IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 7, 2022, accepted April 24, 2022, date of publication April 26, 2022, date of current version May 3, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3170716

Evaluating the Energy Readiness of National Building Stocks Through Benchmarking

ANDREA FERRANTELLI[1,2], **JURI BELIKOV**[1,3], **(Senior Member, IEEE), EDUARD PETLENKOV**[1,4], **MARTIN THALFELDT**[1,2], **AND JAREK KURNITSKI**[1,2,5]

[1]FinEst Centre for Smart Cities (Finest Centre), Tallinn University of Technology, 19086 Tallinn, Estonia
[2]Department of Civil Engineering and Architecture, Tallinn University of Technology, 19086 Tallinn, Estonia
[3]Department of Software Science, Tallinn University of Technology, 12618 Tallinn, Estonia
[4]Department of Computer Systems, Tallinn University of Technology, 12618 Tallinn, Estonia
[5]Department of Civil Engineering, Aalto University, 00076 Aalto, Finland

Corresponding author: Andrea Ferrantelli (andrea.ferrantelli@taltech.ee)

This work was supported in part by the European Regional Development Fund through the Estonian Centre of Excellence in Zero Energy and Resource Efficient Smart Buildings and Districts ZEBE under Grant 2014-2020.4.01.15-0016; in part by the Estonian Research Council through DigiAudit under Grant PSG409, Grant PRG658, and Grant PRG1463; in part by the Estonian Ministry of Education and Research and European Regional Fund under Grant 2014-2020.4.01.20-0289; and in part by the European Commission through the H2020 Project Finest Twins under Grant 856602.

ABSTRACT Evaluating the energy performance of existing buildings is critical for improving the efficiency and resilience of the building stock as a whole. The importance of this information holds at different scales, both locally and at the national and international levels. A major problem arises from the difficulty in obtaining information from existing buildings; often, the only available data are the yearly consumption per unit area, typically corresponding to the energy performance certificate (EPC). This paper shows how to address concerns of practical relevance with a limited number of variables by examining an EPC national database (including the major cities of Tallinn, Pärnu, Tartu, and others) that provides only EPCs, construction/renovation year and heated area. Through a systematic statistical investigation of nearly 35 000 EPCs of educational, office, commercial and other building typologies, we i) characterise the time evolution of EPC classes, ii) evaluate the impact of incentives pre/post-renovations, and iii) create benchmarking tables that allow comparisons of a specific building with the existing stock to identify representative buildings for detailed auditing. The readiness of the Estonian building stock could thus be evaluated by linear fitting. All new and renovated buildings are estimated to achieve the zero-energy building (ZEB) status by 2050; remarkably, for some categories, this will occur already in the present decade if the identified linear trends persist. Additionally, we investigated whether the COVID-19 pandemic has affected building stock readiness by comparing pre- and post-2020 ZEB year fit estimations. Contrary to what was expected, the change in working habits affected some building types only marginally, while the national regulations played a prominent role. Detached private houses exhibited a pronounced worsening in readiness, while the educational and entertainment sectors benefited from specific energy labelling remodulations.

INDEX TERMS Benchmark testing, energy consumption, energy management, energy efficiency, statistical analysis.

I. INTRODUCTION

The energy performance and resilience of the building stock constitute key concepts in the current efforts for reducing CO_2 emissions and, ultimately, for achieving the decarbonisation goals. Consistent with national as well as supranational agendas such as the European Green Deal [1], policymakers, governmental agencies and municipalities need to implement an effective approach to the management and distribution of energy resources.

The European Green Deal was reinforced at the end of 2021, when the European Commission adopted a major revision of the Energy Performance of Buildings Directive (EPBD). The revision aims to accelerate building

The associate editor coordinating the review of this manuscript and approving it for publication was Alexander Micallef.

renovations, reduce GHG (greenhouse gas) emissions and energy consumption, and promote the usage of renewable energy in buildings. A new European Union definition of a 'zero emissions building' will be introduced for application to all new buildings from 2027 as well as to all those that will be renovated from 2030. In particular, net zero GHG emissions (i.e., climate neutrality) should be achieved by 2050. The resolution underlines the need for the existing building stock to be renovated into nearly zero-energy (nZEB) buildings to achieve carbon neutrality by 2050 at the latest [2].

Realising such an ambitious agenda is clearly a very challenging task that relies on, among other factors, an accurate estimation of the national building stocks' current energy efficiency as well as their readiness to achieve the nZEB or ZEB status by 2050. This challenge constitutes the main motivation for the study performed in this paper.

Real-time and detailed information about the main energy consumers and issues is important for achieving such high performance [3]. Such a so-called "building energy audit" [4] usually focuses on numerical models that are calibrated through actual data on operation [5], with application to a single building or further extension to a region or even to an entire country [6]. One of the main purposes is to provide automated analyses and *dynamic* energy performance certificates (EPCs).

An EPC is a key tool for evaluating the energy efficiency of a building, either as measured or estimated with a variety of computational methods [7]. Adopted in the European countries since the early 1990s, EPC databases should ideally provide an objective and unbiased picture of the current energy consumption level. They allow, e.g., allocating a building into energy performance ratings (from A, the best, to F, the worst) for a classification within the market [8]; when considered collectively, they constitute EPC databases that can be used in *building energy benchmarking*.

A benchmark is generally used as a reference to measure performance using a specific indicator; choosing the energy consumption of a building (the EPC) as the indicator leads to the concept of "building energy benchmarking" [9]. By comparing the individual EPC relative to its previous values (i.e., the building's typical condition), one can determine whether the energy and physical performances of the building are consistent. The natural extension is comparing the EPC to those of other similar buildings and/or a reference building (the benchmark); this allows the building to be evaluated regarding whether it performs well in terms of energy consumption, thus providing information to stakeholders and motivating energy retrofits [10].

The purpose is ultimately to define a performance database of peer groups with numerous existing buildings. Among other applications, which shall be addressed by the study at hand, this allows the identification of representative units in a specific category. This information is essential for the auditing and implementation of *dynamic* energy performance certificates, which are a critical tool for accomplishing the European Green Deal objectives.

A. LITERATURE REVIEW AND RESEARCH GAPS

Building energy benchmarking has a variety of applications and synergies with other fields, making the related literature quite substantial. The uses of benchmarking cover all the components of an energy distribution network, from the smaller scale of a single building to an entire "smart city" [11]. Comparison with benchmark energy consumption datasets constitutes an important tool in forecasting a smart building's energy consumption for planning and operating power generation towards efficient smart grid energy management [11]. To this aim, transmission and distribution networks called "power system benchmarks" are implemented by using anonymised clustering, statistical sampling, heuristic algorithms, etc. These were reviewed in [12].

