5 Actual heating energy savings in thermally renovated Dutch dwellings

Explanatory notes

As opposed to the samples studied in the first three papers, all of which were based on cross-sectional data, Chapter 5 was the first to analyse longitudinal data from the social housing dwelling stock between 2010 and 2013, meaning that the research was narrowed down to dwellings that had undergone renovations in order to see whether the theoretical reduction of energy consumption materialised and to what extent. Since in this sample the dwelling's geometry mostly stays the same, the relation between performance gaps before and after renovations provides important insight into the accuracy of the normalisations used in the regulatory calculation model used in energy labelling. Moreover, a comparison of the actual reductions effected by different renovation measures was made in order to show which renovation practices lower energy consumption most effectively.

Submitted for publication to Energy Policy in October 2015.

Abstract

Since previous research has indicated large discrepancies between the theoretical and actual heating consumption in dwellings, it is important to know what savings renovations achieve in reality. The register of the Dutch social housing stock was analysed, containing dwelling thermal performance information of ca. 2 million dwellings between 2010 and 2013. Renovated dwellings were identified, providing insight into the performance gap before and after the renovation and the actual vs. the theoretical energy reduction of renovation measures. Improvements in efficiency of gas boilers (space heating and hot tap water) yield the highest energy reduction, followed by deep improvements of windows. Improving the ventilation yields a small reduction compared to other measures, however, it is still much larger than theoretically expected. High R and low U values of insulation are well predicted, as well as efficient heating systems whereas low R and high U values, local heating systems, changes from a non-condensing into a condensing boiler and upgrades from a natural ventilation system are not well predicted. The study therefore demonstrated that unrealistic theoretical efficiencies of heating systems and insulation values are causing a part of the performance gap.

§ 5.1 Background

Energy Performance of Buildings Directive is, since its first adoption in 2002, the main policy driver in reducing energy consumption in buildings in Europe. By proposing several actions such as a national performance calculation methodology (Article 3), performance certification of new and existing buildings (Article 11 and 12), cost optimality calculation (Article 5), the directive strives to raise awareness and increase investments leading to an accelerated transformation of the dwelling stock. In May 2010, a recast EPBD was drafted as a response to the more ambitious 2020 targets - 20% reduction of energy consumption and CO_2 emissions set by the commission in 2007 and 2009, respectively. To ensure that the directive is paving the way towards achievement of the set goals, monitoring of the dwelling stock efficiency is paramount on the national and European level to prove whether or not the improvements in efficiency are driving towards the desired targets. Monitoring would thus enable member states and the EU to reflect on the adopted policies and apply amendments where necessary. In 2011, registers of performance certificates were established nationally in 11 member states (Economidou et al., 2011) with the share of dwellings it contains ranging up to 24% in both The Netherlands and UK. For this study, we used a non-public register called SHAERE, which includes the annual performance of almost all dwellings of social housing associations between 2010 and 2013. In The Netherlands the social housing stock represents about a third of the total dwelling stock and is supposed to set nation-wide example for lowering the stock's energy consumption. Each year, the associations record the state of most of their dwellings, including their energy performance in the SHAERE register. SHAERE was set up by AEDES, the national organisation of housing associations, to be able to report the progress of energy renovations and improvement of the energy performance of their stock in relation to the 2020 goals laid down in a covenant with the government and the tenants organisation.

The dataset contained about one million dwellings in each of the four years, thereby offering a great opportunity to get insight into the changing energy performance of the dwelling stock. Previously published research conducted on the mentioned register, analysed the renovation pace of the dwellings between the years 2010 and 2013 (Filippidou et al. 2015a, Filippidou et al. 2015b). This paper, builds upon the findings of those papers by observing theoretical and actual heating energy consumption before and after the thermal renovation, which allows to compare performance gap (difference between theoretical and actual gas consumption) before and after renovation, thereby providing a much needed validation of the current label calculation method. Moreover, the theoretical reductions in dwellings where specific measures have been taken are compared with the actual metered reductions.

This helps establish the highest saving of the most commonly implemented thermal measures and enables a comparison of their effectiveness. The outcomes obtained by using different analysis methods are compared, making the analysis robuster and offering an insight into the accuracy of the methods.

Several definitions are used throughout the paper. *Dwelling properties* include 5 dwelling characteristics: type of space heating installation, hot tap water system, ventilation system, window thermal quality and the quality of insulation of roof, floor and wall aggregated as one variable called the insulation of the envelope. *A renovation measure* is defined as a change in at least one of these 5 parameters from one category into another (the continuous properties for insulation and window quality have been categorised). A *validated renovation measure* is a measure that yields the actual energy reduction comparable to the one predicted. A p*re-label* is a complete thermal recording of the dwelling, including all dwellings energy labels, theoretical heating demand and dwelling properties, which was reported to Aedes at least once in the period 2010 – 2013. *Label registration* is the act of submitting the pre-label data to the government thereby obtaining an official label certificate. *Energy index* is calculated according to the national standards on the basis of total primary energy usage, summing up the energy required for heating, hot water, pumps/ventilators and lighting, and subtracting any energy gains from PV cells and/or cogeneration and finally correcting this sum for the floor and envelope area. *The performance gap* is the difference between (average) theoretical and actual gas consumption of a dwelling or group of dwellings.

§ 5.2 State of the art

The SHAERE register was established in 2010 and includes complete thermal performance of the majority of the Dutch social housing dwelling stock, bringing the much anticipated data required for dwelling stock monitoring. First analyses of this dataset, encompassing over 1,2 million dwellings annually have been conducted by Filippidou et al. (2015a and 2015b). Filippidou et al. (2015a) describes the frequencies of 7 renovation measures as recorded in SHAERE in each available year. According to the author, 35,5% of the dwellings had a change in their energy label, 15% had an improvement of a single dwelling property and 12,7% had a change in more than one dwelling property. The author further breaks down the measures among the 757.614 dwellings which had a change in the energy label (the mentioned 35,5%) and established that 16,8% of the dwellings have improved their label class between years 2010 and 2013 resulting in an increased share of A and B labels (well performing) and decreased share of C-G labels. The remaining 18,7% had a deteriorated label class, which was thought to be a consequence of poorly executed dwelling inspection, which led to re-inspection and recalculation. Another study analyses the Dutch dwelling stock and the measures taken based on a survey of about 4000 representative dwellings (Tigchelaar and Leidelmeijer, 2013) who examine the frequency of various dwelling properties in the samples over the years. Based on the studied sample, however, the energy index of dwellings has improved from 2.09 to 1.89 (label E to label D) in the years 2006-2012, which is comparable to the pace of improvement as described by Filippidou et al (2015b), where the index dropped from 1.81 in 2010 to 1.69 in 2013. The sample analysed in the study by Tigchelaar and Leidelmeijer was relatively large, representative, and not limited to social housing associations. However, unlike the study of Filippidou, it did not follow renovations but samples of representative dwellings in each year. The third study is a national monitoring carried out in The Netherlands (Hezemans et al., 2012) on the basis of surveyed label improvements made in a sample stock of specific housing associations. An assumption was made that by implementing two saving measures (insulation of an envelope part or improvement in installation) coincides with 20% reduction in energy use. In the mentioned years together it was established that about 950.000 dwellings were made 20 – 30% more energy efficient. This monitoring was indirect (the assumption that two measures correspond to 20% energy reduction is a very rough one), used survey and not measured data and analysed relatively small samples which affects representativeness. However, it was the best available at that time and the assumption about two measures coinciding with a 20% reduction has been made due to serious gaps in existing knowledge about actual energy saving of renovation measures.

These three studies delivered information about the thermal measures taken in the housing stock but not on their effectiveness to achieve energy savings. Studying the actual energy savings of thermal renovation measures enables a precise evaluation of renovation strategies and subsequently policy effectiveness. Previous research showed that in The Netherlands, well performing dwellings consume more than expected and that poor dwellings consume up to half less than expected (Majcen et al., 2013a, Majcen et al, 2013b) causing the actual energy savings to be smaller in reality than expected. One of the causes of this performance gap is the fact that theoretical calculations are based on the same normalised conditions (for example average indoor temperature) regardless of the dwelling quality, even though in practice it turns out that the indoor environment differs greatly in poor performing dwellings from the one in efficient dwellings. The gap seems to be difficult to explain statistically, mostly due to the complex nature of the variation in actual gas consumption. However, differences in average indoor temperature and in the quality of estimation of insulation and ventilation flow rates in dwellings of different quality and socioeconomic factors were shown to be important factors in explaining this gap (Majcen et al., 2015). Menkveld studies the relation between the energy saving measure taken and the actual energy reduction using the national energy label database, which is dominated by social housing associations (about 70% of social housing and 30% of private dwellings, Majcen et al., 2013a). However, this study observes cross sectional dwelling data (only one record in time available for each dwelling), comparable also with previous analysis done by Majcen et al., 2013a and Majcen et al., 2013b and Tigchelaar and Leidelmeijer, 2013.

Numerous scientific papers have evaluated individual dwellings operational energy use, such as Adalberth, 1996, Winther et al., 1999, Dodoo at al., 2010, Thormark, 2001. However, as a rule these studies are based purely on theoretical operational heating energy, which as shown before can diverge from the actual consumption by as much as 50% less or 30% more. Karlsson et al. (2006) did base their operational energy consumption on real monitoring data of a reference dwelling, but still based energy calculations for different renovation scenarios on the exact same indoor temperature assumptions, which might not yield realistic results. Small scale projects are usually not that interesting for scientific audience since they lack representativeness and the results shown in non-scientific sources (construction companies, housing associations, even local governments) are likely to be skewed with an emphasis on successful examples.

Therefore, there seems to be a lack of studies analysing the efficiency of thermal renovation measures at the stock level. However, the gap in the literature is understandable since no large scale data about the dwelling stock's energy performance and actual energy use was available previously.

Despite this, an objective and representative evaluation of the undertaken saving measures is paramount in order to evaluate and improve the effect of current retrofit policies. This paper complements the results described above.

§ 5.3 Goal and scope

Using the detailed energy performance register coupled with annual actual energy consumption data gathered by Statistics Netherlands at address level, this paper offers an in-depth insight into longitudinal dwelling stock transformations. By studying a large sample of dwellings that underwent thermal renovation we aim to answer two research questions:

1. What is the actual heating energy saving in renovated dwellings for different thermal renovation measures?

What is the performance gap (difference between theoretical and actual gas consumption) in thermally renovated dwellings before and after the renovation? This way, we can not only provide data on actual energy savings but also offer a validation of the calculation method used to calculate the label. Additionally, the various samples studied (see methods section) will enable a comparison of different analytical approaches. Through the use of these methods we can comment on the usability of SHAERE dataset and provide guidelines for future setup of data registers in different European countries. In the results section we present the first results for the total changes in dwelling performance. Each of the thermal renovation properties is then divided in two sections – B and C. Until now (Majcen et al., 2013b and Majcen et al. 2015), the influence of dwelling properties on actual and theoretical gas use was determined cross sectionally, mostly with the use of both descriptive statistics and regression analysis. Since this is the first study using longitudinal data, section A provides cross sectional statistics of data used in longitudinal analysis (B and C). This enables a comparison of cross sectional and longitudinal analyses and validates the results.

In section A we present the actual and theoretical consumptions of dwellings in different label classes cross-sectionally, in the whole available dwelling stock in year 2010 (first available SHAERE record) and 2012 (last useful SHAERE record). This is done in order to place the results among the existing literature on the subject, since existing studies of the performance gap have invariably focussed on cross sectional data. Moreover, this first section gives an idea how the total thermal performance of the whole stock changes through time (how many label changes there are and how much energy consumption changes in each label class).

However, the core of the paper is the efficiency improvement of the dwellings and the actual energy savings following thermal renovations, therefore in parts B and C of the results we select only dwellings which have undergone changes and analyse the theoretical as well as actual reduction of energy consumption before and after renovation. In section B, all dwellings having a change in one specific dwelling property are studied, regardless of whether the other properties have changed or not. This may seem illogical, but in the past, such an approach was applied often in order to obtain significant results despite the small sample sizes. In section C, the dwellings having only a change in this specific property while all others are constant, are studied.

