Benchmarking the measured energy use of Nordic residential buildings and their Zero Energy-readiness

Andrea Ferrantelli[1,2], *Martin* Thalfeldt[1,2], *Jarek* Kurnitski[1,2,3]

[1] FinEst Centre for Smart Cities (Finest Centre), Tallinn University of Technology, 19086 Tallinn, Estonia
[2] nZEB Research Group, Tallinn University of Technology, 19086 Tallinn, Estonia
[3] Department of Civil Engineering, Aalto University, 00076 Aalto, Finland

Abstract. It is well known that buildings are responsible for a nearly 40% share of the total energy consumption; in order to reduce it by improving the energy efficiency of the building stock, it is necessary to first evaluate their performance. Building energy benchmarking provides information to stakeholders and motivates energy retrofits, by evaluating and comparing a building to similar units and/or to a reference building in terms of energy consumption with the minimum amount of data possible.
Towards this end, in this paper we analysed nearly 19000 Estonian Energy Performance Certificates (EPCs) of detached houses. By means of a systematic statistical investigation, we determined the time evolution of EPC labels and evaluated the impact of incentives pre/post renovations, drawing a comprehensive and updated picture of the Estonian detached houses. This allowed evaluating their readiness based on recent trends: unfortunately, new or renovated dwellings are not estimated to achieve the zero-energy status by 2050. Although marginally due also to the use of homeworking during the COVID-19 pandemic, we show that this is mostly determined by changes in the regulations. A benchmarking ranking for each construction type was also created by calculating rating tables based on a 0-100 coefficients scale; this allows comparing with the existing stock any building with known EPC, for energy Audit and other investigations aiming at energy efficiency.

Keywords. Benchmarking, energy consumption, energy management, energy efficiency, statistical analysis
DOI: https://doi.org/10.34641/clima.2022.171

1. Introduction

Building energy audits [1] are an important step in reducing carbon dioxide emissions and the energy consumption of the building stock. These can consist of inspections and surveys aimed at understanding the energy use of the building, starting from e.g. a review of utility bills, to identify opportunities for improving energy efficiency through operational adjustments or system upgrades. Reviewing bills and conducting surveys is called "Simple level Audit" or "Audit Level 1" by ASHRAE, that defines a scale of three stages [2].

As Level 1 can only uncover major problems in the system and does not provide enough data granularity for effective diagnostics and sophisticated statistical analysis, Level 2 and 3 audits are required for a more comprehensive understanding of the building's energy use. These are usually based on data that covers the HVAC system, building envelope etc. and is collected directly through (wireless) sensors. This allows real-time monitoring of the energy usage for efficient interventions on the mechanical and electrical system, as well as testing advanced energy efficiency measures together with their cost-effectiveness [3]. For realising detailed energy audits, it is therefore necessary to acquire a deep and wide knowledge of the energy consumption status of the building stock, to identify eventually problematic sectors or clusters, and to single out representative buildings, or building typologies, for operation monitoring.

Energy Performance Certificate (EPC) databases have a variety of applications [4], as they are directly related to energy consumption. The EPCs in fact are usually defined as the measured or calculated energy consumption of a building during a year. In the European Union (EU), it is common to rate EPCs from A to F (from best to worst) for market classifications, or for building benchmarking. In general, a benchmark is a concept that originated from manufacturing, and it is used to measure the performance of a process. In energy investigations concerning buildings, a building consumption can be used as the indicator to which the benchmark is compared [5].

A variety of methods have accordingly been formulated in the literature, in order to define accurately a benchmarking procedure that could be extended to diverse construction typologies, for evaluating the present status of the building stock and finding the occurrence of problems and anomalies. Roth [6] propose a benchmarking method based on normalised consumption, by using a non-linear statistical analysis of open-data from 10 cities. Park [7] analysed 1072 office buildings focusing on the US counterpart of the EPC certificate, the Source Energy Use Intensity (Source EUI).

This paper follows those earlier studies and analyses into detail the Detached houses cluster of a larger ~35000 buildings EPC database that was investigated in [8] but did not account for this specific cluster into detail. Here we fill that gap and provide the missing information with a complementary paper that completes the study of the full dataset. Aided by the software R [9], we performed a thorough statistical study of the EPC certificates of over 18000 Estonian dwellings, which were divided into three subcategories according to heated area, following the national legislation's regulations.