Energy benchmarks can then be defined through regression analysis, which accounts for correlations of building operation with climate, occupancy density, heated area, etc. [13]. The resulting regression curves are usually very predictive [14]; using key variables allows an objective and effective benchmark normalisation for comparing energy performance. Nevertheless, as observed in [9], the predictive power of this approach relies excessively on the independent variables; the applicability of regression analysis is accordingly limited to datasets with abundant and diverse data.

Building performance simulations (BPSs) are a powerful resource, allowing the prediction of the energy consumption of a building by using thermodynamic principles [15]. However, obtaining reliable energy consumption estimates with BPS is not straightforward, as the actual energy use in buildings can be up to three times the calculated value [16] (this is usually called the "energy performance gap" [17]). The role of occupants' behaviour is remarkably critical in this respect [18], as it strongly affects both HVAC [3], [19] and plug loads [20]. Additionally, the work [21] showed that conventional (and *static*) assumptions are often invalidated by field data of tenant plug and light loads; this was confirmed in [22], where the overestimation of seasonality was also stressed. Domestic hot water (DHW) consumption [23] constitutes a large portion of the overall energy use of the residential building sector [24] as well.

In other words, buildings can no longer be regarded as static and isolated entities depending only on the local climate; rather, their *dynamic* response to energy usage must be accounted for. Data mining techniques are very effective in overcoming this difficulty. They do not need a large number of explanatory variables, nor do they use physical knowledge for estimating the building's energy use as in BPS. These models are built upon empirical training, creating an optimal profile with a minimal amount of input data. An example is given in [25], where sensitivity analysis was postulated as a feature selection problem and building grouping was postulated as a clustering problem. An accordingly defined data-driven framework proved to be more predictive and accurate than the standard Energy Star benchmark system.

The authors of [3] used total and conditioned area, together with DHW open-data from 10 cities, and by using nonlinear methods (random forest and lasso regression), they identified important building characteristics and proposed a benchmarking method based on normalised consumption. [26] analysed 1072 office buildings using the Source Energy Use Intensity (Source EUI), which corresponds to the EPC, as an independent variable. By using correlations, decision tree (DT) and analysis of variance (ANOVA), they developed six types of energy benchmarks for office buildings according to gross floor area (GFA) and the building use ratio (BUR). Finally, the study in [19] formulated a method to compare the energy performance in different climates, using degree days, solar-air temperature and economic insulation thickness to normalise space heating and cooling. Introducing some normalisation factors and performing dynamic whole-year BPS allowed the location of buildings in different climatic areas and supported a comparison of the nearly zero energy building (nZEB) national requirements with European Commission benchmarks.

Despite all these valuable achievements, a systematic assessment of a national EPC database, including a methodology that is specifically tuned towards building energy benchmarking, still needs to be realised. Currently, following the above discussion, an important tool for aiding both building management and authorities in achieving European Green Deal targets is therefore missing.

B. CONTRIBUTION OF THIS STUDY

Inspired by the latest developments, in this paper, we carry out a comprehensive statistical assessment of a substantial EPC database from Estonia, covering all the major cities (Tallinn, Pärnu, Tartu) as well as minor settlements. To highlight a simple yet systematic in-depth analysis that can reveal time trends and data structures, we aim to shed light on the actual performance status of the country's building stock. This approach is all the more useful since, in addition to portraying various energy-related aspects of diverse building typologies, ineffective renovation campaigns or improper energy classifications can be uncovered as well.

Specifically, the key contributions of this work include the following:

1) A comprehensive set of 11 building classes of interest is identified, which are analysed and compared through box plots, EPC time evolution, energy label time trends, distribution fitting, correlations with construction year and heated area, and readiness estimation.
2) The time evolution of EPC classes is identified for several clusters and correlated with normative regulations that have been remodulated through the years.
3) Evaluating the impact of incentives pre/ post-renovations is accomplished by combining a qualitative as well as a quantitative analysis of the time trends of both measured and calculated EPCs for all categories.
4) Derivation of benchmarking tables that allow comparison of a specific building with the existing stock is accomplished by distribution fitting of the full EPC database for each distinct cluster, which in general exhibits nontrivial distributions.
5) The readiness of the Estonian building stock, namely, an estimation of how close it is to achieving the ZEB status, is evaluated by linearly fitting the EPCs for recently built or renovated buildings. The respective ZEB year is then calculated for each and every cluster of the database.
6) By comparing the ZEB year fit estimations that were obtained with pre- and post-2020 data, we also quantify an eventual impact of COVID-19 on the Estonian building stock readiness.
7) To exploit all the data available in the database, we also investigated whether the energy consumption correlates strongly with heated area; previous analysis found that this holds for the gross floor area [26].

In summary, the present study identifies the building typologies that are most problematic, those that are the most dependent on renovation incentives and national regulations, and formulates suggestions and recommendations for a new EPC class scaling. Our analysis is articulated as follows: In Section II, we discuss the EPC dataset, including data cleaning and the national policy for the energy labels; the distribution fitting methods and generation of benchmarking tables are described as well. Section III shows the results in thematic subsections, covering the time trend of EPC classes, the impact of renovation incentives, benchmarking tables, correlations with building age and heated area, and finally the readiness of the Estonian building stock. These findings are then critically analysed in Section IV, and conclusions are drawn in Section V. Finally, Appendix displays the benchmarking tables for the building categories that were selected in this study.

II. METHODS

A. EPC DATABASE AND ENERGY LABELS

The data assessment and statistical analysis illustrated in this paper were performed with R software using various supportive packages [27], explicitly recalled whenever needed.

The EPC database under examination consisted of a total of 34 625 certificates that were issued between 2004 and the first two months of 2022. This database provides a thorough overview of the country's building stock energetic history for nearly twenty years, covering the unprecedented situation generated by the COVID-19 pandemic as well. However, any data released before 2008 are statistically irrelevant, as indicated in Fig. 1. Furthermore, the Estonian legislation radically changed the energy labelling system in 2013; in what follows, we discuss the time interval 2013–2022. In our database, the tabbed variables were building ID, construction year, renovation year, heated area, EPC certificate issue and expiration date, ETA, KEK and resulting energy label.

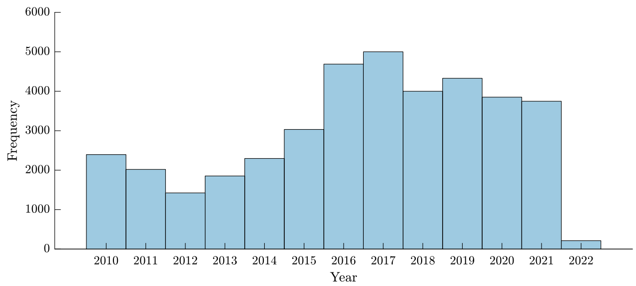


FIGURE 1. Total number of issued energy performance certificates per year for the Estonian building stock, full database including 34 625 entries.

