

Handling data uncertainties when using Swedish energy performance certificate data to describe energy usage in the building stock



Mikael Mangold*, Magnus Österbring¹, Holger Wallbaum²

Chalmers University of Technology, Civil and Environmental Engineering, Sven Hultins gata 8, 412 96 Göteborg, Sweden

ARTICLE INFO

Article history:

Received 30 January 2015

Received in revised form 18 May 2015

Accepted 28 May 2015

Available online 30 May 2015

Keywords:

Building stock

Energy performance certificates

Data quality

Atemp

Building area

Energy usage

ABSTRACT

The EU directive to create Energy Performance Certificates (EPC) for all buildings was implemented in Sweden as a tool to advise building owners on possible improvements and to give energy efficiency visible market value. The Swedish EPCs include measured energy usage. Currently 82% of the buildings have EPCs; this database makes it possible to create overview and to validate models of the building stock in an unprecedented high detail.

However, the process of issuing Swedish EPCs has received criticism from real estate agents, real estate owners, Energy Experts, and Boverket, the national agency responsible for EPC data collection. In order to use the EPC data for describing the building stock it is necessary to assess and remediate the data quality. This has been done by merging the EPC data with databases of the Housing and Urban Development office and one of the larger real estate companies in Sweden, Riksbyggen. The Swedish EPC specific area measurement, Atemp, is found to vary according to methods of derivation. The method of estimating Atemp is improved using a stepwise regression model ($R^2 = 0.979$). This method can be applied to subsets of EPCs depending on the intended way of describing the building stock.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In 2010, buildings accounted for 32% of total global final energy use [1]. The European Directive 2012/27/EU [2] requires member states to have a strategy for refurbishment of the building stock with the target of reducing energy usage by 20% by 2020. The Swedish government has set an additional target of reducing the energy usage in the building sector by 50% by 2050 [3]. Having an overview of energy usage in the building stock is necessary for creating a refurbishment strategy, as engineering-based models are used in refurbishment strategies to predict energy savings for buildings after the application of renovation measures [4–6]. The difficulties with using engineering-based models on a city scale is that reliable data on energy usage on the building level is limited [7].

Abbreviations: EPC, Energy performance certificate; HUD, Housing and Urban Development office.

* Corresponding author. Tel.: +46 702979778.

E-mail addresses: mikael.mangold@chalmers.se (M. Mangold), magost@chalmers.se (M. Österbring), holger.wallbaum@chalmers.se (H. Wallbaum).

¹ Tel.: +46 7022846310.

² Tel.: +46 317721994.

The Swedish energy performance certificates (EPC), managed by The Swedish National Board of Housing, Building and Planning (Boverket), contain measured energy usage data for more than 500 000 buildings. The EPC data can be used to describe the energy usage of the building stock, as it was done for France [8]. However, the data quality of the Swedish EPCs has been criticized [9,10]. The purpose of this paper is to assess and remediate the data uncertainty in the Gothenburg EPC data to be used to produce a comprehensive overview of energy usage in the building stock, as did Kragh and Wittchen [11] for the Danish building stock with different preconditions. The EPC database is compared with billing data from Riksbyggen, a real estate company with a portfolio of 170 000 apartments of which 13 500 are in Gothenburg, as well as data from the Gothenburg Housing and Urban Development office (HUD, Swedish translation: Stadsbyggnadskontoret). This article focuses specially on EPCs for multi-family-dwellings since retrofitting measures are primarily called for in this segment of the building stock [12]. Furthermore most buildings in this segment have an EPC.

In the following chapter the background and previous research in connection with the Swedish EPC is summarized. Thereafter the methods with which the EPC data quality have been analyzed are described, followed by a results chapter that describes findings and detail recommendations for usage of the EPCs in description of building stock or validation of building stock models.

1.1. Background of the Swedish EPC data

The EPC system was proposed by the EU Energy Performance of Buildings Directive [13], in 2002 inspired by the Kyoto Protocol. Prior to that a multitude of national building labeling existed in the member states and the EPC system has been implemented differently depending on existing labeling systems and national conditions [14]. The Swedish EPC system was specified by the Swedish Government [15]. The purpose of the Swedish EPC is to promote energy usage reducing measures by making the actual yearly cost of energy usage visible when buildings are sold [16]. This is different from how other European member states have implemented the EPC [17–19] and has had implications on data required for the Swedish EPC, see Table 1.

On a European level the accuracy of the EPC system is estimated to be 35% [17]. Nevertheless the EPC data has been used to describe building stocks of several member states. Majcen et al. [21] used the Dutch EPC data to assess the development of the Dutch building stock. The biggest problem of the study is reported to be uncertainties regarding data quality. The Danish EPC system has also been criticized for being too expensive and the data unreliable [22], but the data has been used to describe the Danish building stock [11].

A Swedish EPC is more costly than the EPC of other member states [17], and should include a site visit by a certified Energy Expert. The Swedish EPC system has been criticized by practitioners and academics for being inaccurate, expensive [10,23] and not usable for the purpose of advising building owners on how to reduce energy usage [24]. Neither has it achieved the goal of making building purchasers pay more for a building that demands less energy [18,25,26]. However, the data generated in the EPC process can be used for the purpose of generating a comprehensive overview of energy usage in the building stock, if data quality issues are solved. Booth et al. [27] have analyzed ways of handling data uncertainty, but with EPCs in the UK, which contain calculated energy usage and not measured data, as in Sweden.

The contribution of this article is to detail ways of handling uncertainty in EPCs with measured energy usage on a city scale. The article details Swedish EPC data coverage, accuracy and precision in relation to available HUD data and heating, electricity and water billing data.