2. Research methods

2.1 EPC database and energy labels

Here we address a database of 18689 EPCs of Estonian detached houses, with data categories: building id, construction year, renovation year, heated area, ETA, KEK (an EPC certificate in Estonia can either be ETA if calculated, typically with simulation software, or KEK if measured). The buildings are detached and terraced houses that were further subclassified as portions with dedicated entrance, two or three apartment houses and so on. The EPCs were released between the late 1990's and February 2022 and comprised a few negative values (houses with energy generation e.g. solar panels), as well as a number of outliers with exceedingly large consumption.

As the highest EPC value allowed by the Estonian energy labels classification is 1350 kWh/(m^{2}a), we set 1500 kWh/(m^{2}a) as an upper cut-off for the database, resulting in N=18122 EPCs, with Median equal to 138.0 kWh/(m^{2}a), mean M=143.2 kWh/(m^{2}a) and standard deviation SD=51.04 kWh/(m^{2}a). An important characteristic of the Estonian legislation is that the Detached houses EPC data are subdivided into three groups according to the heated area A, corresponding to A<120 m^2, A=120 m^2-220m^2 and A>220 m^2. We thus labelled the three subclusters with D1, D2 and D3; the boxplot in **Figure 1** and **Tab. 1** shows the breakdown of means and data spread.

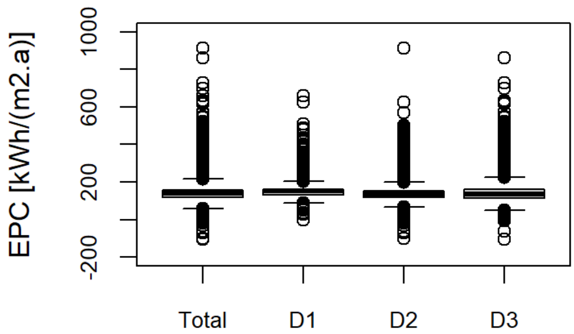


Figure 1 - Boxplot of the EPCs for the database

Tab. 1 – Detatched houses energy classification: number (N), mean (M) and Standard Deviation (SD).

ID	A [m^2]	N	% tot	M	SD
D1	<120	2265	12%	154.8	49.15
D2	120–220	10089	56%	138.9	42.96
D3	>220	5768	32%	146.2	62.68

Such a clustering is reflected in the energy labelling of the national regulation, which runs from A to H as per European Union (EU) standards with the thresholds displayed in **Tab. 2** below.

Tab. 2 – Estonian energy labels for the three categories of detached houses D1, D2 and D3; EPC [kWh/(m^{2}a)].

En. label	D1 (EPC)	D2 (EPC)	D3 (EPC)
A	≤ 145	≤ 120	≤ 100
B	146-165	121-140	101-120
C	166-185	141-160	121-140
D	186-235	161-210	141-200
E	236-285	211-260	201-250
F	286-350	261-330	251-320
G	351-420	331-400	321-390
H	≥ 421	≥ 401	≥ 391

The nZEB level corresponding to class A was first defined in 2013 and then revised in 2018, due to updated cost-optimality calculations and non-renewable primary energy factors for the class A EPC values. The dwellings initially had an EPC value of 50 kWh/(m^{2}a) regardless of the heated floor area A, but the revision increased the corresponding values substantially, also introducing a floor area dependence as well. Furthermore, after 2018 class A must be reached only by the D3 cluster, i.e. dwellings with heated floor area > 220 m^2. D1 and D2 must only meet class B requirements. As we shall explain in Section 3, this turned out to be crucial for the latest years' trends and for the energy readiness of all three building groups.

Such energy "readiness" is the capability of a specific building cluster to reach the ZEB status and zero emissions by 2050 [9]. By fitting the EPC certificates for recent buildings, namely those that were built or renovated after 2000, we looked at the intersect of the fitting curve with the EPC=0 axis. This allowed for a rough estimation of the ZEB year for the three building subclusters D1, D2 and D3, as well as for the full dataset. The linear fit was computed with a simple linear model lm() in R.

As the fit including the years 2000 through 2022 included also the COVID-19 pandemic period, we wondered whether it could be reflected in the ZEB year estimates. A fit of the data from 2000 until the end of 2019 was accordingly performed and results were compared.

In the following, since the national regulation does not distinguish between ETA and KEK, accepting either certificate as an EPC with units kWh/(m^{2}a), we will mostly mention EPC values, referring to ETA or KEK only when this distinction is significant.

