirect rebound effects for UK households, Energy Econ. 52 (2015) S100–S116, https://doi.org/10.1016/j.eneco.2015.08.026.
[21] S. Kelly, D. Crawford-Brown, M.G. Pollitt, Building performance evaluation and certification in the UK: Is SAP fit for purpose?, Renew. Sustain. Energy Rev. 16 (2012) 6861–6878, https://doi.org/10.1016/j.rser.2012.07.018.
[22] F. Flourentzou, A.Y. Ivanov, P. Samuel, Understand, simulate, anticipate and correct performance gap in NZEB refurbishment of residential buildings, J. Phys. Conf. Ser. (2019), https://doi.org/10.1088/1742-6596/1343/1/012177.
[23] C. Deb, M. Frei, A. Schlueter, Identifying temporal properties of building components and indoor environment for building performance assessment, Build. Environ. (2020), https://doi.org/10.1016/j.buildenv.2019.106506.
[24] O. Guerra-Santin, C.A. Tweed, In-use monitoring of buildings: an overview of data collection methods, Energy Build. 93 (2015) 189–207, https://doi.org/10.1016/j.enbuild.2015.02.042.
[25] P.X.W. Zou, X. Xu, J. Sanjayan, J. Wang, Review of 10 years research on building energy performance gap: life-cycle and stakeholder perspectives, Energy Build. 178 (2018) 165–181, https://doi.org/10.1016/j.enbuild.2018.08.040.
[26] J. Love, Understanding the Interactions Between Occupants, Heating Systems, and Building Fabric in the Context of Social Housing, University College London, 2014.
[27] J. Khoury, P. Hollmuller, B. Lachal, S. Schneider, U. Lehmann, COMPARE RENOVE: du catalogue de solutions à la performance réelle des rénovations énergétiques, Office fédéral de l'énergie (OFEN), 2018.
[28] C. Van Dronkelaar, M. Dowson, C. Spataru, D. Mumovic, A review of the energy performance gap and its underlying causes in non-domestic buildings, Front. Mech. Eng. 1 (2016) 1–14, https://doi.org/10.3389/fmech.2015.00017.
[29] W. Tian, Y. Heo, P. de Wilde, Z. Li, D. Yan, C.S. Park, X. Feng, G. Augenbroe, A review of uncertainty analysis in building energy assessment, Renew. Sustain. Energy Rev. 93 (2018) 285–301, https://doi.org/10.1016/j.rser.2018.05.029.
[30] G. Branco, B. Lachal, P. Gallinelli, W. Weber, Predicted versus observed heat consumption of a low energy multifamily complex in Switzerland based on long-term experimental data, Energy Build. 36 (2004) 543–555, https://doi.org/10.1016/j.enbuild.2004.01.028.
[31] M. Hughes, J. Palmer, V. Cheng, D. Shipworth, Global sensitivity analysis of England's housing energy model, J. Build. Perform. Simul. 8 (2015) 283–294, https://doi.org/10.1080/19401493.2014.925505.
[32] Y. Zhang, X. Bai, F.P. Mills, J.C.V. Pezzey, Rethinking the role of occupant behavior in building energy performance: a review, Energy Build. 172 (2018) 279–294, https://doi.org/10.1016/j.enbuild.2018.05.017.
[33] I. Gaetani, P.J. Hoes, J.L.M. Hensen, Occupant behavior in building energy simulation: towards a fit-for-purpose modeling strategy, Energy Build. 121 (2016) 188–204, https://doi.org/10.1016/j.enbuild.2016.03.038.
[34] J. Khoury, Rénovation énergétique des bâtiments résidentiels collectifs : état des lieux, retours d'expérience et potentiels du parc genevois, 2014. https://archive-ouverte.unige.ch/.
[35] V. Fournier, Comparaison des coûts face à l'impact environnemental entre travaux de rénovation sur l'enveloppe et changement de vecteur énergétique de plusieurs projets immobiliers à Genève, EPFL, 2018
[36] C. Hoffmann, A. Geissler, The prebound-effect in detail: real indoor temperatures in basements and measured versus calculated U-values, Energy Procedia 122 (2017) 32–37, https://doi.org/10.1016/J.EGYPRO.2017.07.301.
[37] U. Lehmann, J. Khoury, M.K. Patel, Actual energy performance of student housing: case study, benchmarking and performance gap analysis, Energy Procedia 122 (2017) 163–168, https://doi.org/10.1016/J.EGYPRO.2017.07.339.
[38] C. Ahern, B. Norton, Thermal energy refurbishment status of the Irish housing stock, Energy Build. 202 (2019), https://doi.org/10.1016/j.enbuild.2019.109348 109348.