The Swedish EPC data quality has been evaluated by Claesson [9] using 10 randomly selected EPCs and three interviews with Energy Experts. Claesson [9] conclude that most EPCs contain estimations and distributed values of energy usage which contain uncertainty. Claesson [9] recommends that data plausibility analysis is included as part of the EPC. Stensson [10] compared five EPCs made for the same shopping malls by different Energy Experts, finding that the energy usage varied from 71 to 79 kWh/m² year, primarily due to variations in Atemp.³

The two most used methods of deriving Atemp are to either measure it using the building plan or to derive it from BOA and LOA.⁴ Boverket recommends Atemp to be derived from BOA and LOA for larger multi-family-dwellings by applying a factor of 1.15 to the addition of BOA and LOA, or a factor of 1.25 if there is a heated basement. Prior to this recommendation Göransson [28] made a study on the relationship between Atemp and BOA + LOA based on 151 multi-family-dwellings. Göransson found the measured Atemp

to be 1.35 times larger than BOA + LOA, with a standard deviation (σ) = 0.18.

A purpose of this article is to make measured Atemp comparable with Atemp derived from BOA + LOA to increase the accuracy of a building stock description based on EPC data. To avoid incorrect assumptions when correcting Atemp it is necessary that those assumptions also are valid for the desired building stock description. Therefore a method of finding correction factors, α , using a stepwise regression analysis is presented. Correction factors for Atemp to be used in building stock descriptions in age groups are presented separately.

Liu et al. [29] have made a thorough assessment of data quality on ventilation in the Swedish EPC. Consequently it was decided to not focus further on ventilation. It was also decided to limit the work to not include the recommended energy usage reducing measures, considering the findings of Olofsson [23] and Stensson [10] that this data is of too low quality to draw any conclusions.

2. Materials and methods

By comparing the Gothenburg EPC data with HUD data and Riksbyggen heating, electricity and water billing data, it was possible to analyze the data quality and describe the variance in energy usage. These datasets are detailed in Table 2.

The HUD data is mainly geographical information on the buildings in the city. By graphically dividing the building stock according to subdivision in the EPC data it is possible to observe data quality issues visually, such as over-representation of buildings.

2.1. Area measurements in the EPC

The two most common ways of deriving Atemp for multi-family-dwellings are by measuring it or by estimating it from BOA and LOA. In Fig. 1 Atemp in these EPC groups are plotted against BOA + LOA without any adjustment or removal of outliers. A contribution of this article is to make measured Atemp comparable with Atemp derived from BOA + LOA. This was done in two different ways. The first method is an equation from a stepwise regression analysis which is an improved way of estimating Atemp. The second method is intended to fit a building stock description divided by building age directly.

In Fig. 1, the difference between Atemp and BOA + LOA has been illustrated. The figures contain many data points making odd data points comparatively more visible. The line for Atemp = BOA + LOA (The ratio $X = 1$) have been added as a reference.

A stepwise regression was conducted with the confidence intervals set at 95% containing all variables in the EPC and Atemp from EPCs in which Atemp has been measured ($N = 1163$), Fig. 1, as a dependent. The regression produced an improved equation for the estimation of Atemp from BOA + LOA ($R^2 = 0.979$). The number of cases in the analysis was increased from $N = 777$ to $N = 1163$ by adding BOA + LOA from using HUD data.

The ratio, X , between Atemp and BOA + LOA is expressed for multi-family-dwellings, and the outlier criteria $2 > X > 1$ introduced by Göransson [28] is applied before processing the data. The second method adjusts Atemp statistically with assumptions which must be in accordance with the final intended usage [30]. As a reference, the building stock is therefore divided into age groups in which the averages of X from measured Atemp are used to produce adjustment factors, α , for Atemp derived from BOA + LOA.

These two methods are both simplifications, but can be used when analyzing a building stock statistically. The methods also assume that buildings where Atemp has been measured are representative for the building stock as a whole. Table 3 describes differences and similarities between the two subsets. The method

³ Atemp is a measure of building floor area specifically developed for the EPC in Sweden, and it is calculated differently depending on type of floor space standard. Atemp is defined as the heated floor space including shared spaces and footprints of walls but not including garages.

⁴ BOA and LOA are real estate valuation measures of usable heated floor area for habitation and non-habitation (stair cases, shared areas etc.). BOA and LOA are registered for 90% of the multi-family-dwellings and 74% of the single family buildings.

Table 1
Swedish EPC data overview.

Value category	Data specification	Measurement type
1. Matching, keys, and sorting	National real estate number and index, address, EPC index	
2. Building characteristics	Building age* Complexity* Shared walls with other buildings* Recognition of heritage value	Scale variable [year] Nominal [complex, non-complex] Ordinal [detached, semi-attached, attached] Nominal
3. Building usage	National registration of building usage type code Detailed usage of building*	Nominal Share of building used for the 12 most common types
4. Building area	Interior areas (Atemp ^a , BOA and LOA ^b)* Floors Stair cases Number of apartments Floors below ground Heated garage space	Scales [m ²] Ordinal Ordinal Ordinal Ordinal Scale [m ²]
5. Heating	Energy usage for heating divided in 13 energy sources* Tic box for measurement type Period of measurement	Scales [kWh/year] Nominal [measured, distributed] Interval [year and month]
6. Household electricity and water	Cooling energy usage* Tap water energy usage* Electricity usage divided in: domestic, shared, and non-domestic usage*	All measurements in the category include scale [kWh/year] and nominal [measured, distributed]
7. Ventilation	Ventilation type* Ventilation control conducted*	Nominal [exhaust, balanced, balanced with heat exchanger, exhaust with heat pump, natural ventilation] Nominal [yes, no, partially]
8. Recommended energy usage reducing measures	Tic box for 28 common energy usage reducing measures Estimated decreased energy usage* Estimated cost per saved kWh*	Nominal Scale [kWh/year] Scale [SEK/kWh]

The parts specifically mentioned by EU [13] have been marked with an asterisk **. Required [13] and not covered aspects are: insulation, air-tightness, natural lighting, in-door climate, and building orientation.