2.2 Fitting distributions and benchmarking

Examining the full dataset histogram in **Figure 3** unveils a clear structure, namely three peaks at 110, 130 and 150 kWh/(m^{2}a), and a very long yet unsubstantial right tail. It can be shown that the D1, D2 and D3 EPC datasets reflect this structure, with the exact same three peaks for D2 and only those at 110 and 150 kWh/(m^{2}a) for D1 and D3.

As our main purpose is benchmarking, namely creating a reference profile for each dataset against which one compares the energy efficiency of a corresponding given building, we need a reliable method to determine where the EPC value of this building sits within this general distribution. In this paper and in [8] we followed Ref. [9] by fitting each of the D1, D2 and D3 datasets of ETA/KEK certificates with a probability distribution, which was then integrated to return the empirical cumulative distribution function (CDF).

This allowed mapping a unique EPC at any given ratio of the dataset; the ranges of EPC values correspond to different quantiles of each original EPC distribution. We also added a 10–100 points scale to embed a rating system that quantified the energy efficiency. This resulted in so-called "benchmarking tables" that are reported in **Tab. 5**, **Tab. 6** and **Tab. 7**. Fitting was non-trivial due to the sharply multimodal nature of the data, to which the ordinary normal or gamma distributions did not apply. The tri- or multimodal distributions were thus handled by means of a Gaussian finite mixture with the R package "mclust" [10], that overlapped either two (D1 and D3) or three (D2) normal distributions and integrated accordingly.

3. Results

3.1 Time trend of EPC labels

The yearly breakdown of EPCs that have been issued since 2010 is illustrated in **Figure 2**. It is clearly seen that after the 2013 remodulation there has been a quick increase in certificates, reaching a steady trend after 2015. A histogram of the full database of 18122 Dwellings is also given in **Figure 3**; **Figure 4** features a plot showing EPC values in function of construction year.

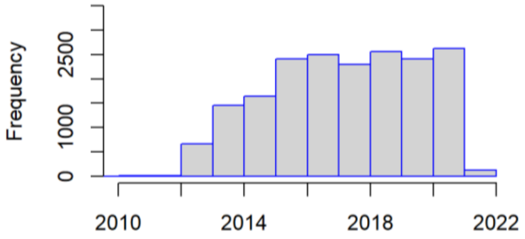


Figure 2 - Yearly breakdown of EPC certificates that were issued since 2010

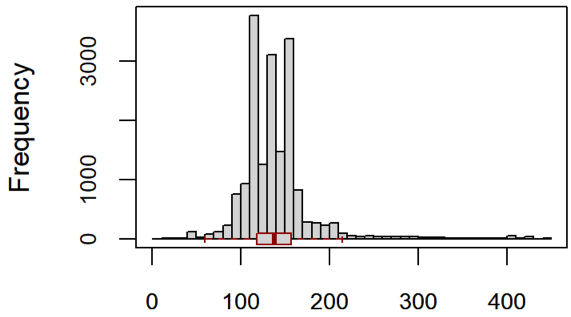


Figure 3 - Histogram of EPC values for the full dataset

A slight accumulation of values below ~300 kWh/(m^{2}a) can be observed in the most recent years, probably determined by the more stringent legislation requirements. Otherwise no clear pattern can be seen, with lower EPCs remaining consistently above 100 kWh/(m^{2}a), i.e. class A. As discussed in Section 2, the EPC certificates for Estonian buildings are comparable until a critical 2013 remodulation, then in 2018 the energy label requirements became more strict; after 01.01.2019 it was required to every newly constructed or renovated building to comply with even stricter bounds.

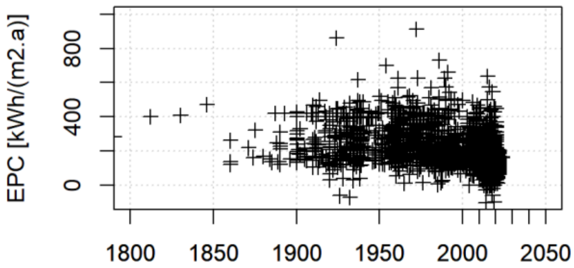


Figure 4 - EPC values for the full dataset in function of construction or renovation year

In this section we accordingly split the data into three clusters according to the EPC certificate release date: 2013-2017, 2018, 2019-2022. The bar plots in **Figure 5** follow this reasoning and illustrate how for the houses of any size, the most prominent class shifted from C in the period 2013-2018 to B since 2019. A dramatic (yet very welcome) increase in Class A certificates, previously critically scarce, can be observed as well after January 2019.