In the methodology section which follows, the process of data handling and subsample selection is outlined and the way of dealing with the data accuracy is explained. The results are presented separately for each examined dwelling property (space heating, hot tap water, ventilation, window quality and insulation). In the discussion section we first compare the three different methods, followed by a discussion of trends noted regarding the effectiveness of different thermal renovation measures, the performance gap and the validation of the calculation method.

TOC

As the Netherlands have an oceanic climate with cool summers and moderate winters, most of the energy consumption comes from heating demand. Natural gas is used as a source of heating in most Dutch dwellings and therefore also label certificates express heating energy consumption in m^3 gas. The actual consumption data is available at Statistics Netherlands in the same units, which is why we chose to study gas consumption as a measure of dwellings thermal performance. This means, however, that the dwellings that make use of electrical installation systems (e.g. heat pumps) were excluded from the analysis.

§ 5.4 Methodology

§ 5.4.1 Dataset properties

The SHAERE register is a raw, full export of the entire energy performance certificate calculation according to the Dutch standard (ISSO, 2009) on the level of dwellings for each year from 2010 on. The data differs significantly from the certificate data stored by the Ministry of the Interior and Kingdom Relations of The Netherlands (label certificates registered by the authorities as used in the studies by Majcen et al. 2013a and 2013b), since it includes all detailed properties required for the calculation of the energy label. However, the data in SHAERE does not consist of registered label certificates, but of so-called pre-labels. A pre-label is a label certificate of a dwelling that may have not been registered at the authorities yet but has nevertheless been recorded internally by a housing association. According to Aedes, pre-labels are updated whenever a renovation measure takes place and are considered accurate because housing associations report to use these pre-labels as an asset management tool (Visscher et al., 2013). Aedes provided the data from 243 Dutch housing associations (in 2011 there were a total 289 associations in The Netherlands) in years 2010, 2011, 2012 and 2013. It is important to note, that social housing represents 33% of the Dutch dwelling stock (Energiecijfersdatabase) and even though some properties differ with the private sector (Majcen et al., 2013a) such a larger sample does offer a great deal of representativeness. The database included dwellings geometry, envelope and installation system characteristics (including detailed information on the quality of insulation, ventilation and heating and hot tap water installation), as well as the theoretical heating energy consumption calculated according to the Dutch ISSO standard (ISSO 82.3, 2009).

In the present paper the dwelling data is available pre-and post-renovation (also called longitudinal data), which probably greatly decreases the variance between groups due to the changes in conditions we do not control for (different household and occupant properties in different groups etc.).

§ 5.4.2 Variable extraction

From the MSSQL SHAERE database, the tables about dwelling information, heating and hot tap water installation information, ventilation and envelope characteristics were merged for analysis, based on the dwelling ID. The type of each construction element (floor, roof, wall, window or door), area, U-value (heat transfer coefficient for windows) or R value (thermal resistance for all other constructions) is known.

To simplify the analysis we computed the average R value for the whole envelope and U value for windows using the formulas below using basic thermodynamic principles.

Insulation values for floor, roof, wall, windows and doors were available as continuous values. To simplify the detection of changes in insulation in between years, these variables were discretised into a finite number of categories. We first considered using the commonly encountered categories of insulation (as described in the Dutch standard ISSO 82.1), but since this yielded distributions highly dominated by the average value, we rather decided to rank the data into 10 categories and use the top and bottom value of each rank class as a basis for the category. We aimed for 10 categories within each label (each containing 10% of records). That way we capture more changes than by using the commonly used insulation groups. The categories are described in Table 1. The categories for R-value may seem to have strange ranges: the maximum R-value is 1,36 which is relatively low. One should keep in mind that an old Dutch dwellings may often have an R-value of 0,19 and insulation is generally brought only on a part of the house (e.g. the roof only or the wall between the window and the floor only) leading to average values that are still low.

The heating installation systems were all gas powered. The least efficient system (η=65%) is a local gas heater, where local means that the heater – a gas stove - is situated in one or two places in the apartment, most commonly the living room. The rest of the bedrooms are in this case not heated. An upgraded version of this system is a gas stove that is used to also heat the bedrooms, this is the gas heater with efficiency between 65% and 83%, regarded as η<83%, this kind of heater is non-condensing. A conventional non-condensing boiler has an efficiency between 83 and 90%, in named in this paper as η>83%. And several high(er) efficiency condensing boilers with efficiencies of 90, 94 and 96%, are referred to as η>90%, η>94% and η>96%.

The heaters for hot tap water are similar, in most cases the heater for space and water is combined, and in cases where it is not combined, the households use a tankless gas boiler for water heating. The methodology predicts several water efficiencies of water heaters – conventional ($\eta<83\%$), improved ($83\%<\eta<90\%$) and high efficiency condensing boiler ($\eta>90\%$).

Regarding ventilation, most dwellings in The Netherlands only have natural ventilation. In the data we also encountered several types of mechanical ventilation, such as, central mechanical exhaust, central demand controlled mechanical ventilation (DCV) controlled by CO_2 sensors, mechanical balance ventilation with heat recovery, decentralised mechanical ventilation with heat recovery, and finally, demand controlled decentralised mechanical exhaust ventilation.

R ENVELOPE EXCLUDING SURFACE RESISTANCE [M^2K/W]	CATEGORISED R VALUE	U-WINDOW [W/M^2K]	CATEGORISED U-VALUE
-0.19	R10	/	
0.19-0.21	R9	/	
0.21-0.25	R8	>4	U8
0.25-0.28	R7	3.7-4.0	U7
0.28-0.34	R6	3.1-3.7	U6
0.34-0.45	R5	2.93-3.1	U5
0.45-0.68	R4	2.9-2.93	U4
0.68-1.01	R3	2.6-2.93	U3
1.01-1.36	R2	1.8-2.6	U2
1.36-	R1	≤1.8	U1

TABLE 1 Categories of insulation values used)

§ 5.4.3 Sample selection

In theory, all dwellings should be pre-labelled and reported to Aedes each year, therefore ideally, each dwelling would have one record for each year of observation starting with 2010 up to 2013, adding up to four records. However, due to several reasons such as changes in associations reporting on the stock (some may cancel or start their cooperation with Aedes), purchases and/or sales of dwellings and demolition and new construction many dwellings have less than 4 records. In principle, more and more dwellings are pre-labelled and reported each year, since more associations decide to participate and the reported dwellings stock continues to grow. If one dwelling had several records in one given year and in case all dwelling properties were equal, we deleted the copies to leave only one record per dwelling. In some instances, not all properties were identical in both records and in that case we deleted both cases as we could not determine which one is more recent (the only time reference in the database is the year of the pre-label, no day or time stamp is available). After deleting those, our dataset was reduced from the initial 5.205.979 to 4.612.020 cases over four years.

After examining frequencies it became clear that the dataset contained a number of dwellings with an unrealistically small or large floor area. Therefore cases where floor area is below 15m^2 and above 500m^2 were deleted, resulting in a further reduced sample of 4.606.749 cases.

Most Dutch dwellings are heated by gas, and in the SHAERE sample almost 90% of the dwelling records (over all four year together) had a gas-powered hot tap water system and 93% had a gas-powered heating system. The rest of the dwellings utilize either district heating (4%) or electricity (6%) for hot tap water and about 7% of the space heating installations are electrical systems. District heating systems had to be removed due to the inaccurate actual annual consumption data for such installations. Electrical heating systems, mostly heat pumps, have been omitted to keep the scope limited and results more accurate. Removing non-gas based and collective systems left us with a sample of 3.729.256 reported pre-labels and further deletion of non-independent dwellings (student rooms, rooms in elderly homes etc.) resulted in a dataset of 3.728.143 pre-labels. As the actual energy consumption data from Statistics Netherlands was not yet available for the year 2014, we narrowed the sample further to the period of 2010 – 2012, resulting in 2.726.600 pre-label reports. For the measures that were taken in 2013 we would namely not be able to find a corresponding actual consumption (see also further in this section).

The actual energy use data provided by Statistics Netherland is collected from the energy companies, which base it on the annual meter readings done by the occupants. The data is therefore sometimes missing and averaged on the basis of similar households and sometimes an extrapolation of monthly values (if the reading are less than a year apart). This can cause inaccuracies that have already been discussed in previous papers (Majcen et al. 2013a, Majcen et al. 2013b, Majcen et al. 2015). The actual gas consumptions were corrected with degree days of the theoretical gas use (Majcen et al. 2013b).

Three types of subsamples were used in order to demonstrate trends with as much accuracy as possible. The abovementioned SHAERE sample of 2.726.600 reported pre-labels corresponds to 1.234.724 individual dwellings. In this dataset, every dwelling contained one or several pre-labels. The number of pre-label certificates from different years is gathered in Table 2.

2010 only	93.797	8%
2011 only	104.959	9%
2012 only	126,599	10%
2010 and 2011 only	151,467	12%
2010 and 2012 only	64,140	5%
2011 and 2012 only	111,255	9%
2010, 2011 and 2012	582,507	47%
Total	**1,234,724**	**100%**

TABLE 2 Number of dwellings having a pre-label in a given year

A Performance gap in the total stock

To show what changes occurred in the social housing stock data globally (section A of the results see Goal and scope), we first analysed the entire sample by coupling it with the corresponding annual actual gas consumption on address level (pre-labels from 2010 were coupled with 2010 actual gas data, 2011 pre-labels with 2011 and so on...). Reports with missing actual gas data were removed using outlier thresholds of 15 and 6000 m^3 gas (Table 3) per year. Part A analyses the theoretical and actual gas consumption in all pre-labels at the end of 2010 (835.313 pre-labels remained after the 891.911 total records were coupled with actual energy use) and in all pre-labels at the end of 2012, which includes also the years prior (1.152.320 coupled records out of the 1.234.724 total data, see Table 2). This means that for 2012, only the latest reports were taken into account. If there are no labels in 2011 and 2012 for example, we assume that there was no modification to the 2010 situation. In this section we compare all available records in 2010 and 2012, meaning that the dwelling that we observe are not identical (nor is the size of the sample). However, this gives a good idea of the changes made in SHAERE dataset globally over the years.

However, a sample of 835.313 (2010), representing 35% of the total social housing stock, can be considered to be well representative. Former studies (Majcen et al. 2013a, Majcen et al. 2013b) were based on such samples. The sample from 2012 is even more representative (ca. 50% of the stock). Therefore, under these assumptions of representativeness a comparison between 2010 and 2012 should lead to valid results about the changes in the dwelling stock.

YEAR	2010	2012
Total pre-label reports	891,911	1,234,724
Valid actual consumption data	835,313	1,152,320

TABLE 3 Pre-label reports with available actual gas consumption data

B Dwellings with a change in at least one dwelling property

In this section, dwellings with at least two pre-labels (sum of row 4 till 7 in Table 2) were selected, in total they amount to 909.369 dwellings. Due to missing actual gas consumption data and the fact that some categories contained less than 30 dwellings (which leads to high 95% confidence intervals and low statistical significance), the sample was reduced to 644.586 dwellings. Sample B is for each property, a subsample of these 644.586 dwellings. For instance, when studying changes in space heating and hot tap water, all dwellings with an improvement in space heating between the first and the last pre-label were selected, leading to a sample of 79.241 dwellings (Table 4). For dwellings with more than two pre-labels, the first and the last one were selected. Since dwelling observations were annual, last actual gas consumption before the first pre-label report year was used as baseline and the first available consumption data after the last pre-label report year. For example, for dwellings having the first pre-label report in 2010, gas data from 2009 was used and for dwellings having their last pre-label report in 2012, gas data for 2013 was used. Another condition was that both actual and theoretical consumptions have to be valid before and after the renovation (between 15 and 6000 m^3).