[39] J. Kragh, J. Rose, H.N. Knudsen, O.M. Jensen, Possible explanations for the gap between calculated and measured energy consumption of new houses, Energy Procedia 132 (2017) 69–74, https://doi.org/10.1016/j.egypro.2017.09.638.
[40] L. La Fleur, B. Moshfegh, P. Rohdin, Measured and predicted energy use and indoor climate before and after a major renovation of an apartment building in Sweden, Energy Build. 146 (2017) 98–110, https://doi.org/10.1016/j.enbuild.2017.04.042.
[41] D. Grossmann, R. Galvin, J. Weiss, R. Madlener, B. Hirschl, A methodology for estimating rebound effects in non-residential public service buildings: Case study of four buildings in Germany, Energy Build. 111 (2016) 455–467, https://doi.org/10.1016/j.enbuild.2015.11.063.
[42] E. Burman, D. Mumovic, J. Kimpian, Towards measurement and verification of energy performance under the framework of the European directive for energy performance of buildings, Energy. 77 (2014) 153–163, https://doi.org/10.1016/j.energy.2014.05.102.

[43] R. De Lieto Vollaro, C. Guattari, L. Evangelisti, G. Battista, E. Carnielo, P. Gori, Building energy performance analysis: a case study, Energy Build. 87 (2015) 87–94, https://doi.org/10.1016/j.enbuild.2014.10.080.
[44] M. Bauer, A. Kuenlin, Bewertung des Experteneinflusses auf die GEAK®-Klassifikation eines Gebäudes, Direktion für Energie (DiREN), 2013
[45] H. Hens, W. Parijs, M. Deurinck, Energy consumption for heating and rebound effects, Energy Build. 42 (2010) 105–110, https://doi.org/10.1016/j.enbuild.2009.07.017.
[46] T. Ekström, R. Bernardo, Å. Blomsterberg, Cost-effective passive house renovation packages for Swedish single-family houses from the 1960s and 1970s, Energy Build. 161 (2018) 89–102, https://doi.org/10.1016/j.enbuild.2017.12.018.
[47] P. van den Brom, A. Meijer, H. Visscher, Actual energy saving effects of thermal renovations in dwellings—longitudinal data analysis including building and occupant characteristics, Energy Build. 182 (2019) 251–263, https://doi.org/10.1016/j.enbuild.2018.10.025.
[48] F. Filippidou, N. Nieboer, H. Visscher, Effectiveness of energy renovations: a reassessment based on actual consumption savings, Energy Effi. 12 (2018) 19–35, https://doi.org/10.1007/s12053-018-9634-8.
[49] A.J. Summerfield, T. Oreszczyn, J. Palmer, I.G. Hamilton, F.G.N. Li, J. Crawley, R.J. Lowe, What do empirical findings reveal about modelled energy demand and energy ratings? Comparisons of gas consumption across the English residential sector, Energy Policy. 129 (2019) 997–1007, https://doi.org/10.1016/j.enpol.2019.02.033.
[50] J. Crawley, P. Biddulph, P.J. Northrop, J. Wingfield, T. Oreszczyn, C. Elwell, Quantifying the measurement error on England and Wales EPC ratings, Energies. 12 (2019), https://doi.org/10.3390/en12183523.
[51] M. Laurent, B. Allibe, C. Tigchelaar, T. Oreszczyn, I. Hamilton, R. Galvin, Back to reality: how domestic energy efficiency policies in four European countries can be improved by using empirical data instead of normative calculation, Eceee Summer Study Proc. (2013) 2057–2070.
[52] T. Mac Uidhir, F. Rogan, M. Collins, J. Curtis, B.P.Ó. Gallachóir, Improving energy savings from a residential retrofit policy: a new model to inform better retrofit decisions, Energy Build. (2020), https://doi.org/10.1016/j.enbuild.2019.109656.
[53] J. Khoury, Z. Alameddine, P. Hollmuller, Understanding and bridging the energy performance gap in building retrofit, Energy Procedia 122 (2017) 217–222, https://doi.org/10.1016/j.egypro.2017.07.348.
[54] W. Reimann, E. Buhlmann, M. Lehmann, Erfolgskontrolle Gebäudeenergiestandards 2014-2015, Bundesamt für Energie (BFE), 2016
[55] S. Wyss, W. Hässig, UFELD: Feldmessungen von U-Werten zur Überprüfung der im Gebäudeenergieausweis (GEAK) hinterlegten U-Werte, Bundesamt für Energie (BFE) (2016).