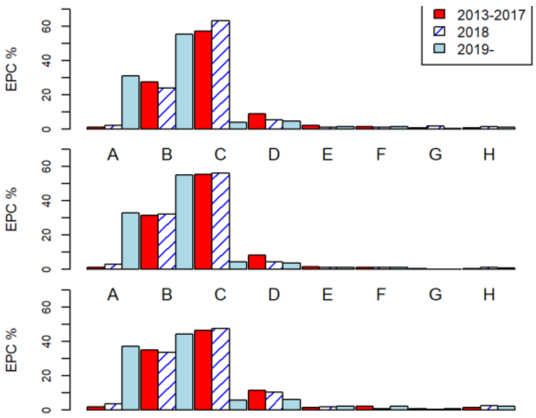
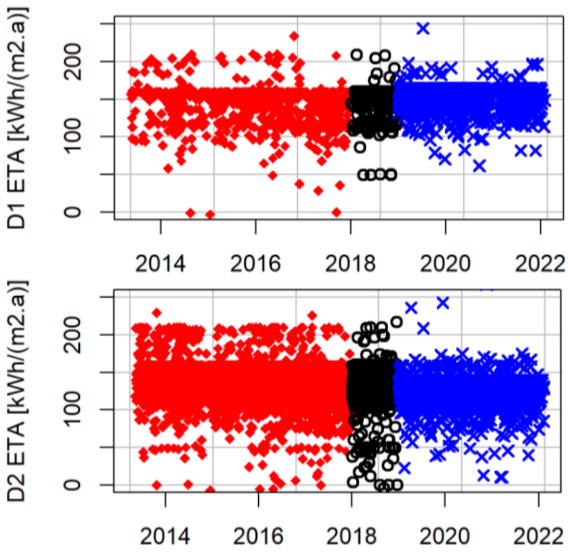


Figure 5 - ETA/KEK certificate classes (percentage over the total) for D1 (top), D2 (middle) and D3 (bottom), grouped by year of certificate release

Another important feature of **Figure 5** is a strong clustering of EPCs towards A, B and C classes. Especially for the D2 dataset, the less energy efficient buildings beyond D Class are negligible.

3.2 Impact of incentives pre/post renovations

How do stricter energy label requirements influence the EPC values? According to **Figure 6**, D1 mostly lies below 160 kWh/(m^{2}a), i.e. within Classes A and B; for D2 the EPCs tend to cluster below 150 kWh/(m^{2}a) after 2019, namely they stay within Class B. D3 exhibits a plateau at about 160 kWh/(m^{2}a), i.e. Class D, consistently with **Tab. 1**.



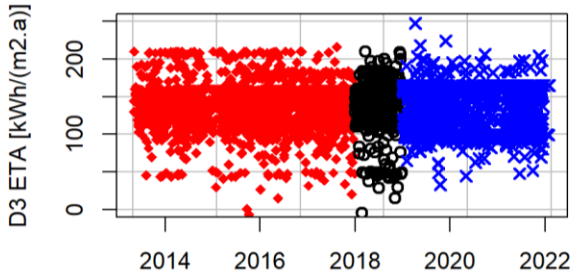


Figure 6 - ETA certificates for D1 (top), D2 (middle), D3 (bottom) versus date of issue: 2013–2017 (red diamonds), 2018 (black circles), 2019–2022 (blue crosses).

Figure 7 features three graphs, one for each Di cluster, where ETA (computed) and KEK (measured) are plotted against construction or renovation year. In all of the cases the ETAs are more frequent after 2013, i.e. when the major remodulation of energy consumption calculations occurred, in agreement with the previous Sections' findings. It is confirmed that a substantial clustering of EPCs below 200 kWh/(m2a) does exist for all building groups.

We observe a substantial increase of certificates since 2013 that is related to a slight reduction in their values. The linear fits (dashed lines), which are computed for EPCs issued after the year 2000, manifest a tendency to a very slow decrease for D1 (the smaller dwellings), while for D2 and D3 the decrease is more pronounced.