As Table 4 shows, the database reveals that some of dwellings in the sample have improved, most stayed the same and a fraction even deteriorated. Since all stock should be reported each year, it is logical that a large fraction remained unchanged as most dwellings do not undergo any change. Deteriorations are more surprising at first sight, but appear to occur due to a re-inspection of dwelling leading to a re-calculation of the label. This occurred due to changes in the inspection procedure or faults in the first inspection. All three installation variables observed have rather few deteriorations – between 1 and 2% whereas insulation values have slightly more (Table 4). Since we suspect these are administrative corrections, we do not show these changes in the graphics and consider only the improvements.

	LABEL CHANGES	SPACE HEATING AND HOT TAP WATER	VENTILATION	U-VALUE WINDOWS	R-VALUE ENVELOPE
Deteriorations	5%	2%	1%	6%	10%
No change	78%	87%	95%	77%	74%
Improvements	17%	12%	4%	18%	15%
Total sample size A	835,313 cases for 2010 and 1,152,320 for 2012				
Total sample size B	109,278	79,241	25,783	116,025	96,688
Total sample size C	/	30,749	4,866	15,744	21,035

TABLE 4 Share of improvements and deteriorations of various dwelling properties and sizes of analysed subsamples

C Dwellings with a change in only one dwelling property

The drawback of the sample selection in the previous paragraph is, that a change in for example heating installation system doesn't mean all other dwelling parameters remain constant. In fact, in most cases, more aspects of the dwelling have changed. In section C renovated dwellings were selected like in section B, but in addition all dwellings having more than one property changed were eliminated, meaning that dwellings have one and only one property changed. Categories with a number of records below 30 were discarded and Table 4 shows the amount of dwellings observed. While the samples in this section are much smaller than in section B, they offer valuable results about the effect of one single measure, which have to our knowledge not been previously described in scientific literature.

§ 5.4.4 Uncertainties

There was one difference between the end uses of theoretical and actual gas consumption, which is gas used for cooking. Actual gas consumption takes it into account and theoretical does not. However, cooking constitutes less than 2% of total gas consumption and it should therefore not affect the results too much.

In the section before, we showed that deteriorations of properties were observed in a small part of the sample (1 to 10%) due to re-inspection and re-calculations. We cannot exclude a comparable amount of improvements being caused by re-inspection and re-calculations rather than by real improvements. This will be taken into account in the analysis of the results. Moreover, also degree days calculation applied to actual gas consumptions (see section 5.4.3) and socioeconomic factor could influence the results (varying household size or composition, economic crisis, changing energy source for cooking etc.). To test these impacts, a control group consisting of unchanged dwellings was studied. Dwellings with 4 pre-label reports (497.088 dwellings) were selected out of the 2010-2013 SHAERE database containing 3.728.143 cases, after removing dwellings with missing actual gas data. From these 497.088 dwellings only the ones which had identical theoretical gas consumption four times were selected. These dwellings had no changed in any of the properties considered in this paper. This subsample contained 15.602 dwellings where no renovation measures took place. Table 5 shows a slight decrease of actual gas consumption of about 1,6% annually. In the identified sample of 15.602 dwellings their standardised actual gas use has decreased with 3,6% in years 2010 – 2013, which means that energy savings below 38 m^3 should not be considered as real improvement but as background noise. The numbers of degree days in the studied years were 3321, 2622, 2879 and 3078 from 2010 up to 2013.

YEAR	2010	2011	2012	2013
Average actual gas use [m³/year]	1054*	1034*	1017*	1016*
Average theoretical gas use [m³/year]	1113	1113	1113	1113
Gas reduction relative to 2010 [m³/%]		20 [1,9]	37 [3,5]	38 [3,6]

*The differences in actual consumption between the four years are significant on a 95% confidence interval.

TABLE 5 Reduction in actual gas consumption between 2010 and 2013 in non-renovated dwellings (N=15,602)

§ 5.5 Results

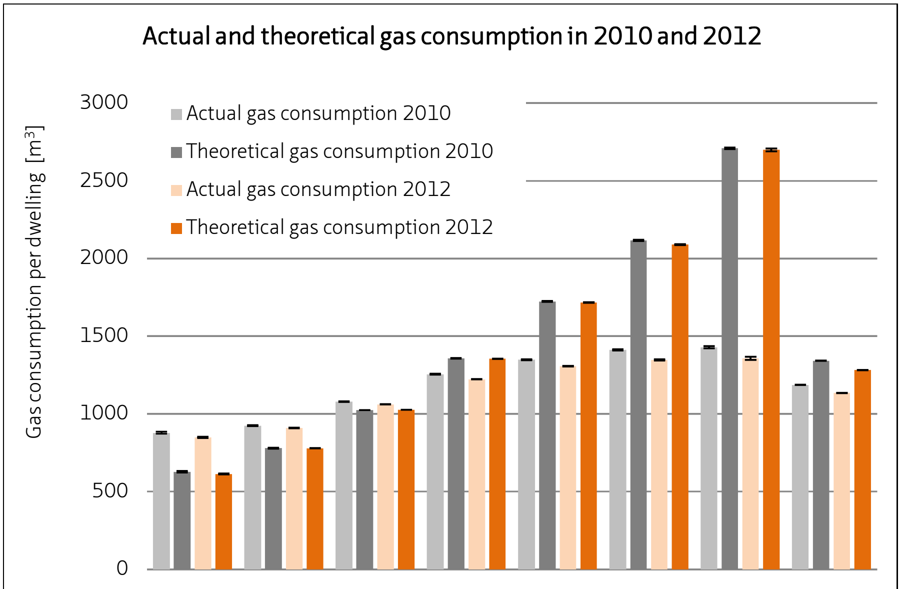
For an easier overview, the results are shown in sections 5.5.1 for label calculation and 5.5.2 to 5.5.7 per renovation measure. Section 5.5.1 consist of part A and B and later sections consist of B and C (like described in methodology section). Finally, section 5.5.8 compares the actual reduction of different measures investigated with method C and comments on their performance gap.

The results are presented in m^3 gas consumption per dwelling and not per m^2 floor area, since previous research demonstrated that although there are some slight differences in average floor sizes of dwellings in different label classes (size of A labelled dwellings is on average 105 m^2 and in other label classes the size is between 90 and 96 m^2), the performance gap does not change significantly whether observed per m^2 dwelling or not (Majcen et al., 2013). Furthermore, samples B and C represent renovations, therefore the floor area remains constant.

§ 5.5.1 Total thermal performance of the dwellings – comparison of label categories

This section shows the actual and theoretical reduction of dwellings which had their energy label improved, meaning that their total energy performance is observed.

A Total stock recorded by SHAERE



		A	B	C	D	E	F	G	TOTAL
N	2010	16295	100718	265625	224814	117435	77767	32659	835313
	2012	36553	173351	383472	286707	145280	89387	37570	1152320
F	2010	2%	12%	32%	27%	14%	9%	4%	100%
	2012	3%	15%	33%	25%	13%	8%	3%	100%

FIGURE 1 Actual and theoretical gas consumption in 2010 and 2012 with 95% confidence intervals

Figure 1 reveals that while actual gas consumption drops from year 2010 to 2012 within all label categories (a drop between 14 and 71 m^3), theoretical remains more constant (the drop in range of +3 and -26 m^3) whereby the differences between theoretical consumption in categories B, D and G are also not statistically significant. Very similar results can be found in previous studies on this subject, using different samples (Majcen et al. 2013a, Majcen et al. 2013b, Majcen et al. 2015), with the results being comparable in terms of annual trends as well as the performance gap across categories. Overpredictions occur in labels D to G and underpredictions in the rest of the categories. While a difference of 2086 m^3 gas can be noted when comparing theoretical consumptions of category A and G, the difference in actual consumption is a mere 508 m^3, almost 4 times less. The difference between the two consumption of category F and G is the most drastic, 609 m^3 for theoretical and only 10 m^3 for actual consumption. Despite the changes noticeable in the performance gap in different label categories, there is only a slight decrease in the performance gap of the total sample – from 156 m^3 in 2010 to a 148 m^3 in 2012. This is because of the increasing number of better performing dwellings and the decreasing number of poor performing dwellings.

The frequencies of label classes change throughout the years (table below Figure 1). Frequencies of well performing dwellings (labels A – C) have increased (in total from 46% to 51%) and there are fewer D – G labels (from 54% to 49%).

Figure 1 shows that in total both average actual and theoretical gas use are lower in 2012 than in 2010. The absolute difference in the actual gas use is 52 m^3 and in the theoretical gas use it is even higher, 60 m^3. The theoretical reduction is a reflection of an improving dwelling stock (as said before, the frequency of good labels is increasing) and the actual gas use probably partly reflects that as well, however, the 38 m^3 background reduction should be disregarded leading to an actual gas reduction is only 19 m^3, three times less than expected. These 19 m^3 are either due to a different sample (many new dwellings were added) or performance improvements within one label category.

B Dwellings with a change in label class

For the results in this section, the sample of 644.586 records described in 4.3 was used. To show how this sample relates to the one in 5.5.1.A, gas consumptions in 2012 of both samples, A as well as B are plotted on Figure 2. Even though the confidence intervals are not plotted for better readability, the differences in consumptions between the two samples are negligible (not significant). This means that in terms of actual and theoretical consumptions on average, sample B is representative for sample A (which is larger).

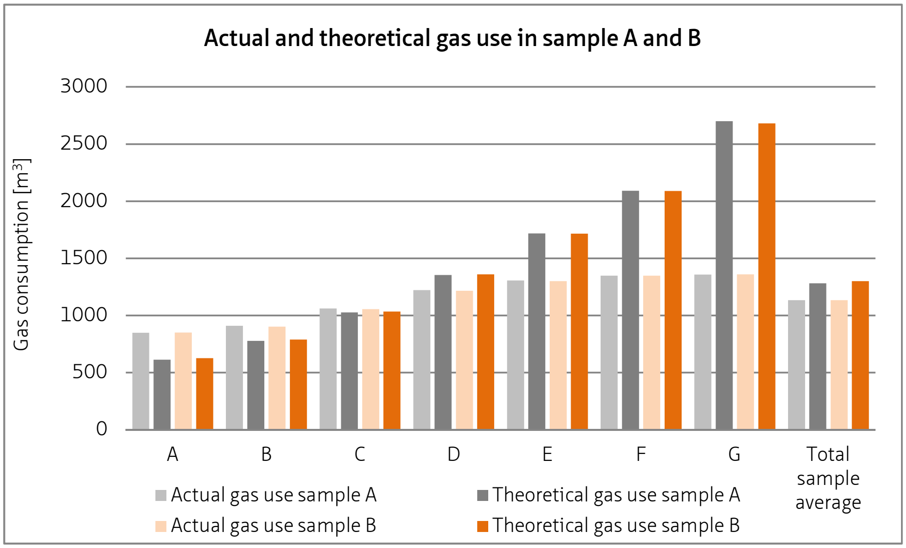


FIGURE 2 Actual and theoretical gas use in sample A and B

Table 6 shows the actual and theoretical gas reduction between years 2010 and 2012 in the selected sample (dwellings having a change in label class). As shown in Table 4, majority of the dwellings that had two labels reported in this period did not change label class, 17% has been improved and 5% have deteriorated. In this section, we focus on the sample of the 17% that have improved showing the actual and theoretical reduction in each of the label changes together with the ratio between them.

When looking at the changes in actual and theoretical gas consumptions at the time of first and second label (sample B) two possibly related phenomena can be noted (Table 6). Firstly, the actual improvement corresponds with the theoretical the best in dwellings that were well performing already before the measure (for example the improvement B to A has a theoretical reduction of 125 m^3 and actual of 129 m^3 whereas a dwelling that went from F to E has a theoretical reduction of 374 m^3 and actual of 136 m^3). Secondly, smaller improvements seem to be better predicted than deep renovations (for example B to A or C to B achieve 103 and 95% of the expected theoretical reduction), while F to B achieves only 27%. Renovations of very poor performing dwellings such as G or F result achieve a smaller % of theexpected reduction, 36% when improving from F to E and only 26% when improving from G to F. Renovating such poor dwellings to an even higher standard is even less well predicted (G to A dwellings achieve 21% (2075 m^3 theoretical and 446 m^3 actual) and G to F realize 26% of expected savings (508 m^3 theoretical and 133 m^3 actual). However, the absolute values prove that deep renovations nevertheless yield a higher saving in m^3 than minor renovations. These findings are in line with previously mentioned cross sectional studies.