TABLE 1. Energy labels for kindergartens and schools according to the estonian directive [8], EPC expressed in [kWh/(m^{2}a)].

Ener. label	Kindergartens	Schools
A	ETA/KEK $\leq$ 100	ETA/KEK $\leq$ 100
B	101 $\leq$ETA/KEK $\leq$ 120	101 $\leq$ ETA/KEK $\leq$ 120
C	121 $\leq$ETA/KEK $\leq$ 165	121 $\leq$ ETA/KEK $\leq$ 160
D	166 $\leq$ETA/KEK $\leq$ 220	161 $\leq$ ETA/KEK $\leq$ 200
E	221 $\leq$ETA/KEK $\leq$280	201 $\leq$ ETA/KEK $\leq$ 250
F	281 $\leq$ETA/KEK $\leq$ 360	251 $\leq$ ETA/KEK $\leq$ 310
G	361 $\leq$ETA/KEK $\leq$ 460	311 $\leq$ ETA/KEK $\leq$ 390
H	ETA/KEK $\geq$ 461	ETA/KEK $\geq$ 391

An EPC certificate is usually expressed in kWh/(m^{2}a), and in the Estonian country regulations, it is regarded as ‘‘KEK’’ if measured or as ‘‘ETA’’ if calculated (usually with simulation software). Since the national regulation does not distinguish between ETA (calculated) and KEK (measured), the names were eventually dropped, and each ETA or KEK value was simply regarded as ‘‘EPC value’’ unless the distinction was essential for the analysis.

In the database, each ETA/KEK was associated with a building typology via a `type_id` field, corresponding to different definitions of energy labels. As per European Union (EU) standard regulations, the latter are rated from A (best) to H (worst), corresponding to energy consumption (=EPC value) bounds that are country-specific. For instance, the current classification for kindergartens and schools, which is discussed in Section II-A, is given in Table 1.

Since approximately 7500 certificates ($\sim$22% of the total) did not specify any `type_id`, they were ignored for the energy labels analysis and included only in Fig. 1. As a result, out of all the $\sim$30 building types, we identified the 17 most representative typologies, such as educational and office buildings (i.e., we disregarded animal feed storage and forestry, hunting or fishing buildings and similar buildings). Afterwards, the 11 typologies with greatest available data were selected for the assessment reported in Section III.

Specifically, the Dwellings cluster (K4) comprises both single detached and terraced houses of various typologies (portions with dedicated entrances, two or three apartment houses, etc.). The energy labels are divided into three groups according to the heated area A, namely, for A<120 m^2, A=120 m^2–220 m^2 and A>220 m^2.

TABLE 2. Building cluster breakdown for the final database featuring 25 979 EPCs, including number N, mean annual energy consumption M and standard deviation SD (both in [kWh/(m^{2}a)]).

ID	Building type	N	M	SD
K1	Kindergartens	540	220.20	81.33
K2	Schools	88	185.9	66.76
K3	Educational (all)	1132	204.1	80.69
K4	Dwellings	18122	143.2	51.04
K5	Apartment buildings	3945	172.6	63.78
K6	Office buildings	1081	197.97	113.83
K7	Commerce, Services	618	228.35	149.93
K8	Entertainment	358	207	92.6
K9	Sports halls	229	246.9	149.06
K10	Welfare buildings	255	232.97	128.45
K11	Hotels, Dormitories	239	197.24	72.23

In Estonia, the nZEB level, i.e., The EPC value corresponding to class A, was first defined in 2013 and then revised in 2018. The reworked cost-optimality calculations and changes in nonrenewable primary energy factors caused the revision of EPC values of class A. Typically, the changes did not exceed 5 kWh/(m^{2}a) except for dwellings, which initially had an EPC value of 50 kWh/(m^{2}a) regardless of the building heated floor area. During the revision, the corresponding values were increased significantly, and currently, they also depend on the floor area. Additionally, the EPC certificate class A must be reached only by dwellings with heated floor areas above 220 m^2, while other types of dwellings must meet class B requirements.

Regarding this cluster, as we are not interested in discussing their energy labels separately, the three subgroups were aggregated together into a single large group since the EPC distribution is independent of the energy labels.

Finally, four types were grouped into two macroclusters since they have the exact same ETA/KEK energy label thresholds according to the legislation. These are 1) commercial and services and 2) hotels, hostels and dormitories. Only a few outliers were ignored, with a common cut-off set at 1500 kWh/(m^{2}a) since the highest threshold for class H in Estonia is 1350 kWh/(m^{2}a). This data selection and clustering resulted in the energy performance analysis of 25 979 buildings, or $\sim$75% of the total original records. A summary statistical description of the respective 11 building clusters, labelled sequentially from K1 to K11, is given in Table 2; the respective box plot is pictured in Fig. 2.

Kindergartens K1 and schools K2 (i.e., elementary and high schools) were first analysed separately because they provide particularly interesting insight into the time evolution of ETA/KEK and energy labels, including the effect of renovations (Sections III-B and III-A). To address the educational sector as a whole, these buildings were then grouped together with university and research facilities to form the ‘‘Educational (all)’’ cluster K3.

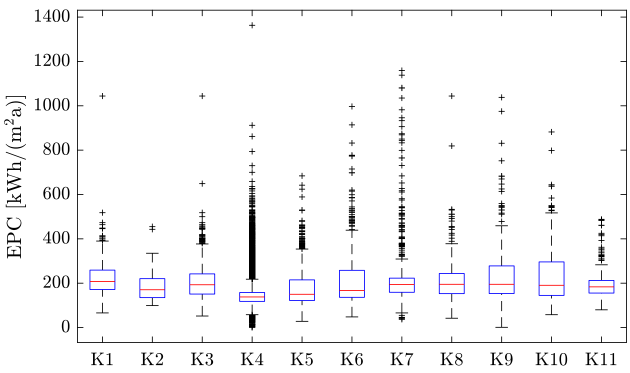


FIGURE 2. Box plot summarising the EPCs for all building categories.

B. FITTING DISTRIBUTIONS AND BENCHMARKING TABLES

Since the EPC database provided only a few variables for each building, essentially only construction/renovation date and heated area, we opted for a benchmarking method that requires the minimum amount of input data.