^a Atemp is a measure of building floor area specifically developed for EPC in Sweden, and it is calculated differently depending on type of floor space standard. Atemp is defined as the heated floor space including shared spaces and footprints of walls but not including garages [20].

^b BOA and LOA are real estate valuation measures of usable heated floor area for habitation and non-habitation (stair cases, shared areas etc.). BOA and LOA are registered for 90% of the multi-family-dwellings and 74% of the single family buildings.

Table 2
Data frequencies: EPC data, HUD data, Company register and billing data.

Dataset	Boverket (EPCs)	Gothenburg HUD	Riksbyggen
Unique unit	An EPC	A coordinate in polygons	A confined inhabited building
<i>N</i>	17 500	64 600	615
Dwellings	189 000	n/a	13 500
Atemp [m ²]	19 300 000	24 100 000 (BOA)	1 058 000
EPC	17 500	17 500	452
BRF ^a	n/a	n/a	80
Bill paying BRFs ^b	n/a	n/a	53

^a Condominium organizations (BRF). Some BRFs include more than one building.

^b Not all BRFs have contracted Riksbyggen for the payment of heating, water and electricity.

Table 3
Comparison of the two data subsets of different.

Row labels	Atemp derived from BOA + LOA	Measured Atemp
Average of Atemp [m ²]	2848	2180
Average of adjusted Atemp [m ²]	3079	2180
Average number of basement levels	0.697	0.673
Average building base footprint [m ²] (From HUD)	575	457
Average number of stories	4.17	4.06
Average number of apartments	33.0	26.5
Average year of EPC issuing	2008.45	2009.27
Average construction year	1957	1957
Energy usage without adjustment of Atemp [kWh/m ² year]	140.5	128.9
Energy usage with adjustment of Atemp [kWh/m ² year]	130.1	128.9

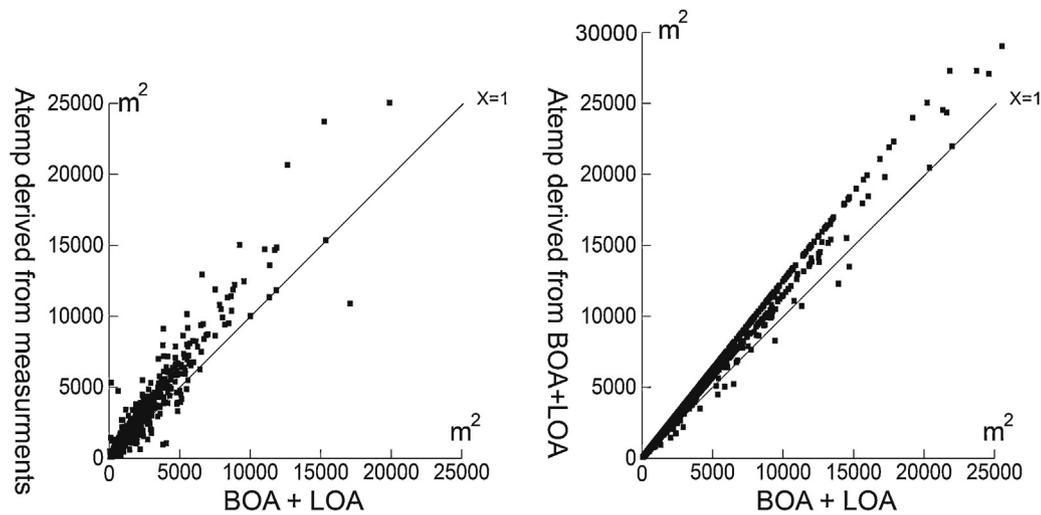


Fig. 1. To the left, Atemp derived from measurements compared with BOA + LOA ($N = 1163$), before removal of outliers. To the right, Atemp derived from BOA + LOA compared with BOA + LOA ($N = 4747$).

Table 4

Description of EPC and comparison with all buildings in Gothenburg (HUD data).

Dwelling type	Boverket (EPCs)	Gothenburg city (HUD)	Overlap of the data sets
Single-family-dwelling	Atemp = 1 660 000 m ² EPC = 11 300	BOA = 8 010 000 m ²	Atemp = 1 160 000 m ²
Row houses	Dwellings = 11 300	BOA = 2 520 000 m ²	Atemp = 617 000 m ²
Multi-family-dwelling	Atemp = 17 500 000 m ² EPC = 6320 Dwellings = 178 000	BOA = 13 600 000 m ²	Atemp = 17 500 000 m ²

of using BOA + LOA to derive Atemp was a method of simplification used later in the EPC issuing process. Atemp for larger buildings is more often derived from BOA + LOA. This is a bias in the applied method.

2.2. Heating, water and electricity billing data

The heating, water and electricity billing data for the 53 BRFs, see Table 2, are from 2012 and 2013. This billing data is compared with the energy usages in the EPC. The different temperature between years were taken into account by adjusting the heat usage to “normal year” by weighting energy used for heating with the degree-days of the month compared with the average degree-days below 17 °C, as done in the EPC. “Normal year” adjustment cannot be done perfectly and it is a source of error. This error also exists within the EPC data since EPC is done continuously over many years. Energy for hot water is calculated from cold water usage with the assumption that 40% of the cold water is heated [31].

A method problem is that the billing data sometimes contained fractions that are not included in the EPC, such as outdoor lighting, engine heaters, etc. This error can also exist for water that is used outside of the building. Heating records are less impacted by this error since all heating is used for heating the building and all is included in the EPC.

In the EPC a separation between energy used by the building and electricity used by organizations in the building is made. In the comparison these two categories are merged since they exist as one measurement in the billing data. This makes it more difficult to assess the data quality of electricity figures in the EPCs.

3. Results

The intention was for all buildings to have an EPC by 2008 [15], but this goal has not been achieved and data collection is

still ongoing. In the Gothenburg building stock 82% of the multi-family-dwellings, 60% of the non-residential buildings, 19% of the row houses and 11% of the single-family-dwellings, are included in the database, see Table 4. A shortcoming is that the EPC alone does not include a distinction between row houses and single-family-dwelling. This distinction exists in the HUD data.