Table 6 also shows the comparison of cross sectional data (section A) vs. the longitudinal data (section B) for renovated dwellings where the label class has changed. Whereas relatively comparable results were obtained when observing larger changes in thermal performance (more than 2 label classes), in changes for only one or two classes (A to B or G to F and G to E) cross sectional methods (section A) seem to strongly underestimate the actual gas saving (G to F 133 m^3 vs. 10 m^3). Longitudinal data (section B) results in actual reductions larger than those of cross sectional data. Dwelling characteristics, which correlate with a particular label class in sample A (for example, more apartments efficient label classes) whereas poor label classes are more dominated by detached and row houses, present a limitation of cross sectional data use, as they cause a comparison of two entities that are essentially very different. Moreover, there is a possibility that behaviour and lifestyles of the occupants in cross sectional data are different in different label classes. The longitudinal data on the other hand, assures that the same dwellings are compared before and after the renovation, which reduces these uncertainties. The occupants could still have moved during this time, but this probably only happened in a fraction of the dwellings (whereas in cross sectional data, the occupants are always different). Moreover, the performance gap expressed as the ration between the actual and the theoretical gas consumption in generally much smaller in sample B than in sample A.

	DWELLINGS WITH AN IMPROVEMENT OF LABEL CLASS, SAMPLE B				WHOLE DWELLING STOCK STATISTIC, SAMPLE A				
	ACTUAL [M³]	THEO-RETICAL [M³]	N	RATIO ACTU-AL/	ACTUAL [M³]	THEORETICAL [M³]	N BEFORE	N AFTER	RATIO ACTUAL/
G to F	133	508	3,576	0.26	10	609	37,570	89,387	0.02
G to E	153	846	2,090	0.18	51	983	37,570	145,280	0.05
G to D	215	1,415	934	0.15	135	1,345	37,570	286,707	0.10
G to C	301	1,742	730	0.17	297	1,672	37,570	383,472	0.18
G to B	354	1,871	348	0.19	449	1,921	37,570	173,351	0.23
G to A	446	2,075	78	0.21	509	2,086	37,570	36,553	0.24
F to E	136	374	2,090	0.36	41	373	89,387	145,280	0.11
F to D	135	674	934	0.20	125	735	89,387	286,707	0.17
F to C	227	1,091	730	0.21	287	1,063	89,387	383,472	0.27
F to B	371	1,379	348	0.27	439	1,312	89,387	173,351	0.33
F to A	510	1,688	78	0.30	499	1,477	89,387	36,553	0.34
E to D	127	323	934	0.39	84	362	145,280	286,707	0.23
E to C	187	626	730	0.30	246	690	145,280	383,472	0.36
E to B	342	920	348	0.37	398	938	145,280	173,351	0.42
E to A	392	1,107	78	0.35	458	1,104	145,280	36,553	0.42
D to C	150	242	730	0.62	161	328	286,707	383,472	0.49
D to B	217	473	348	0.46	313	577	286,707	173,351	0.54
D to A	318	718	78	0.44	374	742	286,707	36,553	0.50
C to B	157	165	348	0.95	152	249	383,472	173,351	0.61
C to A	137	310	78	0.44	213	414	383,472	36,553	0.51
B to A	129	125	2,499	1.03	61	165	173,351	36,553	0.37

*The orange highlights signify a more than twice as high ratio of method B compared to method A.

TABLE 6 Actual and theoretical heating energy savings corresponding to different label steps made

§ 5.5.2 Space heating and hot tap water

This section shows the actual and theoretical energy reduction in dwellings which had an improvement in the space heating and hot tap water installation. The two systems are viewed together despite the fact that in SHAERE database, these were two separate variables. However, during the preliminary analyses many illogical combinations of space heating and hot tap water were observed, such as a combined high efficiency hot tap boiler together with local gas heater. Such an installation is impossible in practice, since 'combined' boiler means that it is used also for heating. Because of this hot tap water and heating were analysed together, only looking at the dwellings with a logical combination of the two systems. Furthermore, for better readability we only show the results for dwellings which had an improvement in both heating and hot tap water systems and not just in one. To ensure statistical significance, groups with less than 30 cases are omitted from the figures.

A Dwellings with a change in heating and hot tap water system

The most common change among this measure in the observed sample is the replacement of the space heating boiler from improved $\eta>83\%$ efficiency boiler to a condensing boiler with $\eta<96\%$ efficiency and at the same time changing a combined improved (CI) efficiency tap water boiler with high efficiency (CH) one (this is in fact one system, last column in Figure 3). More than half of the studied dwellings within this measure have undergone such a renovation which makes this result very robust. The actual reduction is about two thirds of the theoretical.

Decrease in gas consumption is much smaller than expected in most dwellings with renovated heating and hot tap water systems (Figure 3). Roughly, the results can be divided into two groups, one group being the dwellings with a hot tap water boiler improved from an on-demand tankless boiler to a combined boiler and the other group where hot tap water combined boiler has been improved in efficiency (five last columns of Figure 3). In the first group, the difference between theoretical and actual reduction is in general larger than in the second group. If we look at the changes in heating installation, there seems to be few correlation between the extent of efficiency improvement and the actual gas reduction. Changing a local gas boiler has an actual gas consumption far below the theoretical. A pattern can be detected if one keeps in mind that boilers with efficiencies $\eta<83\%$ and $\eta>83\%$ are non-condensing and other boilers ($\eta<90\%$, $\eta<94\%$, $\eta<96\%$) are condensing. It seems that changes towards a higher efficiency within the category of non-condensing boilers are well predicted (second and tenth column). Similarly, also improvements in efficiency within the category of condensing boilers are reasonably well predicted (eighth and ninth column). In the changes of efficiency between non-condensing and condensing group the predictions are worse (in the last four columns). In some cases the reduction in actual consumption seems to be negative despite the large theoretical reduction (fourth column). It could also be that such group contains a complex of apartments which were recorded at the same time and contain a systematic error.

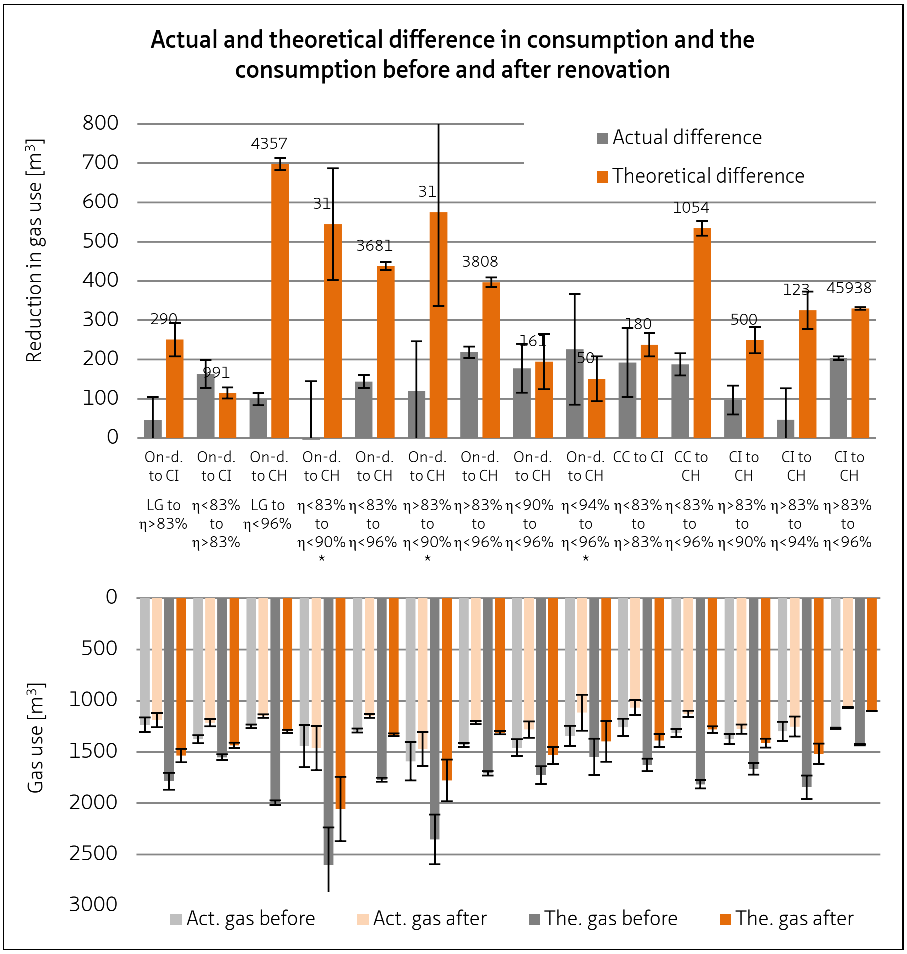


FIGURE 3 Actual and theoretical reduction and number of cases (parenthesis upper figure) and consumption before and after renovation (below) in dwellings with a renovated hot tap water and heating installation system – sample method B (N>30). On-d.= on-demand tankless boiler, CC/CI/CH = combined conventional/improved/high efficiency boiler , LG = local gas heater * This group is not present in the results of section C.

Figure 3 does not show whether for example, a high actual performance gap is more a consequence of poorly predicted consumption before after the implementation of the measure. This means in practice, that one cannot tell whether a low performance gap is indicative of a low performance gap after renovation, which was observed previously by Reynaud (2014). Therefore consumptions before and after are plotted on the bottom of Figure 3. It seems that the dwellings which were poorly predicted before the renovation remain poorly predicted after renovation, however, on average all dwellings seem to be better predicted after the renovation, which is in accordance with the previously noted fact that better performing dwellings are better predicted. Also, the heating systems with η<96% efficiency seem to be well predicted (see light orange and light grey bars), especially where the samples are larger (third and last column). Moreover, it shows once again that in cases where local gas stove was changed with a more efficient system, the gap generally decreases - this is due to mentioned ill-assumption of heated floor area in case of local gas heaters that was shown in (Delghust et al. 2015 and Majcen et al. 2013b): in houses with local gas heaters, generally only one or two rooms are heated, whereas the calculation are based on heating of all rooms. In general, the more efficient the heater the better the prediction.

B Dwellings with a change in only heating and hot tap water change

In this sample of 30.749 cases only the heating and hot tap water installation had changed according to the information in SHAERE database. Among all the studied measures, heating and hot tap water have the most similar samples in section B and C (79.241 and 30.749 dwellings), which means one can expect the most comparable results: when the heating system is changed, there are usually no other measures taken. The difference between the theoretical and actual reduction seems slightly less drastic (see last column of Figure 3 and Figure 4) but despite from that, the results are indeed comparable.

It seems that dwellings are again better predicted after renovation than before. Visually, there does not seem to be a correlation between the size of the performance gap before and after the renovation.