Following [9], we fitted the ensemble of ETA/KEK certificates with a probability distribution. Integrating the fitting function returned the empirical cumulative distribution function (CDF), whose domain is mapped into a smooth curve. The smoothness identifies a unique EPC at any given ratio (or percentage) of the given dataset and quantifies the ranges corresponding to different quantiles of the EPC dataset distribution.

Given its EPC, it is therefore possible to track very precisely where a specific building is placed in the energy consumption spectrum of a comparable category. Additionally, a 10–100 point scale was added to embed a rating system that was directly proportional to energy efficiency. The final result is a so-called "benchmarking table".

As are specifically concerned with mapping the energy efficiency of the Estonian building stock as our case study, we used as an upper cut-off an EPC value of 500 kWh/(m^2a), which coincides with class H for all the building clusters considered here. Any EPC that exceeded this value was regarded as an out-of-scale outlier and was dropped from the fit distribution calculations.

Fitting and integration were accomplished with the R package *fitdistrplus* [28], which allows the user to choose from among a selection of discrete (binomial, negative binomial, geometric, hypergeometric, Poisson) as well as continuous distributions (normal, lognormal, exponential, gamma, beta, uniform and logistic). To identify the best fitting distributions, we first employed the Cullen and Frey graph, otherwise known as the skewness-kurtosis plot, as well as four classic goodness-of-fit plots [29]. These comprise 1) a density function of the fitted distribution overlapped with the histogram of the empirical distribution; 2) a CDF plot of the empirical and fitted distributions; 3) a Q-Q plot of empirical (y-axis) against theoretical quantiles (x-axis); and 4) a P-P plot of the empirical distribution function computed at each data point (y-axis) versus the fitted distribution function (x-axis).

For many building categories, the ETA/KEK distribution was right-skewed, so we chose a maximum goodness-of-fit estimation (MGE) method that gives more weight to data at one tail of the distribution [28]. The more common maximum likelihood estimation (MLE) was used instead whenever the tails were irrelevant.

For the cases in which choosing was difficult, additional quantitative parameters were the log-likelihood, AIC (Akaike information criterion) and BIC (Bayesian information criterion). When comparing fitting distributions, the log-likelihood and AIC should be larger, while the BIC should be smaller. In the presence of distinct modes that required a multimodal distribution (cluster K4), the R package *mixtools* [30], which fits the data based on maximum likelihood with a combination of a Newton-type and EM algorithms, was used.

This procedure was repeated for each building cluster that is reported in Table 2, labelled from K1 to K11.

III. RESULTS

A. EPC CLASS TREND VS TIME

In Estonia, the methodology for defining energy labels has been changed twice. These labels are comparable until 2013, when the energy efficiency parameters for the labelling were remodulated, until the end of 2017. The requirements then became stricter in 2018, and after 01.01.2019, every newly constructed or renovated building needed to comply with stricter rules in terms of energy labels. For instance, schools K2 with ETA/KEK $<$ 100 kWh/(m^2a) belong to class A only if the certificate was released before 01.01.2018; otherwise, after that date, the A class corresponds to ETA/KEK $<$ 90 kWh/(m^2a). Figure 3 illustrates the effect of these different legislations by breaking down the ETA/KEK certificates for Estonian kindergartens K1 (top), apartment buildings K5 (middle), and office buildings K6 (bottom), according to the certificate issue year. For K1, there seems to have been an improvement from class D in 2013–2017 to class C ($\sim$60%) in 2018, when the classification became stricter. Class C for 2018 was particularly dominant. After 2019, however, the largest % was D again, with a large reduction in class C certificates and a corresponding increase of class E. The only positive outcome is that class B doubled with respect to 2018, and class A increased as well.

Apartment buildings K5 present a more optimistic outcome, as the majority of counts oscillate between class B and C. After 2019, there was also a massive increase in class A buildings, following incentives. Note, however, the large variance in the EPCs, with classes E and F being consistently large throughout the period.

Finally, office buildings K6 show a somewhat odd behaviour: while for 2013–2017, the largest class was C,

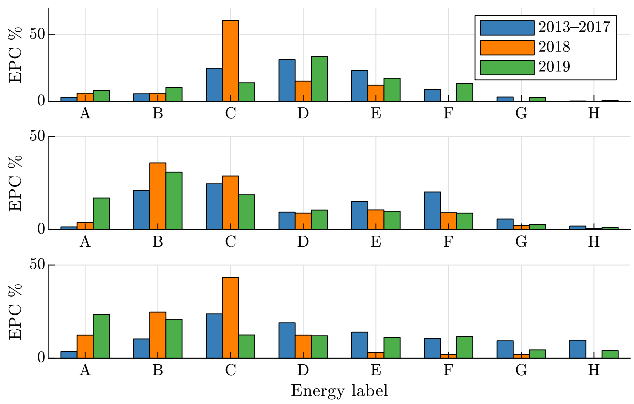


FIGURE 3. ETA/KEK certificate classes (percentage over the total) for kindergarten K1 (top), apartment buildings K5 (middle) and office buildings K6 (bottom), grouped by certificate issue year: 2013–2017 (blue), 2018 (orange), 2019–2022 (green).

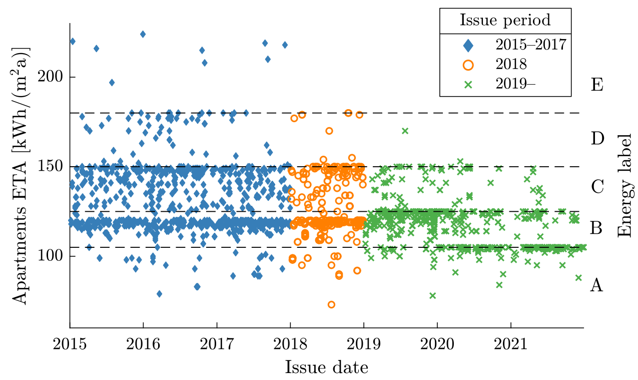


FIGURE 5. ETA certificates for apartments K5 versus date of issue: 2015–2017 (blue diamonds), 2018 (orange circles), 2019–2022 (green crosses). The dashed lines denote separation ranges with respect to the energy label.

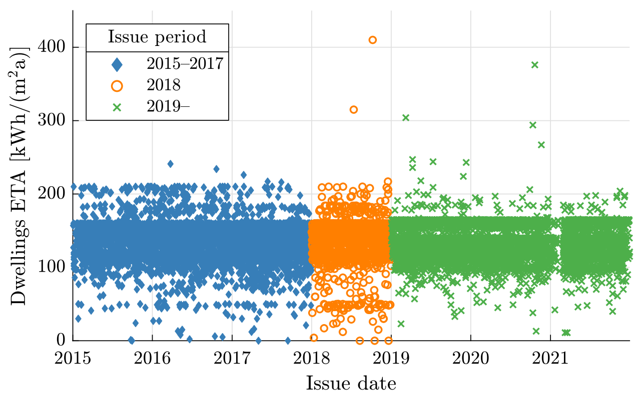


FIGURE 4. ETA certificates for dwellings K4 versus date of issue: 2015–2017 (blue diamonds), 2018 (orange circles), 2019–2022 (green crosses).