HUD data matched with the EPCs made it possible to separate row houses, and to visually assess the data, see Fig. 2. From comprehensive images data uncertainties were identified and investigated more in detail. In the illustration of single-family-dwelling two problems beyond simple errors in registration were found. In 6 areas in Gothenburg all single-family-dwelling have an EPC. This would make it possible to describe variations in higher detail, but it would be necessary to reduce the weighting of the EPCs when describing the building stock using these 6 areas.

3.1. Atemp analyses

Göransson [28] made one of the most detailed studies of the relationship between measured Atemp and BOA + LOA. Göransson [28] used the ratio X of Atemp to BOA + LOA and introduced the practice of removing buildings with X smaller than 1 and larger than 2. For the Gothenburg building stock 40 of 1163 buildings were removed when this method was applied. Göransson [28] expressed the distribution of the area ratios using quartiles, Q . In Table 5 the standard deviation, σ , is added for comparisons.

Atemp is only derived from BOA + LOA for multi-family-dwellings. Atemp for other residential buildings is measured. In the sample of Göransson [28] the median ratio between BOA + LOA and Atemp is 1.35. This implies that the factors 1.25 and 1.15, recommended by Boverket for deriving Atemp from BOA + LOA, underestimates Atemp of Swedish buildings. In accordance with

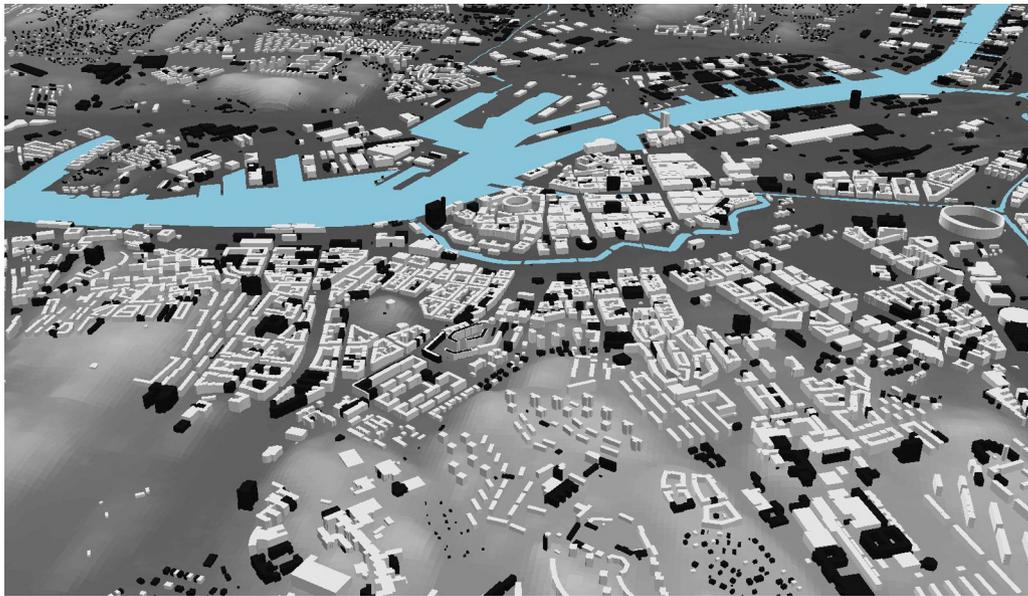


Fig. 2. Illustration of the geographic data of Gothenburg city center. White buildings have an EPC and black buildings do not.

Table 5
Comparison of X in different building groups with different methods of deriving A_{temp} .

Atemp method	Göransson [28]	Multi-family-dwellings			Single-family dwellings	Row houses
	Measurement	Measurement	BOA + LOA	BRA	Measurement	Measurement
N	151	1123	4747	130	4672	3204
Average		1.344	1.212	1.383	1.361	1.171
σ		0.211	0.0493	0.172	0.312	0.229
Q_3	1.44	1.48	1.250	1.483	1.598	1.263
Q_2 (median)	1.35	1.335	1.250	1.371	1.312	1.058
Q_1	1.29	1.181	1.150	1.268	1.059	1.000

this, the average ratio, X , is systematically larger for buildings with A_{temp} derived from measurements, see Table 5.

A_{temp} can be predicted in a more accurate way than the standard method suggested by Boverket by applying a stepwise regression analysis for specific groups of EPCs with measured A_{temp} set as the dependent. Table 6 details the results from a stepwise regression analysis for all Gothenburg EPCs, with the confidence intervals set at 95%, with measured A_{temp} as the dependent and using the other variables available in the EPC as independents.

According to the regression analysis in Table 6 the regression equation can be used to better estimate A_{temp} based on BOA, LOA and number of apartments, A , see Eq. (1).

$$A_{temp_{adjusted}} = 69.8 + 1.152BOA + 1.174LOA + 8.969A \quad (1)$$

However, Eq. (1) is a simplification that does not take variation in building groups into account. Building stock is often described using age groups or building types. Construction year does correlate significantly with A_{temp} but in a non-linear manner. In Table 7

Table 6
Non-standardized coefficients, B , for the models in the stepwise regression analysis which included all EPC variables.

	Model 1	Model 2	Model 3
Constant	145.4	89.28	69.78
BOA [m ²]	1.356	1.303	1.152
LOA [m ²]		1.154	1.174
Number of apartments			8.969
R^2	0.944	0.971	0.979
R^2 change	0.944	0.027	0.008

All coefficients are significant at the 0.001 level (2-tailed).

adjustment factors α have been derived for groups of buildings based on the average ratio X of buildings with measured A_{temp} . As seen in the graph to the right in Fig. 1 and in Table 5 the existence of a basement heated above 10 °C is assumed to be of importance in the method of calculating A_{temp} from BOA + LOA recommended by Boverket.