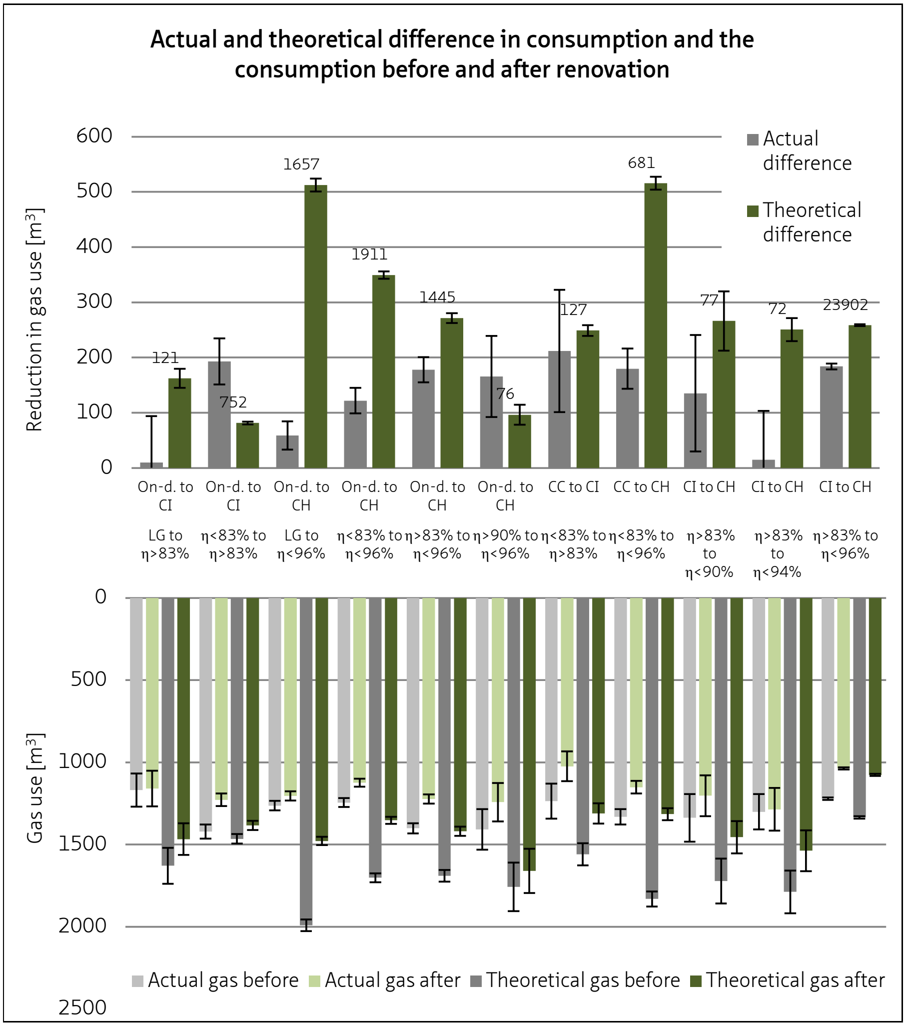


FIGURE 4 Actual and theoretical reduction and number of cases (parenthesis upper figure) and consumption before and after renovation (below) in dwellings with a renovated hot tap water and heating installation system – sample method B (N>30). On-d.= on-demand tankless boiler, CC/CI/CH = combined conventional/improved/high efficiency boiler , LG = local gas heater). Actual reduction of the first and before last column is below the background reduction.

§ 5.5.3 Ventilation

This section shows the actual and theoretical reduction of dwellings which had an improvement in the ventilation installation. We excluded the groups of dwellings which contained less than 30 cases to ensure statistical significance.

A Dwellings with a change in ventilation system

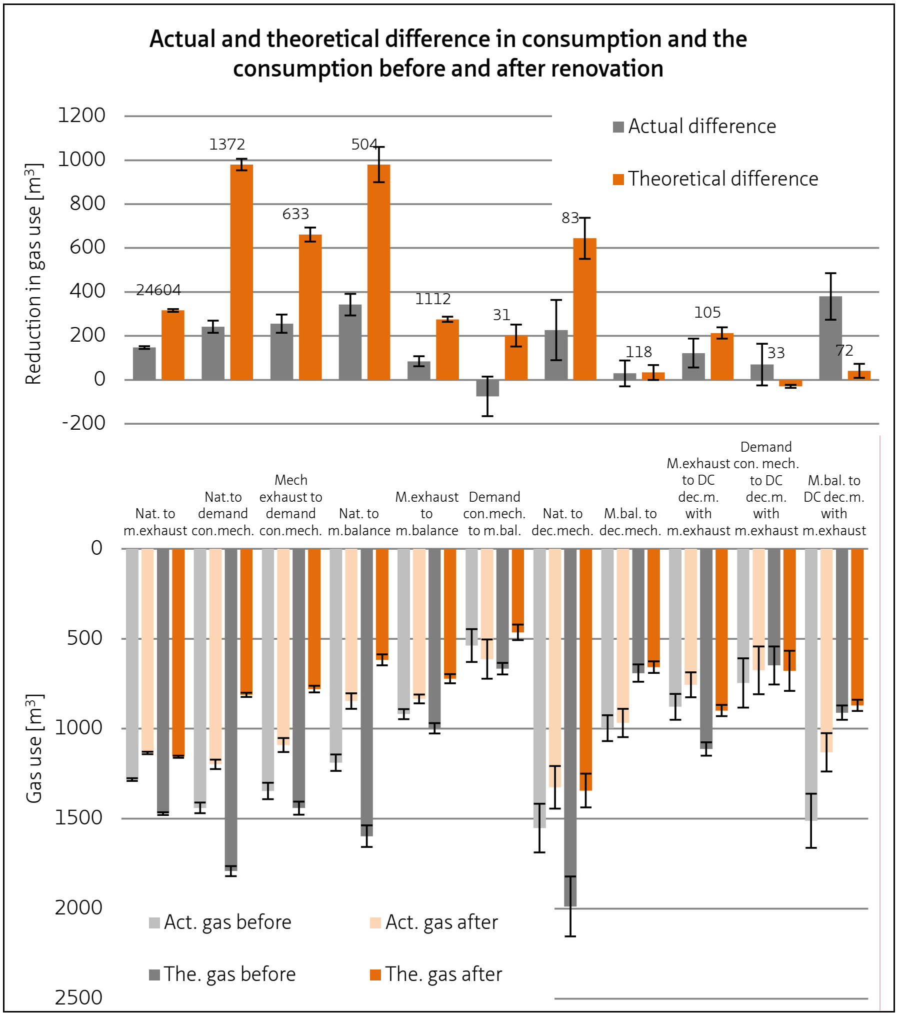


FIGURE 5 Actual and theoretical reduction and number of cases (above) and consumption before and after renovation (below), in dwellings with renovated ventilation system – sample method B (N>30)

The most common change in this category is replacement of natural ventilation with mechanical exhaust ventilation. In this and also most other categories, the decrease in gas consumption is much smaller than expected with the exception of converting a mechanical balanced ventilation system to a demand controlled decentralised mechanical ventilation with mechanical exhaust. Converting a naturally ventilated dwelling into one with mechanical exhaust (the most common renovation) ventilation yielded 147 m^3 of the expected 316 m^3 gas reduction. Other renovation from a natural ventilation system also yielded half to a third of the expected savings.

It is rather interesting, that many categories go from overprediction of gas use (this is typically the case for natural and mechanical exhaust ventilation), to underprediction after the renovation. This creates the large 3-4 fold ratio between theoretical and actual reduction (Figure 5 above). Like Figure 3 and Figure 4, also Figure 5 does not show a correlation between performance gap before and after renovation. Rather this gap seems to correlate well with the type of system (the energy performance of less efficient systems is overpredicted and efficient ones are underpredicted).

B Dwellings with a change in only ventilation system

As opposed to Figure 5, Figure 6 seems to suggest the savings when changing from natural to mechanical exhaust ventilation to be at least three times as high as expected. In Figure 5 we have seen the performance gap in dwellings that changed from natural to mechanical exhaust ventilation system to decrease substantially and the actual gas consumption was overpredicted both before and after renovation. Both these phenomena are not observed in Figure 6. The theoretical gas consumption barely reduces after the renovation, which is logical, since in the calculation method mechanical and natural ventilation both use exactly the same air flow rates. In practice it could be that the savings are achieved at the expense of the air flow rates. Mechanical balance ventilation makes use of heat recovery, which explains the theoretical reduction in the third column, however, the fact that the actual reduction is so much less could mean that heat recovery does not work at the rate assumed by the calculation method. Since in the second column the ventilation is also upgraded to a balance system, it is not clear why the two theoretical consumption are so different, that may relate to project specific data and the small amount of cases. Column three states with statistical significance that actual reduction when replacing mechanical exhaust with balance ventilation is less than a quarter of the expected. Also the last column gives an interesting result, since there is an actual increase in consumption of the systems which are expected to have a reduction. The implemented demand ventilation system does have lower theoretical air flow rate, which explains the theoretical reduction. A validation of air flow rates could solve these problems in the future. A possibility is also that this last category of on-demand decentralised ventilation with exhaust ventilation is not interpreted by the inspectors correctly due to its complexity which could lead to frequent input errors.

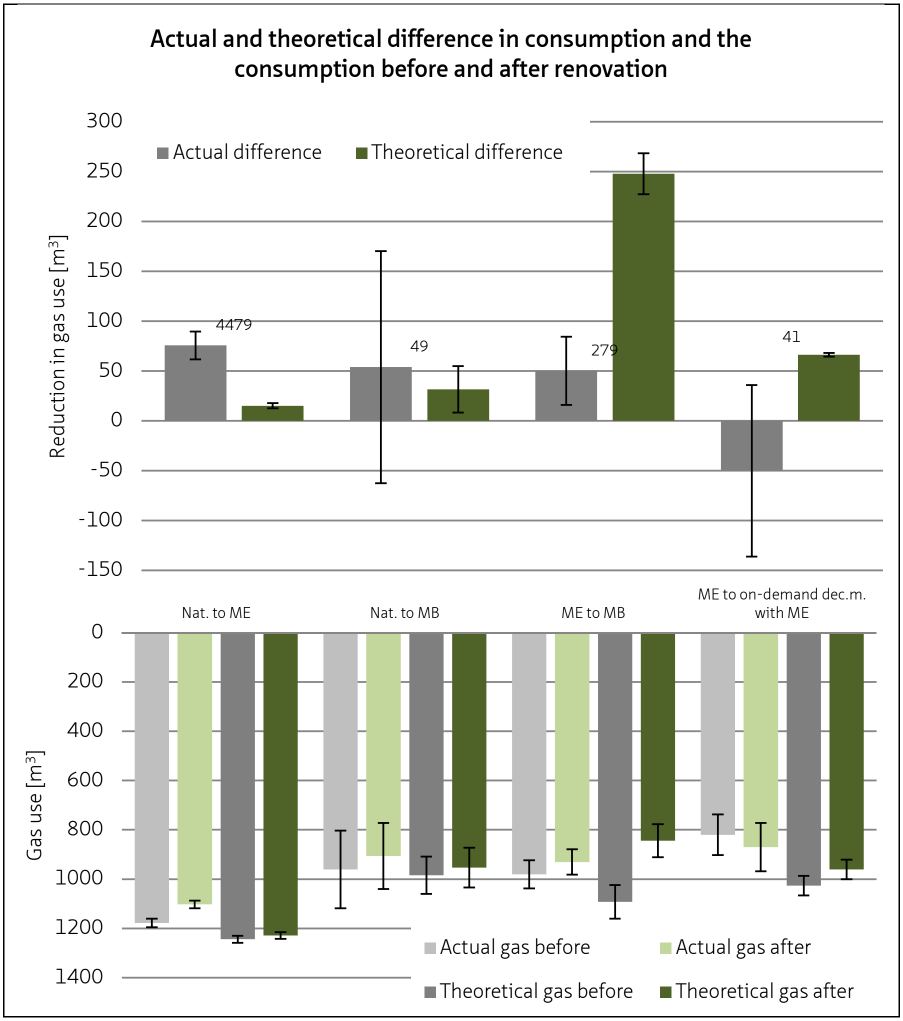


FIGURE 6 Actual and theoretical reduction and number of cases (above) and consumption before and after renovation (below), in dwellings with replaced ventilation system – sample method C (N>30). Nat. to ME = Natural to mechanical exhaust ventilation, Nat. to MB = Natural to mechanical balanced ventilation, ME to MB = Mechanical exhaust ventilation to mechanical balanced ventilation, ME to on-demand dec.m. with ME = Demand-controlled mechanical ventilation with mechanical exhaust

§ 5.5.4 Changes in window quality

This section shows the actual and theoretical gas reduction of dwellings which had an improvement in the window quality. In this section insulation quality as described in Table 1 are used. To keep the results in Figure 7 readable, changes of windows to window insulation category U5 and U4 are not shown. These do follow the same pattern and they have been included in the results of section 5.5.6.