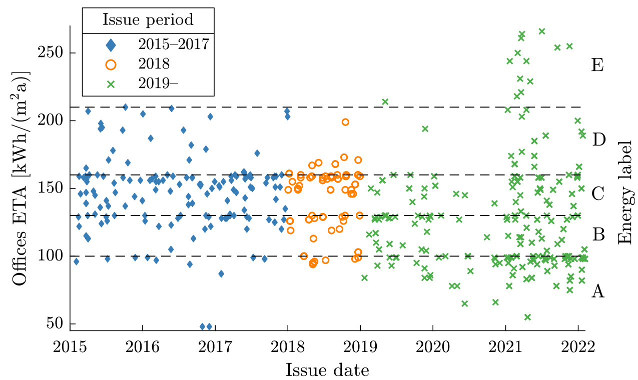


FIGURE 6. ETA certificates for office buildings K6 versus date of issue: 2015–2017 (blue diamonds), 2018 (orange circles), 2019–2022 (green crosses). The dashed lines denote separation ranges with respect to the energy label.

with others up to H being quite equally represented, the year 2018 very clearly reflects the stricter requirements, with a flattening of the counts after class E and a radical increment of class A certificates (from ∼3.5% to 12%). Undoubtedly, this result mirrors a sector that is more strictly controlled than the other two sectors. Surprisingly, while A was the largest class, with B being a close runner up, high-consumption buildings above the D class reappeared with non-negligible percentages.

B. EVALUATION OF THE IMPACT OF INCENTIVES PRE/POST-RENOVATIONS

Continuing the discussion from the previous section, we now seek to verify whether there was an accumulation of ETA (i.e., calculated) certificates towards lower values after 2018, since in that year, the energy label requirements became stricter for all building clusters. Examining dwellings K4, apartments K5 and office buildings K6, we obtained the plots shown in Figs. 4, 5 and 6. The K4 cluster features ETAs starting from 0 kWh/(m^{2}a).

One can observe only a slight accumulation of values for dwellings and apartments, while office buildings seem to maintain a uniformly wide range throughout. This pattern is confirmed by the standard deviations, which hold respectively for 2013–2017, 2018, and 2019–2022 as 39.39, 46.39, 23.82 for K4, 19.8, 16.86, 41.37 for K5 and 29.63, 24.45, 27.27 for K6. Remarkably, this visualisation shows the grouping of ETA certificates into classes A, B and C very clearly (the 50 kWh/(m^{2}a) level discussed in Section II-A is quite evident in Fig. 4 as well).

Figure 7 illustrates the ETA/KEK values for newly built or renovated kindergartens K1. During the entire year 2018, the minimum requirement for renovated kindergartens was class B, i.e., ETA/KEK $\leq$ 120 kWh/(m^{2}a), becoming class A, i.e., ETA/KEK $\leq$ 120 kWh/(m^{2}a) after 01.01.2019 (Table 1). In Fig. 7, the ETAs (calculated) tend to accumulate below 200 kWh/(m^{2}a), yet this value is rather large and corresponds to class D (Table 1). Therefore, those simulated values do not seem to have been underestimated for complying with the normative. On the one hand, a high scattering of both ETA and measured values (KEK) is observed, meaning that those

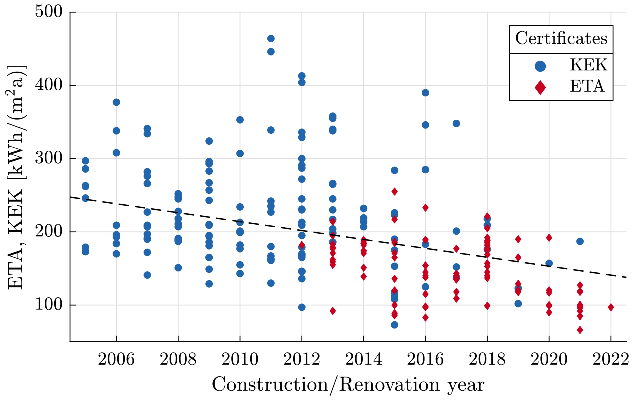


FIGURE 7. EPC values plotted by certificate year (red diamonds for ETA, blue dots for KEK) for recently constructed or renovated kindergartens K1. The fitting line (dashed) is computed for EPCs that were issued after 2000.

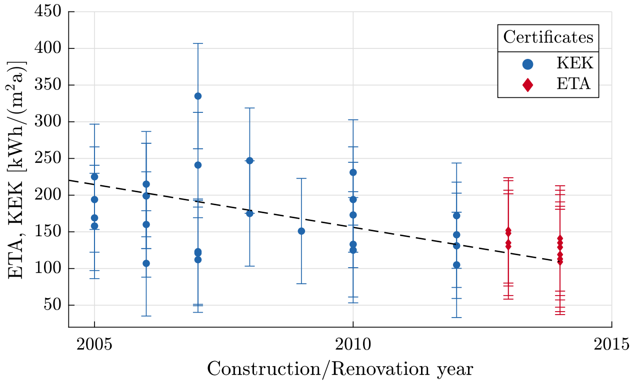


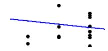
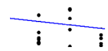
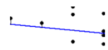
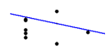
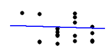
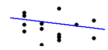
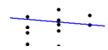
FIGURE 8. EPC values plotted by certificate year (red diamonds represent ETA, and blue dots represent KEK) for recently constructed or renovated schools K2. The fitting line (dashed) is computed for EPCs that were issued after 2000. The error bars correspond to a standard deviation of 71.77.

interventions were not completely successful. On the other hand, a fit performed over both ETAs and KEKs recognises a linear pattern. It accordingly seems that new construction or renovations of kindergartens overall managed to reduce energy consumption, despite a high degree of variance.

Similarly, the fit in Fig. 8 for other types of recently built or renovated schools, cluster K2 (including elementary and secondary), shows a more consistent linear pattern with less variance; constructions and renovations seem to have been moderately improving with time, regardless of a few outliers. This finding is encouraging, since extrapolating the fit to more recent years clearly shows that class A can be reached.