Eq. (1) and the adjustment factors in Table 7 were applied to EPCs where A_{temp} was derived from BOA + LOA to produce the graph in Fig. 3. These two figures can be compared with Fig. 1.

A building stock is also commonly divided in building types. Göransson measured how X varied in building types. Unfortunately building types are not registered in the EPC or the HUD data. Based on the work of Göransson [28]. Table 8 expresses X for measured A_{temp} and adjustment ratios, α , for groups of different building types.

The average ratio X has changed during the years EPCs have been issued, see Fig. 4. In Fig. 4 data points with less than 30 EPCs have been removed, and 86% of the multi-family-dwellings were issued during 2008 and 2009. During the first years of EPCs issuing there was a practice amongst Energy Experts of setting A_{temp} to equal BOA for single-family-dwellings and row houses. This needs to be taken into account when using EPC data to describe the single-family-dwelling building stock using EPC data.

3.2. Heating, water and electricity billing data analyses

Heating, electricity and heat for hot water are given in used energy [kWh/m² year] in the EPC. The Riksbyggen billing data has been recalculated into energy usage and “normal” year adjustment

Table 7
Adjustment factors, α , making Atemp for different building age groups derived from BOA + LOA comparable with Atemp derived from measurements.

Building age group	N, measured Atemp	N, total population	X, measured Atemp	α for buildings without basement	α for buildings with basement
1900	80	245	1.355	1.178	1.084
1910	25	122	1.309	1.138	1.047
1920	118	494	1.375	1.195	1.100
1930	207	950	1.380	1.200	1.104
1940	170	743	1.451	1.261	1.160
1950	123	884	1.366	1.188	1.093
1960	133	958	1.390	1.208	1.112
1970	52	839	1.259	1.095	1.007
1980	16	268	1.315	1.144	1.052
1990	38	455	1.359	1.181	1.087
2000	65	363	1.327	1.154	1.062

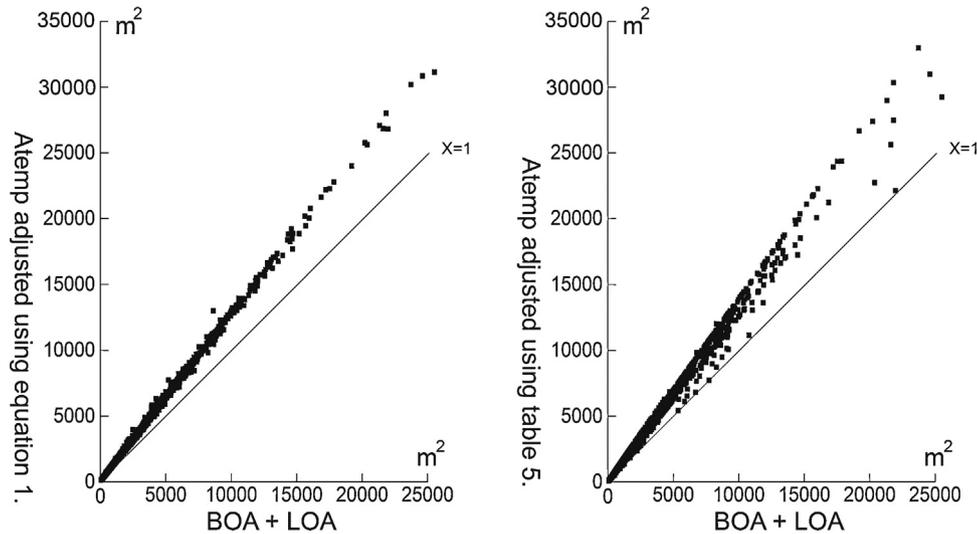


Fig. 3. To the left, Atemp derived from BOA + LOA adjusted using Eq. (1). To the right, Atemp derived from BOA + LOA adjusted using Table 7 (N = 4727).

Table 8
Ratios making Atemp derived from BOA + LOA comparable with Atemp derived from measurements and BOA + LOA for different building types.

Building type	N	X, measured Atemp	α for buildings without basement	α for buildings with basement
Slab block (Lamellhus)	50	1.38	1.2	1.104
Enclosed block (Sluten kvartersbebyggelse)	33	1.33	1.157	1.064
Point block (Punkthus)	31	1.32	1.148	1.056
Slab block (Skivhus)	30	1.31	1.139	1.048
Gallery access block (Loftgængshus)	7	1.42	1.235	1.136

has been made for the heating bill. Fig. 5 illustrates a comparison between annual energy for heating found in the EPC and the billing data. Each dot in Fig. 5 represents one housing organization, BRF. In total 356 EPCs with 9900 apartments and 907 000 m² Atemp are included in this figure.

The two outlier data points below the more well correlated cluster motivates the recommendation to make an outlier analysis of EPC energy usage based on building type. The ratios in Fig. 5 support the usage of an outlier criteria of 2.5 σ for ratios between EPC and billing data. However, depending on the

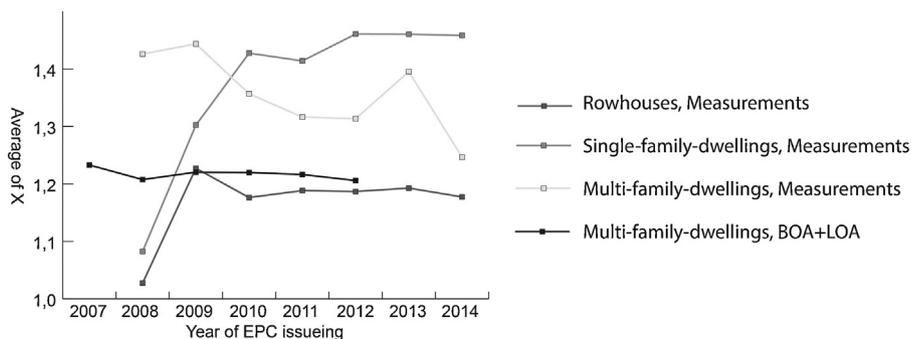


Fig. 4. Average ratio between Atemp to BOA + LOA for dwelling groups divided by the years the EPCs were issued.