A Dwellings with a change in window quality

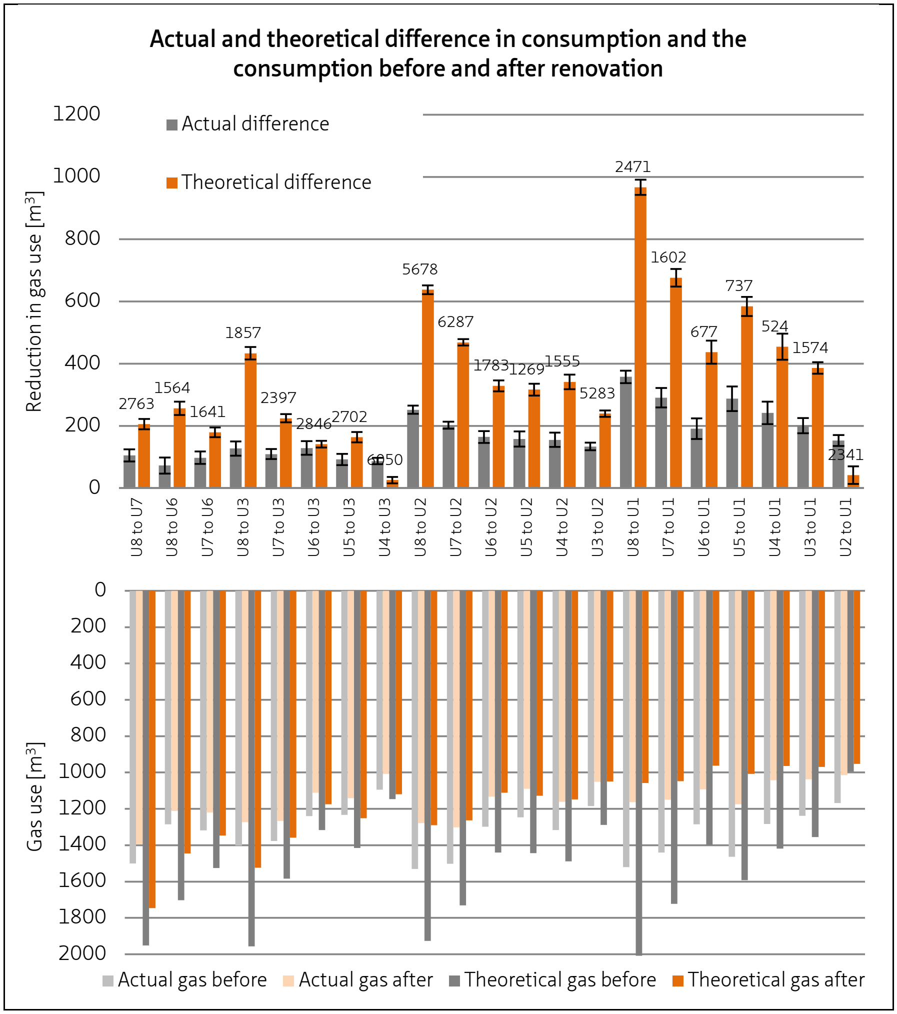


FIGURE 7 Actual and theoretical reduction and number of cases (above graphic) between the first and second pre-label in dwellings with renovated windows (U-value) – sample method B (N>30) Confidence intervals in the bottom graphic are omitted in the bottom graph for better readability

As opposed to previous measures, in this section there is no specific measure that stands out in terms of frequency. This is a feature of window as well as envelope replacements, probably partly because average insulation values were analysed (section 5.4.2). Replacing the glazing never comes close to the expectations, but rather to about half of the predicted saving. Dwellings which were subject to a deeper renovation of windows exhibit a larger reduction in actual gas use (U8 to U1 yielded 357 m^3 reduction out of the expected 966 m^3 reduction). U8 to U7 yielded 105m3 reduction out of the expected 206 m^3 reduction and U2 to U1 290 m^3 out of the theoretical 676 m^3. There are, however, some inconsistencies, such as the group of dwellings which had windows improved from U5 to U1, which saved more than the group with more drastic renovation of U6 to U1. It is possible that a certain group of dwellings contains a large residential dwelling block which had specific renovation characteristics which skews the result of a particular category.

Another thing noticeable from Figure 7 is, that dwellings which had their windows replaced to a more moderate standard (U3-U5) and did not start out with the worse window quality (U8), but rather a U6-U7, exhibit the best match between actual and theoretical reduction. It is nevertheless questionable whether these changes were real renovations or administrative corrections, since such windows are these days not considered standard anymore.

The bottom graphic in Figure 7 shows that the positive performance gap (overprediction) observable before renovation everywhere except in the very last column(category U2 to U1) is just as present after the renovation for all categories except U2 and U1. It seems that dwellings with U value U3 and higher always consume less than predicted whereas others consume more.

B Dwellings with a change in window quality only

Figure 8 reveals that dwellings that had a drastic change in window quality (U8-U2,U7-U1) tend to have an actual gas reduction lower than the theoretical. This phenomenon was seen before in Figure 7 – where just like in Figure 8, the least drastic changes were the best predicted. Some more moderate changes have an actual reduction closer or exceeding the predicted one (U6 to U3, U5 to U2), which is also the case for some small improvements (U2 to U1 or U8 to U7). One also needs to keep in mind that in some cases the actual gas reduction seems to be smaller than the background gas reduction (see section 5.4.4), for example U4 to U1.

Looking at the absolute gas consumption before and after renovation one can see (bottom graph in Figure 8) that the overpredictions observed in bottom chart of Figure 7 in categories U3 and larger is less visible (in some categories they are still notable but in much smaller scale than previously). Also the underpredictions noted previously for U2 and U1 no longer appear consistently. One can therefore hypothesise that the trends seen in Figure 7 were mostly a consequence of a high correlation of window insulation value with other measures taken.

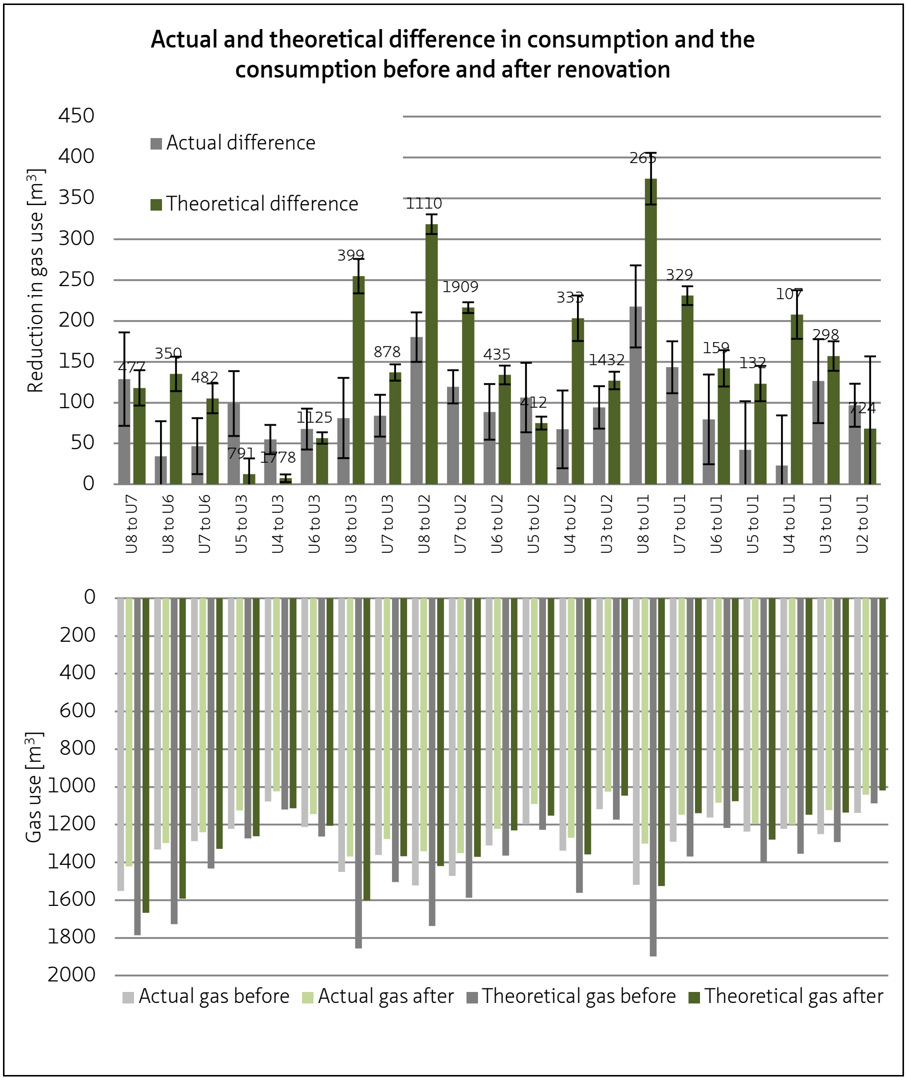


FIGURE 8 Actual and theoretical reduction and number of cases (above) between the first and second pre-label in dwellings with replaced windows (U-value) – sample method C (confidence intervals are omitted in the bottom graph for better readability), N>30. Confidence intervals in the bottom graphic are omitted in the bottom graph for better readability

§ 5.5.5 Changes in envelope quality

This section shows the actual and theoretical reduction of dwellings which had an improvement in the envelope, excluding the groups of dwellings which contained less than 30 cases to ensure statistical significance. The insulation values as described in Table 1 are used. To keep the results in the Figure 9 below readable, we do not show changes of envelope to insulation category R2 and R3. These results follow the same pattern so not much is lost by not conveying those results, which are included in section 5.5.6.

A Dwellings with a change in envelope quality

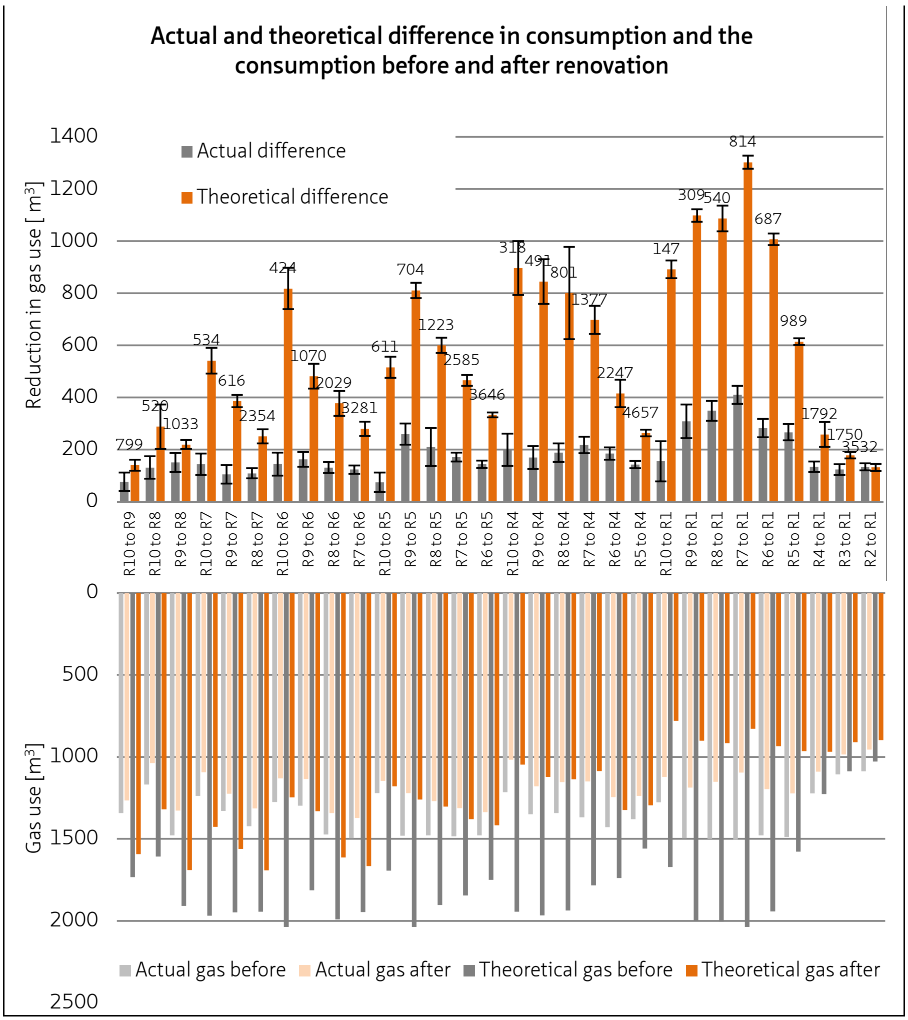


FIGURE 9 Actual and theoretical reduction and number of cases (above) between the first and second pre-label in dwellings with replaced envelope insulation (R-value) – sampling method B, N>30. Confidence intervals are omitted in the bottom graph for better readability

Just as in case of window renovations, there is no measure that stands out in terms of frequency like in the installation measures. The least drastic changes again result in the actual reduction closest to the theoretical, just like in window insulation measure. Even drastic changes yield at most about a third of the expected saving. Roughly, strong overprediction occurs in R5 to R10 and slight underprediction in R1 to R4.