Table 3 compares the above cases with the time trend of ETA/KEK for several other building typologies. Specifically, adding research centres and university buildings to form the educational (all) cluster K3 reduces the renovation success that was accomplished with K1 and K2. One can also see that dwellings K4 do not exhibit any improvement with time, but this is not unexpected, as they should be subdivided into three subgroups. The same holds for the

TABLE 3. Time trend of EPCs for buildings that were built or renovated after 2000, including a forecast for the year when the ZEB status should be reached.

Type	Trend	Slope	p-value	ZEB year
K1		-7.408	1.158×10^{-14}	2039
K2		-11.667	6.476×10^{-06}	2023
K3		-7.4659	2.2×10^{-16}	2036
K4		-0.5567	0.04302	2269
K5		-5.3305	$< 2.2 \times 10^{-16}$	2043
K6		-11.8959	$< 2.2 \times 10^{-16}$	2030
K7		-6.415	$< 2.2 \times 10^{-16}$	2027
K8		-3.525	0.01978	2069
K9		-8.834	9.386×10^{-7}	2037
K10		-14.377	$< 2.2 \times 10^{-16}$	2029
K11		-4.979	1.397×10^{-6}	2050

entertainment sector K8, which entails a grouping of very different buildings, from theatres to discos. The commercial and services sector K7 instead shows a promising trend that forecasts an early 2027 ZEB year. Finally, office buildings K6 exhibit a distinct improvement that is in line with that of schools K2 and welfare K10.

With only the exceptions of K4 and K8, which are structurally different from the other datasets as noted above, the p-values are very small (below 5%), providing some degree of reliability regardless of the substantial EPC variance that is observed in most clusters.

C. BENCHMARKING TABLES AND HISTOGRAMS FOR AGGREGATED EPC CLUSTERS

As discussed in Section II-B, the benchmarking tables are computed by distribution fitting [9]. In this section, we report and comment on a few example results, while the full set of benchmarking tables for various building typologies of the Estonian database is listed in Appendix.

A distribution fitting plot featuring diagnostic graphs as described in II-B is displayed in Fig. 9 for educational database K3 (namely, kindergartens K1 and schools K2 aggregated with university and research buildings). The EPC distribution shape agrees with most literature for comparable datasets (see, e.g., [9], where a gamma was also adopted).

A lognormal distribution provides the best fit for most building clusters from K5 to K11, consistent with [26], where it was recognised to be the best fitting distribution for office buildings in Korea. A bimodal distribution (i.e., a distribution with two distinct modes, or peaks) best fitted K8 and K9, with a trimodal distribution holding for dwellings K4 and apartment buildings K5 (both private sectors with no

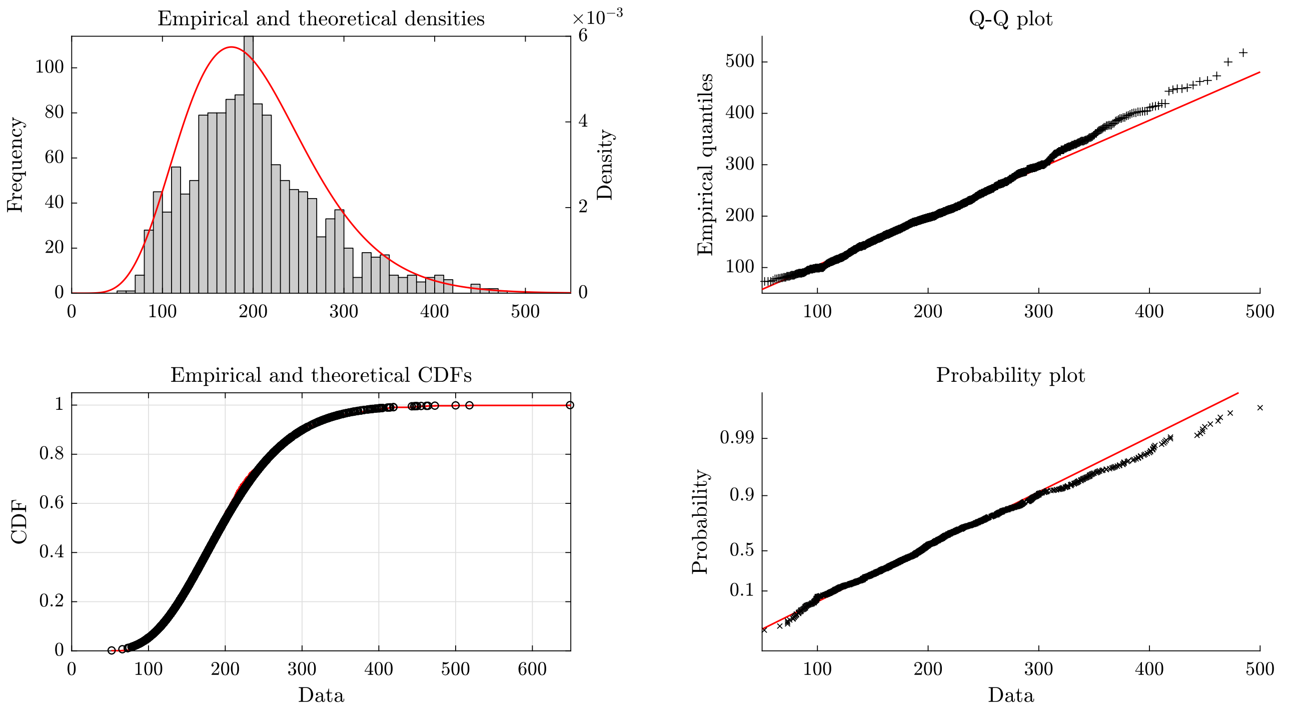


FIGURE 9. Histogram and gamma distribution fit, CDFs and diagnostic plots for the educational buildings cluster K3.

common policy). This finding is expected, since K4 is divided into three subgroups of detached houses according to heated area, which accounts for three modes and defines three distinct energy labels classes in the Estonian normative.

Generally, a lognormal is the continuous probability distribution of a random variable whose logarithm is normally distributed [31]. This distribution has wide applications to natural processes (such as the growth of organisms) and to human behaviour; it is also associated with entropy in energy processes [32]. A lognormal distribution is a reasonable match for the data at hand due to its asymmetrical descent to zero counts at the boundaries. Such asymmetry is explained by observing the accumulation of ETA/KEK towards smaller values, namely, classes B and C, yet with statistically subdominant class A certificates and a long tail of D class and higher EPCs (shown in Fig. 3).

The gamma distribution also provided a reasonably good fit in most cases, but the lognormal distribution had better log-likelihood and BIC values, although the difference could sometimes be as small as 0.2% or 1%. Using the gamma would have been in complete agreement with [9], and since the difference was rather minimal overall, we can confirm their result. A comparison of three different fits for apartment buildings is given in Fig. 10.