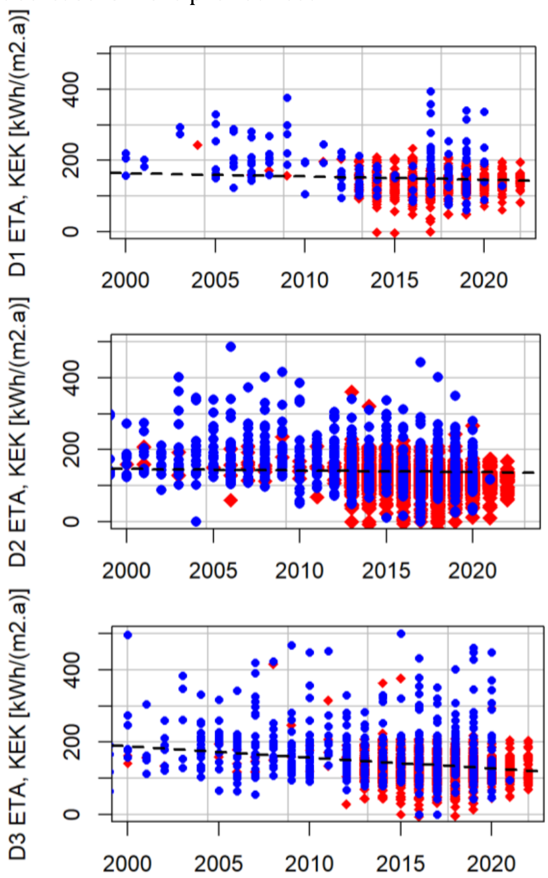


Figure 7 - Top to bottom: EPC values (red diamonds for ETA, blue dots for KEK) versus construction or renovation year for D1, D2 and D3 respectively

3.5 Readiness of the Estonian detached houses

By prolonging the linear fit in **Figure 7**, a rough estimation of the year (ZEB year) when the three building clusters should reach the Zero Energy Building (ZEB) status can be performed. This is illustrated in **Figure 8** below for the full dataset of 18122 EPCs (red solid line for 2000-2022 data.

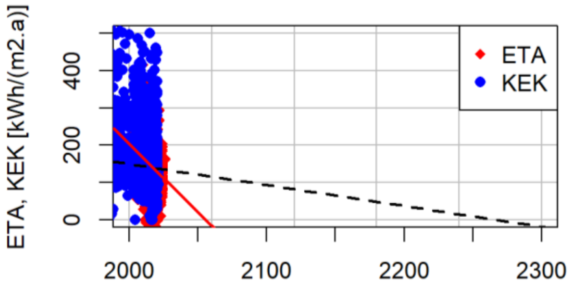


Figure 8 - Estimations of ZEB year for the full 18122 EPCs database, red diamonds for ETA and blue dots for KEK, against construction or renovation year. Dashed line: post-COVID 19, red solid line: pre-COVID-19

Identifying the intersection of the fit with the x-axis, namely the ZEB year, gives 2269 for the full dataset. Distinguishing among the three building categories gives 2178 for D1, 2351 for D2 and 2062 for D3. These are listed in **Tab. 3**.

3.4 Correlations with age and heated area

Pearson correlations between EPC value, heated area A and construction or renovation year have been performed for the three datasets; these are reported in **Tab. 3**. For completeness, also the ZEB year and linear fit slope are added.

Tab. 3 - Correlations for EPC values against Area and construction/renovation Year, ZEB year and fit slope

Dataset	EPC vs A	EPC vs Year	ZEB Year	Slope
D1	-0.082	-0.556	2178	-0.93
D2	-0.181	-0.020	2351	-0.42
D3	-0.217	-0.311	2062	-3.05

The correlations are very weak for all datasets.

3.2 Effect of the COVID-19 pandemic

In order to estimate whether the pandemic has impacted the energy consumption in detached houses, we have compared two ZEB year estimations. The pre-COVID fit was computed with 2000-2019 data, while the post-COVID fit regarded 2000-2022 data, as in **Tab. 3**. The result for the full database and for D1, D2, D3 is reported in **Tab. 4**.

Tab. 4 - Estimates of the COVID-19 impact on the ZEB year for all the building clusters

Data	pre-COVID	Post-COVID	Diff.
TOTAL	2056	2269	+213
D1	2071	2178	+107
D2	2052	2351	+299
D3	2057	2062	+5

At first sight, one might conclude that it was the forced homeworking since 2020 to unquestionably increase the energy consumption of private dwellings to abruptly. However, by breaking down the EPCs into ETA (calculated) and KEK (measured), it can be shown that the main driver of the increase were the regulations, which directly affect the simulations parameters returning the ETAs. We recall indeed from Section 2 that after 2018, class A must be reached only by the D3 cluster, while D1 and D2 must only meet class B requirements. This is perfectly in line with our results, as for D1 and D2 a fit using only ETAs marks a very sharp ZEB year delay. D3 instead exhibited even a slight 7-year ZEB improvement, according to an only-ETA fit.

