Energy and Buildings 102 (2015) 328–336

Contents lists available at ScienceDirect

Energy and Buildings

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

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

Mikael Mangold*, Magnus Österbring[1], Holger Wallbaum[2]

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.

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].

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.

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@chalmer.se (M. Österbring), holger.wallbaum@chalmers.se (H. Wallbaum).
[1] Tel.: +46 7022846310.
[2] Tel.: +46 317721994.

http://dx.doi.org/10.1016/j.enbuild.2015.05.045
0378-7788/

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/m^2 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.