In Appendix, we report the benchmarking Tables 7, 8, 9, 10, 11, 12, 13, 14 and 15 for the selected building categories of the Estonian database that have been summarised in Table 2. As remarked above, the dwellings K4 were particularly intriguing, as they are split into three groups with different thresholds for the energy classes.

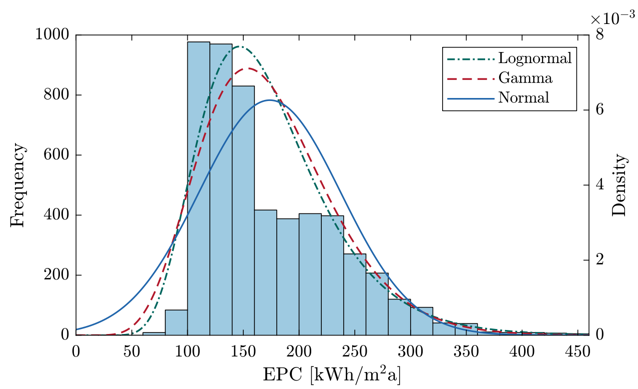


FIGURE 10. Comparison of three fitting distributions for apartment buildings K5: gamma, normal and lognormal.

Their ETA/KEK distribution is accordingly fitted nontrivially by a trimodal normal distribution (Fig. 11) and is also noticeable in Fig. 4. Although the fit in Fig. 11 is not ideal, it can very precisely capture the leftmost mode while well describing the area subtended by the other two peaks.

As a last check of the Estonian dataset as a whole, let us examine the distribution of ETA/KEK means and standard deviations for the 11 categories in Table 2. The "Means" set is left-skewed and was fitted with maximum likelihood by a Weibull distribution (shape=9.569, standard error=2.305; scale=215.43, standard error=7.136). The distribution of standard deviations SD was instead fitted by a right-skewed lognormal distribution (mean log 4.485 and sd log 0.4, log-likelihood −53.09), see Fig. 12.

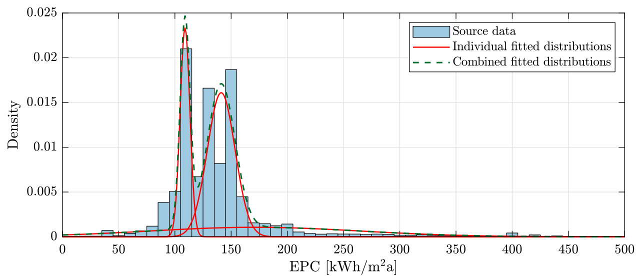


FIGURE 11. Distribution and trimodal fit for dwellings K4.

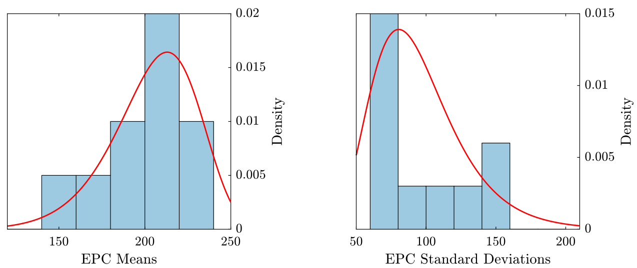


FIGURE 12. Distribution fitting for the means M (Weibull) and standard deviations SD (lognormal) from Table 2.

These results are not statistically relevant with such a paucity of data. They confirm, however, how the fitting method adopted in this study can be accurate even with a very small dataset. Furthermore, Fig. 12 shows that the EPC means of all those different building types tend to cluster between 200 and 240 kWh/(m^{2}a), while the standard deviations tend to be smaller than 75 kWh/(m^{2}a). This might signify that for all 11 building clusters, the ETA/KEK values are not very spread out, and the variance is kept under control. Indeed, each cluster shows only a very small percentage of outliers, usually not reaching 1% of the total data for that building type. With this large amount of data, a normal distribution would instead portray a larger data spread around the mean, unless the SD is very small (see Table 2). Overall, the means are reasonably centred in the lower-middle part of the distribution, and the presence of relatively large SDs accounts for the large variance that was observed in some categories. This is expected, as the dataset spans almost two decades, including several remodulations of the performance bounds, as explained in Section II-A.

D. CORRELATIONS WITH AGE AND HEATED AREA

To answer the question "How are building age and heated area correlated with energy consumption?", we computed correlations between EPC value and building construction or renovation year, as well as heated area, for all clusters. The results are summarised in Table 4, remarkably illustrating that there is no correlation with area for any cluster; this finding is in contrast with some literature [3], [26], although the comparison is not exact since we considered only the *heated area*, while other works addressed the total or *gross floor area* (GFA). Neither the construction nor renovation years show a correlation with ETA/KEK, meaning that an older building is not automatically less energy efficient. This is particularly true for the entertainment sector, which includes large and old theatres and cinemas. For instance, a theatre built in 1879 with a heated area of 2781.4 m^2 has a measured KEK of 187 kWh/(m^{2}a), which is not large.

TABLE 4. Pearson correlation table for all building clusters: ETA/KEK vs construction or renovation year and ETA/KEK vs heated area.

ID	ETA/KEK vs year	ETA/KEK vs area
K1 Kindergartens	−0.301	−0.266
K2 Schools	−0.259	−0.378
K3 Educational (all)	−0.257	−0.281
K4 Dwellings	−0.038	−0.019
K5 Apartments	−0.293	−0.131
K6 Office buildings	−0.293	−0.138
K7 Commercial	−0.096	−0.005
K8 Entertainment	−0.039	0.041
K9 Sports halls	−0.206	−0.022
K10 Welfare buildings	−0.399	−0.044
K11 Hotels, Dormitories	−0.393	0.020

Note how the data granularity of dwellings K4 is also reflected in this context, with very low correlation values. Only the educational sector seems to exhibit a marginal impact of heated areas. Correlations with construction or renovation year are naturally more sizeable overall but are not as high as expected.

E. READINESS OF THE ESTONIAN BUILDING STOCK

Summarising the results of the previous sections, we can now attempt to forecast when the Estonian buildings available in the database will be ready to comply with the European regulations in terms of energy efficiency. In other words, we now investigate when the ETA (or KEK) values of new or recently renovated buildings will reach the 0 kWh/(m^{2}a) and A class levels by 2050.

To answer this question, the linear interpolation model presented in Fig. 7 was extrapolated to the future, with the result shown in Fig. 13. It can be seen that, according to a fit that accounts only for certificates issued since 2000, kindergartens will likely reach the ZEB status in 2039. Estimates for all 11 building clusters are reported in Table 3.