The KEK (measured) values, conversely, are directly related to occupancy and to COVID-19. These showed only a small delay for the ZEB year for D1 and D3, of order ~5 years, whilst for D2 the year estimation did not even change. In other words, the pandemic effect proved to be very limited, if not even absent.

3.3 Benchmarking tables

The distribution fitting, as explained in Section 2, was performed by means of gaussian mixture [10]. In **Figure 9** we report the result for D1, with a trimodal distribution (overlap of three Gaussian distributions) giving the best fit. D2 also required a trimodal fitting (**Figure 10**) with more distinct modes: the peaks are located at EPC=110, 130 and 150 kWh/(m^{2}a); these correspond to the midpoints of class A, B and C respectively (see **Tab. 2** and **Figure 5**). This is consistent with the full dataset as well, as illustrated in **Figure 3**, because D2 is the largest subcluster, corresponding to 56% of the entire detached houses database. The same holds for D3, which exhibits the same three peaks plus a fourth one at 90 kWh/(m^{2}a), shown in **Figure 11**.

Only D1, namely those buildings with heated area smaller than 120 m^2, carries only two sharp peaks, one at 110 and the other at 150 kWh/(m^{2}a), corresponding to classes A and C.

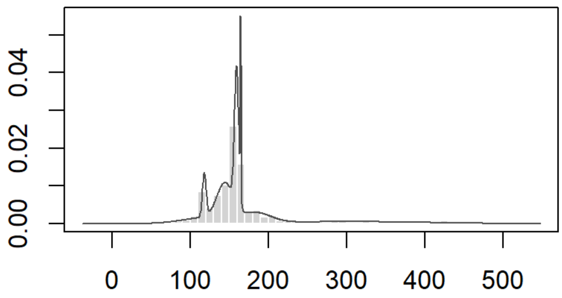


Figure 9 – Trimodal density distribution fit for D1

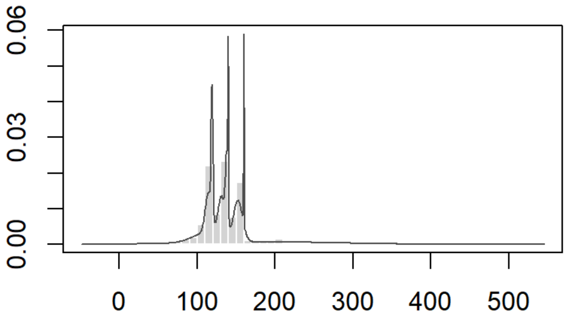


Figure 10 – Trimodal density distribution fit for D2

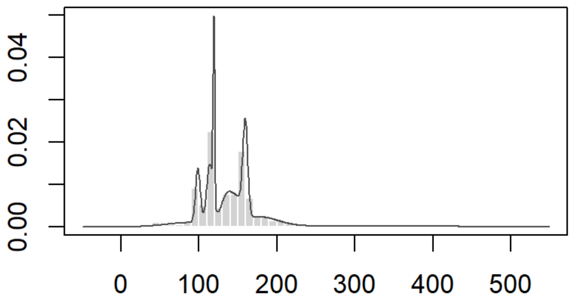


Figure 11 – Multimodal density distribution fit for D3

Through distribution fitting, we were able to construct benchmarking tables for the three clusters D1, D2 and D3 by integrating each distribution and retriving the according cumulative percentage.

Tab. 5 - Benchmarking table for Estonian detached houses D1 with A<120m^2, EPC [kWh/(m^{2}a)]

Score	Cumul. (%)	EPC ≥	EPC <
100	0	0	64.78
99	1	64.78	87.51
98	2	87.51	105.38
95	5	105.38	116.23
90	10	116.23	119.08
85	15	119.08	124.85
80	20	124.85	132.67
75	25	132.67	137.85
70	30	137.85	145.67
60	40	145.67	153.66
50	50	153.66	157.73
40	60	157.73	160.1
30	70	160.1	163.6
20	80	163.6	182.19
10	90	182.19	∞

Tab. 6 - Benchmarking table for Estonian detached houses D2 with A=120-220 m^2, EPC [kWh/(m^{2}a)]