[56] Conferenza Cantonale dei Direttori dell'Energia, CECB – Web Page, (2020). http://cecb.ch/ (accessed January 7, 2020).
[57] S.H. Chan, J.H. Chen, Y.H. Li, L.M. Tsai, Gly1057Asp polymorphism of insulin receptor substrate-2 is associated with coronary artery disease in the Taiwanese population, J. Biomed. Sci. 19 (2012) 1–8, https://doi.org/10.1186/1423-0127-19-100.
[58] N. Roberts, J.B. Thatcher, Conceptualizing and testing formative constructs: tutorial and annotated example, Data Base Adv. Inf. Syst. 40 (2009) 9–39, https://doi.org/10.1145/1592401.1592405.
[59] G.M. Huebner, I. Hamilton, Z. Chalabi, D. Shipworth, T. Oreszczyn, Explaining domestic energy consumption – the comparative contribution of building factors, socio-demographics, behaviours and attitudes, Appl. Energy 159 (2015) 589–600, https://doi.org/10.1016/j.apenergy.2015.09.028.
[60] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B. (1996), https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
[61] A. Field, Discovering Statiscs Using SPSS, SAGE, 2009.
[62] The European Parliament and the Council of the EU, Directive 2002/91/EC of the European parliament of the council of 16 December 2002 on the energy performance of buildings, in: Off. J. Eur. Communities 2003. L1/65-71, 2003.
[63] SIA, SIA 2031 – Energy certificate for buildings, Swiss Society of Engineers and Architects (SIA), 2016.
[64] Facteurs de EnDK, pondération nationaux pour l'évaluation des batiments, Bundesamt für, Energie (BFE) (2016).
[65] S. Cozza, J. Chambers, M.K. Patel, Measuring the thermal energy performance gap of labelled residential buildings in Switzerland, Energy Policy. (2019), https://doi.org/10.1016/j.enpol.2019.111085 111085.
[66] FSO, Construction and housing - key figures, Federal Population Census, Buildings and dwellings statistics, 2017.
[67] K.N. Streicher, P. Padey, D. Parra, M.C. Bürer, M.K. Patel, Assessment of the current thermal performance level of the Swiss residential building stock: Statistical analysis of energy performance certificates, Energy Build. 178 (2018) 360–378, https://doi.org/10.1016/j.enbuild.2018.08.032.
[68] S. Cozza, J. Chambers, A. Geissler, K. Wesselmann, C. Gambato, G. Branca, G. Cadonau, L. Arnold, M.K. Patel, GAPxPLORE: Energy Performance Gap in Existing, New, and Renovated Buildings, Swiss Federal Office of Energy (SFOE), 2019.
[69] SIA, SIA 2028 – Dati climatici per la fisica della costruzione, per l'energia e per l'impiantistica negli edifici, Swiss Society of Engineers and Architects (SIA), 2010.
[70] B. Iglewicz, D. Hoaglin, Volume 16: How to Detect and Handle Outliers, ASQC Basic Ref. Qual. Control Stat. Tech., 1993.
[71] Odyssee Project, Enerdata – Odysee: European Energy Efficiency Databse, 2018.
[72] Conferenza Cantonale dei Direttori dell'Energia CDE, Modello di prescrizioni energetiche dei cantoni (MoPEC), 2014.
[73] F. Wilcoxon, Individual comparisons of grouped data by ranking methods, J. Econ. Entomol. 39 (1946) 269, https://doi.org/10.1093/jee/39.2.269.
[74] D. Majcen, Predicting energy consumption and savings in the housing stock: A performance gap analysis in the Netherlands (PhD. thesis), Delft University of Technology, Faculty of Architecture and the Built Environment, 2016.
[75] Swiss Federal Office of Energy, Energy Strategy 2050 Once the New Energy Act Is in Force, Bundesamt für Energie (BFE), 2018.
[76] Prognos, Die Energieperspektiven für die Schweiz bis 2050, Bundesamt für Energie (BFE), 2012.
[77] D. Calì, R.K. Andersen, D. Müller, B.W. Olesen, Analysis of occupants' behavior related to the use of windows in German households, Build. Environ. (2016), https://doi.org/10.1016/j.buildenv.2016.03.024.
[78] E. Cuerda, O. Guerra-Santin, J.J. Sendra, F.J. Neila, Understanding the performance gap in energy retrofitting: measured input data for adjusting building simulation models, Energy Build. (2019), https://doi.org/10.1016/j.enbuild.2019.109688.
[79] K. Wesselmann, Manuel utilisateur de l'outil en ligne CECB, 2017.
[80] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, 2013. doi:10.1016/j.peva.2007.06.006.