Table 9
Gothenburg EPC data quality issue differentiation, result summary and recommendations.

Value category	Data quality issue	Suggested measures when using EPC data to create overview of a building stock
1. Matching, keys, and sorting	Over-representation of row houses with community associations	<ul style="list-style-type: none"> • EPC data is best used in city or district scale where more data on the buildings is available, including graphical observations • Matched HUD data can be used to weight for row house areas that are overrepresented
2. Building characteristics	1% processing error when EPC data was compared with Riksbbyggen data	<ul style="list-style-type: none"> • Building age was described with less error in the EPC than other datasets
3. Building usage		<ul style="list-style-type: none"> • Of the different registered building usages the EPC data contain more detailed usage divided in fractions
4. Building area	Atemp is calculated differently depending on standard methods applied by the Energy Expert	<ul style="list-style-type: none"> • Atemp derived from BOA + LOA should be adjusted using the methods described in this article to be comparable with Atemp derived by other means • Attention should be given to EPCs for single-family-dwellings made before 2008. Atemp was on average larger during this period than later years • Atemp should be compared with BOA + LOA and cases where BOA + LOA is larger or equal to Atemp, or Atemp is twice that of BOA + LOA should be seen as outliers • For single-family-dwellings the year the EPC was made should be adjusted for
5. Heating	When energy usage is distributed on many buildings granularity is lost Efficiency of heat pumps not provided Measurement period not always registered (458 of 15 400)	<ul style="list-style-type: none"> • Energy for heating are found to deviate the least. Outlier outside of 2.5σ in the building types should be given extra attention • Usage of one heating type alone is recommended if possible.
6. Household electricity and water	Energy Experts divide electricity differently Heating for tap water sometimes not measured (1100 of 17 500)	<ul style="list-style-type: none"> • Electricity should be presented separately from heating when describing a building stock • Heating for water below 10 kWh/m² year for multi-family-dwellings should be removed as outliers.

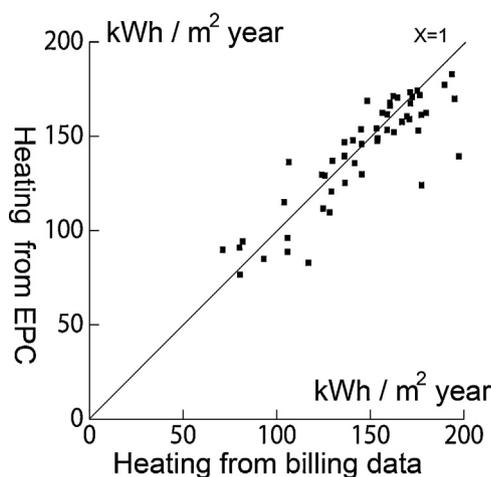


Fig. 5. Comparison of district heating billing data and the data registered in the EPC.

way EPCs are grouped this criterion identifies different EPCs as outliers. Yet, the authors recommends to use an outlier criterion of 2.5σ for further investigation, especially since the heavy energy users are relevant to target to reduce energy usage [32].

The heat usage in Fig. 5 is supplied as district heating. In the EPC data 90% of the multi-family-dwellings in Gothenburg have district heating. A source of data quality problems can be removed if the heating sources are not mixed when describing energy usage in the building stock.

When energy usage is distributed on many buildings granularity is lost. It is not possible to see which other buildings are taking part in the distribution of energy used and the Energy Expert is asked to make assumptions. Assumptions are always needed when there are non-dwelling activities included in this energy to be distributed. Especially electricity usage is complicated to report in a uniform manner, as required in the EPC. Data should be entered for shared building electricity (elevators, lighting), household electricity (domestic), business electricity (e.g. shops on ground level), and electricity for cooling. Energy Experts employed by Riksbbyggen

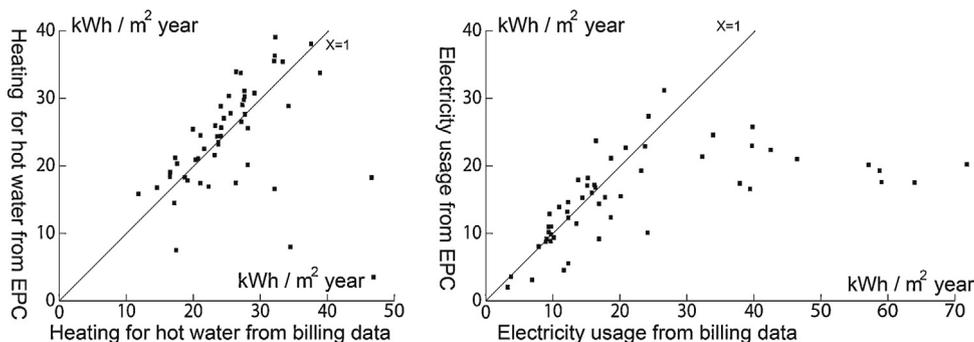


Fig. 6. To the left, comparison of heating for water usage from billing data and the data registered in the EPC. To the right, comparison of heating for total electricity usage from billing data and the data registered in the EPC.