Score	Cumul. (%)	EPC ≥	EPC <
100	0	0	48.73
99	1	48.73	75.15
98	2	75.15	96.39
95	5	96.39	109.13
90	10	109.13	113.41
85	15	113.41	116.82
80	20	116.82	118.51
75	25	118.51	119.67
70	30	119.67	128.78
60	40	128.78	135.84
50	50	135.84	139.32
40	60	139.32	146.23
30	70	146.23	155.29
20	80	155.29	160.61
10	90	160.61	∞

Tab. 7 - Benchmarking table for Estonian detached houses D3 with A>220 m^2, EPC [kWh/(m^{2}a)]

Score	Cumul. (%)	EPC ≥	EPC <
100	0	0	53.98
99	1	53.98	68.61
98	2	68.61	93.97
95	5	93.97	98.88
90	10	98.88	104.39
85	15	104.39	111.84
80	20	111.84	115.32
75	25	115.32	118.07
70	30	118.07	120.46
60	40	120.46	136.84
50	50	136.84	149.32
40	60	149.32	157.68

30	70	157.68	161.85
20	80	161.85	187.84
10	90	187.84	∞

The benchmarking tables for the D1, D2 and D3 datasets are given in **Tab. 5**, **Tab. 6** and **Tab. 7**. The table for the full dataset of detached houses was reported in [8].

4. Discussion

The improvement in energy efficiency of Estonian detached houses is quite evident by looking at **Figure 4**, which displays the time trend of EPCs for the full database. The EPC values have been decreasing steadily during the past six-seven years. This is mirrored by the EPC labels classification that is broken down by heated area, **Figure 5**: a clear shift from C to B class since 2019 exists for all the clusters D1, D2 and D3, together with a substantial increase of A class certificates. One ought to conclude that this should be the effect of renovation incentives. Nevertheless, **Figure 6** and **Figure 7** provide a closer look into the matter. **Figure 6** shows an evident decrease of EPC values for D2 (the largest subcluster by far, 56% of the full database), while for D3 the trend is mostly unchanged and D1 experiences even larger values for 2019-2022. Considering again **Figure 5** is quite enlightening if one looks at the Class A share compared to Class B. This is substantially larger for D3, while D2 and D1 have a much smaller amount of A Class EPCs compared to B Class EPCs. This clearly suggests that renovation campaigns and stricter energy class requirements were successful for larger houses of heated area larger than 220 m^2, while they probably have not been influential for smaller buildings. Notice also how in **Figure 6**, D1 and D2 show a plateau at resp. ~150 and ~160 kWh/(m2a), i.e. at Class B and Class C. D3 exhibits the same behaviour at 160 kWh/(m^2a), the middle of Class D.

Figure 7 is mostly interesting in regard to the structural difference between ETA (computed) and KEK (measured) values. Basically all the outliers for any Di subcluster are KEK, while the calculated EPCs tend to converge rather nicely with the passing of time. However, a steady increment in the lowest ETA values does exist (**Figure 7**) and looking at **Tab. 3**, the absence of correlation between construction (or renovation) year and EPC signify that the whole picture is not so simple. One cannot even suspect the ETA values to have been deliberately underestimated for complying with the regulations, or even the existence of some systematic error in the simulation software.

Keeping in mind the above speculation, a very simple and objective summary of the energy consumption status of the Estonian detached houses is given in **Tab. 3** and **Tab. 4**. The ZEB year provides indeed a simple yet powerful means of judgement: out of the three building subclasses, only D3 (i.e. largest detached houses) are forecast to reach the ZEB status within the next 40 years, in any case well beyond 2050. **Tab. 4** in particular illustrates how the sudden occupancy increment during the COVID-19 emergency was *not* causing the worsening of ZEB year forecasts, if compared to 2000-2019 estimates. On the other hand, the national regulations post-2018 favoured a clustering of EPCs towards class B from class A for D1 and D2, which can be somehow noticed also in **Figure 7**. The effect of regulations and incentives could not be more manifest.

The distribution fitting in **Figure 9** and the benchmarking tables clearly show that the EPCs concentrate within classes A and C, with very long tails beyond class D carrying only a few values. Overall, the three distinct building groups do not exhibit any clustering towards class A. Nevertheless, this seems likely to happen at some point during the next decades anyway at least for D3. Smaller houses, namely the D1 and D2 subgroups, give a very pessimistic forecast as they will tend to cluster around class B for a long time, as in **Figure 5**, unless further restrictions on energy consumption are imposed in the near future.