B Dwellings with a change in envelope quality only

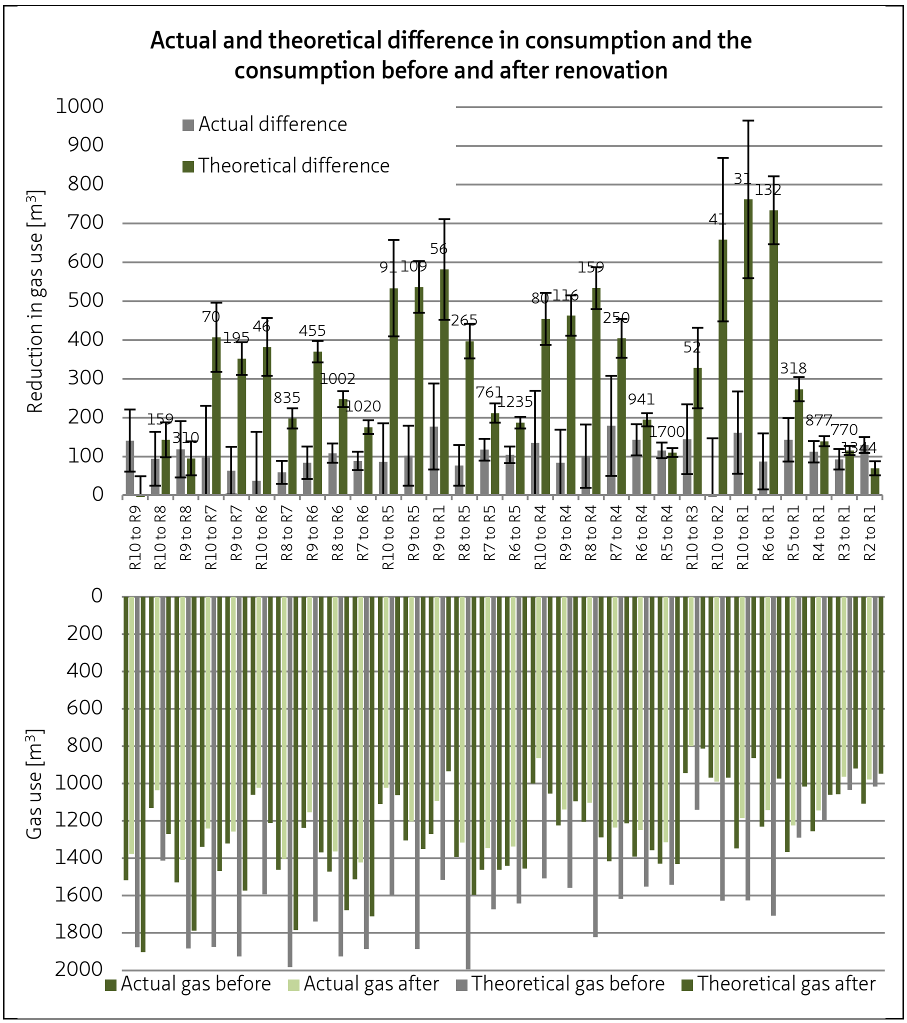


FIGURE 10 Actual and theoretical difference between the first and second pre-label in dwellings with changed envelope insulation (R-value) – sampling method B (N>50). Confidence intervals are omitted in the bottom graphic for better readability

The R value of the envelope is an average value of floor, wall and roof and due to averaging there are fewer dwellings with drastic improvements of the envelope, mostly they only improve by one or two categories. This might seem dissapointing, but in a dwelling with envelope of 300m^2 and an R value of 0,4 insulating the roof (10% of total area) with R=2,5, leads to a new R value of 0,31, which corresponds to a change for one category only (R5 to R6).

The results are similar to those for improving U value of the envelope – small changes are well predicted and actual reduction is close or surpassing the theoretical whereas deeper changes result in actual reduction being much lower than predicted. The better the dwelling is insulated, the easier it is to achieve the envisioned saving, as in general, the gap between predicted and actual consumption is larger in insulations R5 and higher (bottom graphic of Figure 10).

§ 5.5.6 Actual consumption savings among different measures

One of the objectives of the paper was to see which measures are most effective in achieving energy savings. Several tables in this section demonstrate average reduction rates for separate measures. First of all, averages of various measures are calculated in

Table 7 and Table 8 taking into account all the groups containing more than 30 records, the first summing up the results of sample B and the later of sample C. Sample B studies a larger sample, therefore the totals and numbers of dwellings within a measure are, logically, higher. Interestingly, the measure which achieves the largest actual cumulative as well as individual saving in sample B is window replacement and in sample C it is the replacement of heat and hot tap water system. In both samples envelope improvement is in the second place and ventilation improvement the last. Looking at savings in the two tables, both actual as well as theoretical consumption reductions are higher in sample B than in C, which makes sense, since there is a large chance that dwellings in sample B had another renovation measure taken. Comparing the numbers of dwellings in each measure group (last column) reveals that the group of heating and hot tap water has the highest similarity in both samples, since sample B had 60.960 dwellings in this group and sample C 30.749, which is more than half. This means that more than half of the dwellings with a change in heating and hot tap water had no other dwelling change, whereas the other smaller half, did. About two thirds of dwellings with envelope improvement also had other measures taken (21.035 in sample C vs. 62.955 in sample B) and about three quarters of dwellings with window improvement also had other measures taken (15.744 in sample C vs. 61.233 in sample B). The measure which was most usually combined with others was ventilation improvement, which also explains the drastic difference in reduction of this group in sample B and C. When comparing the ratios of actual vs. theoretical reduction, one first notices a higher average ratio in sample C than that of sample B. This means that dwellings with a single renovation measure have on average a better predicted reduction than those with combined measures. The most remarkable considering individual measures, is the gas reduction in dwellings with an improved ventilation systems, achieving a 2,5 times higher reduction than predicted.

RENOVATION MEASURE	CUMULATIVE SAVING (TOTAL SAMPLE)		INDIVIDUAL SAVING (PER DWELLING)			N
	TOTAL ACTUAL GAS REDUCTION	TOTAL THEORETICAL GAS REDUCTION	AVERAGE ACTUAL GAS REDUCTION [M³]	AVERAGE THEORETICAL GAS REDUCTION [M³]	RATIO ACTUAL /THEORIETICAL GAS REDUCTION	
Ventilation	11%	11%	148	327	0.45	26,325
Windows	33%	30%	203	363	0.56	61,233
Envelope	25%	29%	147	352	0.42	62,955
Heating and hot tap water	31%	30%	190	365	0.52	60,960
Total [m³]	37,177,026	75,269,315				211,473
Average			176	356	0.49	

TABLE 7 Totals and averages of actual and theoretical gas reduction for different renovation measure using sample B - non-exclusive measure (groups with N>30)

RENOVATION MEASURE	CUMULATIVE SAVING (TOTAL SAMPLE)		INDIVIDUAL SAVING (PER DWELLING)			N
	TOTAL ACTUAL GAS REDUCTION	TOTAL THEORETICAL GAS REDUCTION	AVERAGE ACTUAL GAS REDUCTION [M³]	AVERAGE THEORETICAL GAS REDUCTION [M³]	RATIO ACTUAL /THEORIETICAL GAS REDUCTION	
Ventilation	4%	1%	73	29	2.52	4,848
Windows	16%	14%	96	134	0.72	15,744
Envelope	23%	25%	104	180	0.58	21,035
Heating and hot tap water	57%	56%	172	279	0.62	30,749
Total [m³]	9,367,264	14,622,945				72,376
Average			131	188	0.70	

TABLE 8 Totals and averages of actual and theoretical gas reduction for different renovation measure using sample C – unique measure (groups with N>30)

	ACTUAL REDUCTION [M³]	N	RATIO
U8 to U1	218	265	0,6
η<83% to η>83% and CC to CI	212	127	0,9
η<83% to η>83% and On-d. to CI	193	752	2,4
η>83% to η<96% and CI to CH	184	23,902	0,7
U8 to U2	180	1,110	0,6
η<83% to η<96% and CC to CH	180	681	0,3
η>83% to η<96% and On-d. to CH	178	1,445	0,7
η>90% to η<96% and On-d. to CH	166	76	1,7
U7 to U1	143	329	0,6
R5 to R1	143	318	0,5
η>83% to η <90% and CI to CH	135	77	0,5
U8 to U5	133	253	0,5
R2 to R1	130	1,344	1,9
U8 to U7	129	477	1,1
R8 to R3	128	90	0,2
U3 to U1	126	298	0,8
η<83% to η<96% and On-d. to CH	122	1,911	0,3
R4 to R1	113	877	0,8
R8 to R6	109	1,002	0,4
R8 to R4	101	159	0,2
U8 to U4	99	111	0,4
U2 to U1	97	724	1,4
R3 to R1	93	770	0,8
R6 to R1	87	132	0,1
U8 to U3	81	399	0,3
U6 to U1	80	159	0,6
R8 to R5	77	265	0,2
Natural to mechanical exhaust	76	4,479	5,0
LG to η<96% and On-d. to CH	59	1,657	0,1
R8 to R7	59	835	0,3
Natural to mechanical balance	54	49	1,7
Mechanical exhaust to mechanical balance	50	279	0,2
U5 to U1	42	132	0,3
U8 to U6	34	350	0,3
U4 to U1	23	107	0,1
η>83% to η <94% and CI to CH	15	72	0,1
LG to η>83% and On-d. to CI	10	121	0,1
Mechanical exhaust to on-demand decentralised mechanical with mechanical exhaust	-50	41	-0,8

TABLE 9 Actual consumption reduction per dwelling of various single renovation measures

Table 9 shows the actual gas reduction, the number of dwellings and the ratio between actual and theoretical consumption reduction. The highest reduction is achieved by drastically improving the U value of the windows (U8 to U1). The actual reduction of such a change (Table 9 first row left) is below the theoretical and the number of dwellings in this category is rather low. The category containing the most dwellings, is the one where heating systems were replaced from a η>83% to η<96% and hot tap water installation renovated from improved to high efficiency. The actual reduction of this group is also below the expected. The measures achieving the most reduction are therefore drastic improvements of window quality and an improvement of the efficiency of heating and hot tap water system (not a replacement of a local system).

Measures that achieve an actual reduction higher that the theoretical seem to mostly be less drastic changes, such as insulation improvement from R2 to R1 or window improvement from U8 to U7 or U2 to U1. Also notable is the underprediction of the reduction in dwellings where natural ventilation was replaced by mechanical exhaust and it is questionable whether such dwellings still have a sufficient quality of indoor air after the renovation. The two heating installation improvements that yielded a reduction higher than theoretical (third and eight row of Table 9) are both within a certain boiler type (in first case non-condensing and in the second, condensing), improvements in between these categories have an actual consumption lower than the theoretical one. This probably means that some of the calculation factors used for efficiencies of gas boilers do not reflect the real efficiency correctly.