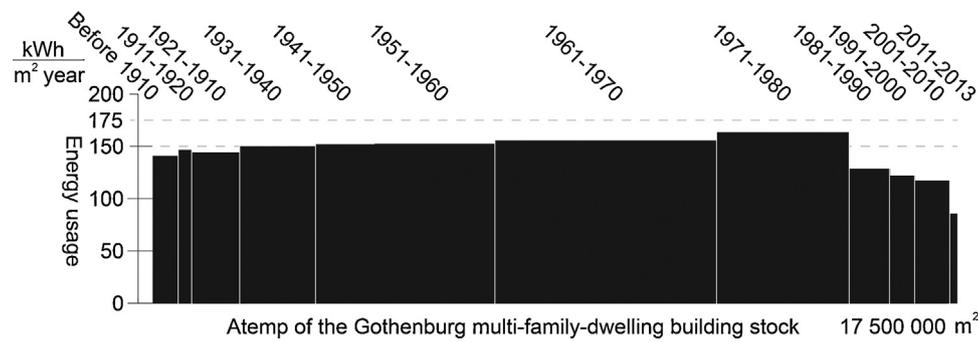


Fig. 7. Energy usage in the Gothenburg multi-family-dwellings divided into age groups with merged heating and electricity usage and without adjusting Atemp.



Fig. 8. Energy usage in the Gothenburg multi-family-dwellings with Atemp adjusted using Table 7 and separated heating (black) and electricity (gray) usage.

reported that uncertainties to which group electricity should be reported and that meters seldom measure only one type of usage so assumptions have to be made. This may lead to data quality issues.

In Fig. 6 the higher billing data electricity use is explained by the BRFs paying electricity bills for activities not included in the EPC (streetlights or car heating), which could not be identified without a site visit. Based on these observations it is recommended that electricity and heating should be presented separately when analyzing energy usage in the building stock.

Billing data for water usage is better correlated with the EPC data than the electricity usage, but there are data points that fall outside of the pattern. This observation further reinforces the recommendation [9] to make a plausibility check and remove outliers. Water usage can vary greatly between dwellings with fewer residents [33], but in larger buildings a more constant water usage, best predicted by the number of residents, can be assumed [34]. The author recommends values below 10 kWh/m² year for hot water heating to be removed as outliers for multi-family-dwellings.

In the exercise of comparing data from the Riksbyggen and EPCs outliers ($\sigma = 2.5$) were identified and given extra attention. Energy Experts employed by the Riksbyggen gave the following reasons for identified deviations.

- In buildings with no private electricity measurements and billing, electricity usage is considerably higher.
- Usage of electricity in laundry rooms is not calculated as contributing to the heating of the building, even though this can be the case.
- Outdoor lighting and engine heating sometimes need to be excluded from the metered energy usage based on assumptions.
- In cases when electricity is measured with one meter for multiple buildings assumptions must be made by the Energy Expert. How these assumptions are made can vary between Energy Experts.

4. Discussion

The data quality issues and ways of handling these issues, based on the analyses in this article, have been summarized in Table 9. Table 9 can be compared with Table 1 in Section 1.1.

In this article factors for adjustment of Atemp derived from BOA+LOA have been presented which makes Atemp larger. According to Eq. (1) the final factors for BOA and LOA are close to the 15% increment suggested by Boverket. One explanation for this is that Boverket based their recommendation on a sufficiently representative sample of the building stock [28]. Another explanation could be errors in reporting regarding the way Atemp was derived. This latter explanation is a methods problem that cannot be excluded.

A larger Atemp will make the energy performance of the buildings seem more efficient. If a building stock description based on the method presented here is compared with building stocks in other countries it will be necessary to adjust the comparison for the country specific building area unit. As a reference for such comparisons Atemp is 34% larger than the Swedish sellable apartment space, BOA. The RESQUE [35] project state that Atemp is 70% of gross floor area. The Energy Performance Certification have mainly been a European project [14], however since 2008 there are initiatives to establish EPCs for buildings in tropical contexts [36].

The EPC data was improved by applying measures described in Table 7. Figs. 7 and 8 illustrate the result of improving the EPC data by adjusting Atemp and separating domestic electricity from heating when applied in a description of the building stock. To be noticed that the integral of Figs. 7 and 8 are the same.

Using the method of finding adjustment factors for groups of buildings has the benefit that the assumption made is clearly defined by being specific to the selected groups. However, the adjustment factors can then only be used to describe the building

stock in accordance with the assumed groups of buildings, as done in Figs. 7 and 8. When a stepwise regression model is used then Atemp is predicted more accurately for specific buildings but the results should still only be used when describing the building stock statistically.

The EPC data contain measured energy usage data and building characteristics for 99% of the multi-family-dwellings from the Million Homes Program era. The Million Homes program was a large national initiative during the period 1961–1975 focused on building one million dwellings to cover an urgent housing need [37]. Buildings from the Millions Homes Program era has been mentioned as a priority for refurbishment and energy usage reducing measures [12,38]. However, the data quality issues motivate an analysis of the data quality before using the data for modeling or describing the building stock.

Building types are sometimes used to separate the building stock when energy usage is analyzed and modeled. The methods of analyzing uncertainties and errors described in this paper can be applied on the chosen building types. Future research can also make use of the data in the EPCs to validate engineering-based models of energy usage. In such analyses the energy usage reducing measures in the EPC would be useful. The data quality of that part of the EPC is difficult to assess without site visits. It is rather recommended that these data quality issues are solved in the EPC issuing process.

5. Conclusion

Data generated in the Swedish EPC process can be used to describe and model the building stock, but there are data quality issues that need to be handled. This article has summarized the data quality issues and has developed ways of improving data quality for different ways of describing the building stock. Building area measures were found to be a parameter that is difficult to use accurately in a building stock description since there are many formats of area measurements that have been used at different times. Analyzing area measures statistically makes it possible to account for the differences when describing the building stock.

Acknowledgements

The authors would like to thank: Hans Nilson at Riksbyggen for technical assistance and helpful discussions, Christer Rosfjord at Boverket for the time spent assembling the Gothenburg EPC dataset and Eric Jeansson at HUD for giving us access to the Gothenburg spatial city plan. This work was supported by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning – Formas, within the Homes for Tomorrow project.