5. Conclusion

This study examined a large dataset of EPC certificates of nearly 19000 Estonian detached houses (single detached or terraced dwellings, portions with dedicated entrance, two or three apartment houses etc.). This allowed portraying the status of the energy performance of old as well as new or recently renovated buildings.
A thorough statistical investigation unveiled general characteristics as well as specific features of the subgroups D1, D2 and D3, which were created by the national regulation bodies according to heated area. We found that the cluster of largest houses D3, with an area >220 m^2, benefitted from renovation campaigns and is expected to reach the ZEB status in 2062. Smaller dwellings instead showed an EPC clustering towards B Class values, which implied no ZEB status in sight unless stricter regulations, or a coordinated effort, will come into force.
The above considerations are confirmed by a dedicated assessment of the COVID-19 pandemic, which compared ZEB year estimations by fits based on two distinct periods: 2000-2019 (pre-COVID) and 2000-2022 (post-COVID). Contrary to expectations, the forced remote working from home in 2020-2022 was only marginally responsible for the massive delay in the ZEB year forecast of as much as 300 years. We have shown that the culprit was instead the 2018 governmental regulations, inducing a shift in the ETAs from Class A to Class B for some categories. The role of legislation will be thus critical in the future, to ensure that most detached homes will reach the ZEB status by the year 2050 or soon afterwards.
Another novel result is constituted by benchmarking

tables that are obtained via distribution fitting. These can be used to compare a given building with the entire building stock, for evaluating its energy efficiency with a dedicated rating system. Such tables also allow identifying representative buildings for e.g. real-time consumption monitoring; therefore they can be an important tool for detailed energy auditing.
Naturally, the analysis here provided should be updated with newer datasets, extended to other building typologies and refined with more sophisticated statistical methods, especially on the side of readiness forecasting and distribution fitting.

6. Acknowledgement

Support is acknowledged from the European Regional Development Fund via the Estonian Centre of Excellence in Zero Energy and Resource Efficient Smart Buildings and Districts ZEBE, grant 2014-2020.4.01.15-0016, from the Estonian Research Council through the grants PSG409, PRG658, and DigiAudit, from the Estonian Ministry of Education and Research and European Regional Fund (grant 2014-2020.4.01.20-0289), and by the European Commission through the H2020 project Finest Twins (grant No. 856602).

References

[1] A. Magrini, L. Gobbi and F. d'Ambrosio, "Energy audit of public buildings: The energy consumption of a university with modern and historical buildings. Some results," *Energy Procedia,* vol. 101, pp. 169-175, 2016.

[2] ASHRAE, "Procedures for Commercial Building Energy Audits, 2nd Edition," ASHRAE, Atlanta, 2011.

[3] M. C. Baechler, "A Guide to Energy Audits," Pacific Northwest National Lab. (PNNL), Richland, WA (United States), 2011.

[4] A. Hamburg and T. Kalamees, "How well are energy performance objectives being achieved in renovated apartment buildings in Estonia?," *Energy And Buildings,* vol. 199, pp. 332-341, 2019.

[5] W. Chung, Y. V. Hui and M. Lam, "Benchmarking the energy efficiency of commercial buildings," *Applied Energy,* vol. 83, no. 1, pp. 1-14, 2006.

[6] J. Roth, B. Lim, K. J. Rishee and D. Grueneich, "Examining the feasibility of using open data to benchmark building energy usage in cities: A data science and policy perspective," *Energy Policy,* vol. 139, no. 111327, 2020.

[7] H. S. Park, M. H. Lee, L. H. Kang, T. H. Hong and J. W. Jeong, "Development of a new energy benchmark for improving the operational rating system of office buildings using various data-mining techniques," *Applied Energy,* pp. 225-237, 2016.

[8] A. Ferrantelli, J. Belikov, E. Petlenkov, M. Thalfeldt and J. Kurnitski, "Evaluating the energy readiness of national building stocks through benchmarking," *IEEE Access*, vol. 10, pp. 45430-45443, 2022.

[9] H. G. Kim and S. S. Kim, "Development of Energy Benchmarks for Office Buildings Using the National Energy Consumption Database," *Energies,* no. 13, p. 950, 2020.

[10] L. Scrucca, M. Fop, T. B. Murphy and A. E. Raftery, "mclust 5: clustering, classification and density estimation using Gaussian finite mixture models.," *The R Journal,* vol. 8, no. 1, p. 289–317, 2016.

[11] R Core Team;, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2020.