MEASURES RESULTING IN HIGHEST CUMULATIVE SAVING	ACTUAL GAS REDUCTION *N [M³]	N	% OF TOTAL REDUCTION IN STUDIED SAMPLE
Heating boiler η<83% to η>96% hot water from improved to high-efficiency boiler	4,396,716	23,902	38%
Natural to mechanical exhaust ventilation	340,404	4,479	3%
Heating boiler η<83% to η<96% hot water from on demand to high-efficiency boiler	257,204	1,445	2%
Heating boiler η>83% to η<96% hot water from on demand to high-efficiency boiler	233,094	1,911	2%
U8 to U2	199,800	1,110	2%
R2 to R1	174,720	1,344	2%
Heating boiler η<83% to η>83% hot water from on demand to improved efficiency boiler	145,277	752	1%
Heating boiler η<83% to η>96% hot water from conv. to high-efficiency boiler	122,457	681	1%
R8 to R6	109,218	1,002	1%
R4 to R1	99,100	877	1%

TABLE 10 Cumulative actual gas consumption reduction of the studied sample

Results in Table 9 are informative in terms of the efficiency of individual measures, however, the problem is that many of these results have poor statistical significance due to the low sample size (the confidence bands can be seen in previous sections). To emphasise the measures which yield the most savings in the studied sample, Table 10 sorts the measures according to the cumulative saving – the sum of the savings of all dwellings in a particular category. This is of course strongly dependent on the sample, but if we consider the studied sample representative it is impressive how much actual gas reduction (38%) comes from replacing the heating and hot tap water system and that 3% of savings come from upgrading the natural ventilation system. Probably the popular measures are the most cost-effective ones.

§ 5.6 Discussion

The results section showed results using three sampling methods. Cross sectional method (A) was only used for dwellings total thermal performance (energy label) and comparison with method B yielded similar results in terms of performance gap (see ratio column in Table 6) unless looking at small changes (mostly one label step) of very poor or very well performing dwellings (e.g. G to F or B to A). Summarizing, longitudinal data is essential when examining the effect of single renovation measures. Albeit carefully, cross sectional data can be used for estimating deep improvements in overall performance (roughly, more than one label class).

The reason could be that in those extreme labels (G or A), cross sectional method compares entities that are not comparable – for example, dwellings in A label are significantly larger than B dwellings (Majcen et al. 2013a), or they could be characterised by a much larger number of occupants. Longitudinal methods do not ensure that analysed dwellings have not undergone a change in household – the chance is, however, much smaller than in cross sectional data, where we know households to be different in each dwelling group. However, even though the ratio of the performance gap across label classes is roughly similar, the actual gas consumption reduction is consistently larger using longitudinal data than cross-sectional data. This highlight the importance of longitudinal data collection for better estimation of actual gas reduction.

If the theoretical consumption before and after renovation would be comparable using method B and C, it would mean that sample B represents well the theoretical consumption of the observed measure. This is, however, almost never the case, since sample B includes a number of cases where also other measure have occurred. Comparing method B and C for renovation measures in fact yielded roughly comparable results when it comes to dwelling insulation (window and envelope) and very different results when looking at installation systems. It seems that better performing systems in general exhibit a smaller performance gap, such as boilers with a higher efficiency, mechanical ventilation and better insulation. Two very notable performance gaps were the one in local gas heater and on-demand tankless water boilers and naturally ventilated buildings. The most extreme example are dwellings with a changed ventilation system where the performance gap ratio in method C is 4 times the ratio of method B. This proves that when analysing single measures, one should definitely ensure other properties are constant making the results of method C are therefore a better basis for conclusions regarding performance gap and actual reduction of the measures. The problem of this method is, however, that we (currently) cannot find enough data to provide significant results for many of the possible combinations of measures, which should be improved in the future with expansion of SHAERE.

The average actual gas reduction in sample B is 176 m^3, which represents 15,5% of the total consumption (see Figure 2) and corresponds to one or several implemented measures. The theoretical reduction of this same sample, 356m^3 makes 27,4% of the theoretical total consumption (Figure 2). For single measures (sample C) the actual and theoretical gas reductions are 131 and 188m^3 which makes up for a reduction of 11,6% and 16,9%. Hezemans et al. from 2012, who assumed that two measures coincide with a 20% reduction, was therefore quite close to reality, although the actual average value is somewhere between 11,6 and 15,5%.

There are some uncertainties regarding the results. According to Aedes, pre-labels are updated whenever a renovation measure takes place and are considered accurate, however, the fact that a number or deteriorations were identified within SHAERE demonstrates that this is not entirely true. This could probably improve in the future as the database grows, however, it is a major uncertainty in this study. This study was done purely on social housing sector and moreover excluded certain heating types (heat pumps), which has consequences for representativeness of the results. Another situation in which a dwelling was not considered in this paper is if during the renovation, its address changes, which is the case in a number of deep renovations. At the time of the study, it was not possible to find out the extent to which this occurs. Moreover, certain parameters such insulation of wall, floor and roof have been aggregated in this paper and would be interesting to analyse independently using continuous instead of categorical values. In section C we analysed the change in one of the dwelling properties, however, we neglected the impact of others (even though constant). For example, it might be significantly different whether the dwellings which had a renovated installation system was very well or poorly insulated. In the future, other statistical methods (correlation tests, regression analysis) should be tested on similar large data, since this allows to include more variables and also enables the use of control variables. In the upcoming studies, one could also limit oneself to deeper performance changes. Here we observed all changes (also small ones, within one label category), however, the results might be more robust selecting a subsample where one or even two label steps have been taken – especially in line with the uncertainties regarding administrative corrections in the data.

§ 5.7 Conclusions

To conclude, several main findings are summarized below.

- In terms of gas reduction by single measures, improvements in efficiency of gas boilers (space heating and hot tap water) yield the biggest energy reduction, followed by deep improvements of window quality. Improving the ventilation system yields a relatively small reduction compared to other measures, however, it is still much larger than theoretically expected.

- In terms of the performance gap between actual and theoretical consumption, high R and low U values of insulation are well predicted, as well as efficient heating systems. On the other hand low R and high U values, local heating systems, changes from a non-condensing into a condensing boiler and upgrades to a natural ventilation system are not well predicted. In Majcen et al., 2013b, it was shown that departures from the standard average dwelling temperature were causing a part of the performance gap and in the present paper it is shown that efficiencies of heating systems and insulation values are also causing a part of the gap.

- This poses the question of how well the standard values are really defined in the calculation method. It could be that excessively low efficiencies have been attributed to inefficient systems simply because of misconception and lack of knowledge, as from an economical point of view, it is more logical to invest effort into estimating the performances new systems. However, not knowing the real efficiencies of older systems results in a performance gap.

- However, since actual consumption data on the level of individual dwelling is becoming available these inconsistencies become visible. The standard values should either be revised or alternatively, one should utilise the available actual gas consumption values in order to make better estimates (Majcen et al., 2015).

- Large datasets such as the SHAERE investigated in this paper are now arising across Europe and few experience is available about how to handle them. The results of large samples are statistically robust and representative, however selecting subsamples from the data offers insight into specific combinations of measures and allows identification of best practices. Energy performance registers should be made publicly available, possibly already coupled with actual consumption data.

- It is of utmost importance to ensure that building performance databases are of sufficient quality and have trustworthy input data. Ensuring such level of quality is not simple, even if dwellings are used for asset management by large housing companies (associations). This paper has highlighted the importance of analysing dwelling stock registers for both the validation and evaluation of energy label calculation. However, in The Netherlands, a simplified label came into force in 2015 next to the existing, complete label. This changed a lot in this field, since the simplified label requires no inspection at all and can be filled in online by the owner of the house himself. The implications of this simplified label are not yet clear, just as it is not clear yet, whether housing associations will continue to inspect a dwelling and get a complete energy label or not.

- Further study should also include costs of the different renovation measure. The results of this paper showed that windows and installation system upgrades provide a high actual reduction, and the remaining question is which of the two is more viable economically. This question is relevant also in the framework of cost effectiveness of nZEBS according to EPBD.

Overall, this paper has shown once more that the calculation method currently in use cannot be considered accurate if compared to actual consumptions. The question that remains is how to, under these circumstances, determine the effectiveness of a specific renovation measure, which is of importance on dwelling level and even more so on the level of the whole stock. If theoretical methodology is to be used as baseline without the use of actual consumption at some point in the process, realistic standard values have to be prescribed.

§ 5.8 References

Adalberth, K., 1997. Energy use during the life cycle of single-unit dwellings: Examples, Building and Environment, Volume 32, Issue 4, Pages 321-329.

Delghust, M., Roelens, W., Tanghe, T., Weerdt, Y.D., Janssens, A. 2015. Regulatory energy calculations versus real energy use in high-performance houses, Building Research & Information. Pages 1-16.

Dodoo, A., Gustavsson, L., Sathre, R. 2010. Life cycle primary energy implication of retrofitting a wood-framed apartment building to passive house standard, Resources, Conservation and Recycling, Volume 54, Issue 12, Pages 1152-1160

Economidou, M., Atanasiu, B., Despret, D., Ingeborg, J.M., Rapf, N.O. 2011. Europe's buildings under the microscope, Country-by-country Review of the Energy Performance of Europe's Buildings, BPIE, 2011.

Filippidou, F., Nieboer, N., Visscher, H., 2015a. Energy efficiency measures implemented in Dutch non-profit housing sector, ECEEE 2015 Summer Study proceedings, Hyeres, France.
Filippidou, F., Nieboer, N., Visscher, H., 2015b. The energy renovation pace of the Dutch non-profit housing sector, Submitted to Energy Policy in July 2015
Hezemans A., Marquart E., Monné T., Monitor Energiebsparing Gebouwde Omgeving 2012, Agentschap NL, Juni 2012.
ISSO 82.3 Publication Energy Performance Certificate—Formula Structure (Publicatie
82.3 Handleiding EPA-W (Formulestructuur'), Senternovem, October 2009.
Karlsson, J.F., Moshfegh, B., 2007. A comprehensive investigation of a low-energy building in Sweden, Renewable Energy, Volume 32, Issue 11, Pages 1830-1841
Majcen, D., Itard, L., Visscher, H., 2013a. Actual and theoretical gas consumption in Dutch dwellings: What causes the differences? Energy Policy 61, 460–471.
Majcen, D., Itard, L., Visscher, H., 2013b. Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications, Energy Policy 54, 125–136.
Majcen, D., Itard, L., Visscher, H., 2015. Statistical model of the heating prediction gap in Dutch dwellings: Relative importance of building, household and behavioural characteristics, submitted to Energy and Buildings in June 2015
Menkveld, M., Leidelmeijer, K., Vethman, P., Cozijnsen., E. 2012. Besparingsgetallen energibesparende maatregelen, ECN, May 2012
Raynaud, M. 2014. Evaluation ex-post de l'efficacité de solutions de rénovation énergétique en résidentiel, Doctoral thesis, MINES ParisTech Centre Efficacité énergétique des Systèmes.
Thormark, C. 2002. A low energy building in a life cycle—its embodied energy, energy need for operation and recycling potential, Building and Environment, Volume 37, Issue 4, Pages 429-435
Tigchelaar, C., Leidelmeijer, K. 2013. Energiebesparing: Een sampenspel van woning en bewoner – Analyse van de module Energie WoON 2012, ECN, August 2013
Visscher, H., Majcen, D., Itard, L. 2013. Gebruik van de SHAERE-database voor het monitoren van het Convenant Energiebesparing Huursector, Technische Universiteit Delft, Faculteit Bouwkunde, OTB - Onderzoek voor de Gebouwde Omgeving.
Winther, B.N., Hestnes, A.G., 1999. Solar Versus Green: The Analysis of a Norwegian Row House, Solar Energy, Volume 66, Issue 6, Pages 387-393.