References

- [1] IPCC, Working Group III – Mitigation of Climate Change, Chapter 9 Buildings, 2014.
- [2] EU, Directive 2012/27/EU on Energy Efficiency, 2012.
- [3] Swedish Government, National Program for Energy Efficient and Energy Smart Construction, 2006.
- [4] A. Fouquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: a review, *Renew. Sustain. Energy Rev.* 23 (July) (2013) 272–288.
- [5] N. Fumo, A review on the basics of building energy estimation, *Renew. Sustain. Energy Rev.* 31 (March) (2014) 53–60.
- [6] H. Zhao, F. Magoulès, A review on the prediction of building energy consumption, *Renew. Sustain. Energy Rev.* 16 (August (6)) (2012) 3586–3592.
- [7] A. Mastrucci, O. Baume, F. Stazi, U. Leopold, Estimating energy savings for the residential building stock of an entire city: a GIS-based statistical downscaling approach applied to Rotterdam, *Energy Build.* 75 (2014) 358–367.
- [8] P. Florio, O. Teissier, Estimation of the energy performance certificate of a housing stock characterised via qualitative variables through a typology-based approach model: a fuel poverty evaluation tool, *Energy Build.* 89 (February) (2015) 39–48.
- [9] J. Claesson, Results and Methods Development of Energy Performance Certification of Buildings (Utfall och metodutvärdering av energideklaration av byggnader. CERBOF Projekt 72), 2011.
- [10] S. Stensson, Energy efficiency in shopping malls some aspects based on a case study (PhD dissertation), Chalmers University of Technology, 2014.
- [11] J. Kragh, K.B. Wittchen, Development of two Danish building typologies for residential buildings, *Energy Build.* 68 (January (Part A)) (2014) 79–86.
- [12] Formas, The Million Homes Program Develop or Demolish (Miljonprogrammet – utveckla eller avveckla), 2012.
- [13] EU, Directive 2002/91/EC On the Energy Performance of Buildings, 2002.
- [14] IPEEC, Building Energy Rating Schemes Assessing Issues and Impacts, Building Energy Efficiency Taskgroup, 2014.
- [15] Swedish Government, Statute on Energy Performance Certificates for Buildings (Förordning om energideklaration för byggnader), 2006.
- [16] Boverket, Energy in the Built Environment – Technical Performance and Calculations (Energi i bebyggelsen – tekniska egenskaper och beräkningar), 2010.
- [17] BPIE, Energy Performance Certificates across Europe, From Design to Implementation, The Buildings Performance Institute Europe, 2010.
- [18] P. Cerin, L. Hassel, N. Semenova, Energy Performance and Housing Prices, 2012.
- [19] K.E. Thomsen, K.B. Wittchen, Implementing the Energy Performance of Buildings Directive, Certification Core Theme 1, 2010.
- [20] Boverket, Guide to the Energy Performance Certificate Form, Bl 1189, 2014.
- [21] D. Majcen, L. Itard, H. Visscher, Monitoring dwelling stock efficiency through energy performance register: trends in Dutch social housing, Presented at the World SB 14 “Sustainable Building: Results”, 28–30 October 2014, Barcelona, Spain, 2014.
- [22] L. Murphy, The policy instruments of European front-runners: effective for saving energy in existing dwellings? *Energy Effic.* (2013) 1–17.
- [23] K. Olofsson, Interpret My Interpretation for Truthfulness: A Quality Study of the Energy Performance Certificates Impact on the Value of Commercial Properties (Tolka min tolkning för sanningens skull: En kvalitativ studie av energideklarationens påverkan på värdet av en kommersiell fastighet) (Master thesis), Umeå University, 2013.
- [24] Boverket, Utvärdering av systemet med energideklarationer, 2009.
- [25] M. Bonde, H.-S. Song, Is energy performance capitalized in office building appraisals? *Prop. Manag.* 31 (3) (2013) 200–215.
- [26] M. Wahlström, The Residential Value of Energy Efficient Housing, 2014.
- [27] A. Booth, R. Choudhary, D. Spiegelhalter, Handling uncertainty in housing stock models, *Build. Environ.* 48 (2012) 35–47.
- [28] A. Göransson, Recalculation Between BOA+LOA and Atemp for Multi-Family-Dwellings, Account of Conducted Measurement Work, Profu, 2007.
- [29] L. Liu, B. Moshfegh, J. Akander, M. Cehlin, Comprehensive investigation on energy retrofits in eleven multi-family buildings in Sweden, *Energy Build.* 84 (December) (2014) 704–715.
- [30] A. Bryman, E. Bell, *Business Research Methods*, Oxford University Press, USA, 2007.
- [31] T. Levander, L. Stengård, Metering of cold and hot water usage in 44 households (Mätning av kall- och varmvattenanvändning i 44 hushåll), Swedish Energy Agency, Eskilstuna, 2009, pp. 26.
- [32] R. Galvin, Targeting ‘behavers’ rather than behaviours: a ‘subject-oriented’ approach for reducing space heating rebound effects in low energy dwellings, *Energy Build.* 67 (December) (2013) 596–607.
- [33] M. Pullinger, A. Browne, B. Anderson, W. Medd, Patterns of Water: The Water Related Practices of Households in Southern England, and Their Influence on Water Consumption and Demand Management, Lancaster University, 2013.
- [34] M. Mangold, G. Morrison, R. Harder, P. Hagbert, S. Rauch, The transformative effect of the introduction of water volumetric billing in a disadvantaged housing area in Sweden, *Water Policy* 16 (2014).
- [35] RESCUE, District Cooling Customer Measurement Analysis Verification of and ECI For Market Assessment, 2014.
- [36] K. Wagner, Generation of a tropically adapted energy performance certificate for residential buildings, *Sustainability* 6 (12) (2014) 8415–8431.
- [37] T. Hall, S. Vidén, The Million Homes Programme: a review of the great Swedish planning project, *Plan. Perspect.* 20 (3) (2005) 301–328.
- [38] H. Lind, K. Annadotter, F. Björk, L. Högberg, T. af Klintberg, Sustainable Renovation Strategy in the Swedish Million Homes Programme: A Case Study, 2014.