Finally, since our database completely covers the years 2020 and 2021, we investigated whether the effect of the COVID-19 pandemic on energy consumption is reflected in the ZEB year estimations. A comparison of the ZEB year forecast by using pre- and post-COVID data is given in Table 5.

IV. DISCUSSION AND LESSONS LEARNT

Evaluating the readiness level of the Estonian building stock is the first immediate application of the analysis performed in this study. Table 3 illustrates that if the renovation and construction trends of the last 20 years continue, 73% of the examined building typologies will fulfil the ZEB

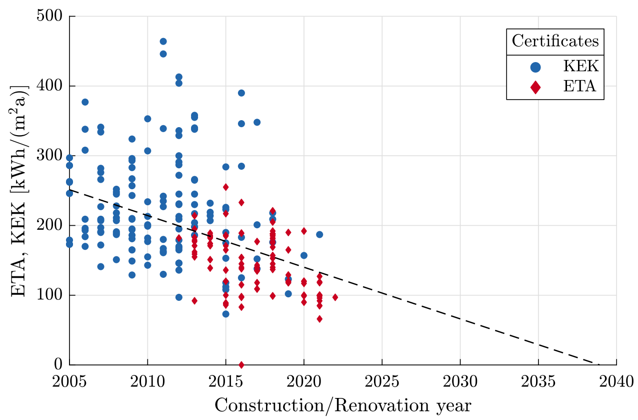


FIGURE 13. Extrapolation of ETA (red diamonds) and KEK (blue circles) for recently constructed or renovated kindergartens K1 until 2050, indicating 2039 as the year when the ZEB status should be reached.

TABLE 5. Effect of the COVID-19 pandemic on the ZEB year estimation: data until January 2020 for pre-COVID, data until March 2022 for post-COVID.

ID	Pre-COVID	Post-COVID	Diff.
K1 Kindergartens	2045	2039	−6
K2 Schools	2024	2023	−1
K3 Educational (all)	2038	2036	−2
K4 Dwellings	2056	2269	+213
K5 Apartments	2042	2043	+1
K6 Office buildings	2031	2031	+0
K7 Commerce, services	2027	2028	+1
K8 Entertainment	2074	2069	−5
K9 Sports halls	2032	2038	+6
K10 Welfare buildings	2028	2029	+1
K11 Hotels, Dormitories	2050	2050	+0

requirements before 2050. Moreover, hotels and dormitories K11 will miss the deadline by only three years. Note also that the trend for dwellings K4 is not relevant, as dwellings are subcategorised into three groups with different energy requirements.

From the above, one can conclude that the improvement of the Estonian building stock energy performance is sufficient and encouraging. Schools, office buildings, sports halls and welfare buildings should achieve the ZEB status within the next fifteen years; more diverse typologies, such as apartment buildings, are likely to reach this status well before 2050.

The distributions in Fig. 3 shed additional light in this respect from a different viewpoint. Publicly funded constructions (schools/kindergartens), which are rarely addressed in a quantifiable way, have targeted energy performance below minimum requirements. This might be the reason why schools K2 show a remarkably early 2023 as a ZEB year.

Privately funded residential (apartment buildings/detached houses) and nonresidential buildings (offices), together with hotels and dormitories, exhibit a more scattered pattern, which determines longer ZEB status times. This situation is probably due to a lack of the overall coordination that is peculiar to the public sector's management.

Although the correlations with year and heated area are weak, Table 4 confirms these considerations from yet another perspective. Focusing on the ETA/KEK vs year column, one can see that dwellings K4 and entertainment K8 indeed have the lowest correlation. Seeking some additional correlation between the two sets of ZEB year (Table 3) and EPC correlation with year (Table 4) returns a value of 0.65. Although not impressive, this value nevertheless indicates some underlying pattern and confirms an expected result.

Conversely, the Pearson correlation for the standard deviation (namely, the EPC spread) in Table 2 versus the ZEB year is -0.32. This value is significantly smaller, and reasonably so: rather than the scattering of EPC status, the time trend of the mean is significantly more relevant. This is all the more evident when looking, e.g., at welfare buildings K10, which are forecast to be ZEB in 2029 due to a high −14.377 slope value, whilst exhibiting one of the largest SDs of the entire dataset. The slope value, in fact, correlates with the ZEB year, with a value of 0.68. While relevant, the whole picture is therefore slightly more complex and should take into account a combination of factors.

Figures 5 to 8, together with the estimation in Fig.13, show that the share of new buildings with targeted energy performance levels that are below minimum requirements increases over time if better energy performance certificate classes have been defined.

The benchmarking method proposed in Sections II-B and III-C has an evident applicative value. Once the EPC of a building, namely, its yearly consumption per square metre, is known, comparing its energy performance with the respective building category is very straightforward. As an example, if an Estonian office building records an ETA or KEK equal to 130 kWh/(m^2a), it receives a score of 75 points based on Table 10. This means that it is placed within the cumulative 25% of the office building stock K6, which is a solid result.

First, this means that as the procedure is based on certificates' distributions, a quantitative description of the national EPC database can be accomplished easily and in detail, including all different building typologies. The above has illustrated that reaching useful conclusions about the performance of diverse clusters is straightforward.

Second, the R code used for fitting distributions and constructing benchmarking tables is computationally cheap and can be easily implemented in spreadsheet-like tools. These tools can be used by municipality managers and employees for a variety of purposes. One immediate application is in the selection of underperforming buildings for detailed energy audits [33]. Such considerations have general validity, as any analogous code written in a different language (e.g., Python) is expected to be equally applicable with low computational demand.

As a general remark, Fig. 1 suggests that one cannot find a distinction in the number of certificates issued before requirements became stricter. The number of ETA or KEK increased progressively after 2013, when the energy labelling

TABLE 6. Relative difference of pre- and post-COVID ZEB year estimation, broken down into ETA and KEK: ">" or "≫" for increment, "<" or "≪" for decrement, "=" if unchanged.

ID	ETA (regulations)	KEK (COVID-19)
K1 Kindergartens	<	<
K2 Schools	=	=
K3 Educational (all)	<	=
K4 Dwellings	≫	>
K5 Apartments	<	>
K6 Office buildings	<	<
K7 Commerce, services	≪	=
K8 Entertainment	≪	>
K9 Sports halls	≫	=
K10 Welfare buildings	>	<
K11 Hotels, Dormitories	≪	=

was remodulated with stricter requirements; the largest counts can be found in 2016 and 2017, which are at the end of the 2013–2017 period. The number of certificates per year since 2018 also seems to be quite constant, regardless of the introduction of stricter bounds.

Finally, regarding the effect of the COVID-19 pandemic, the results in Table 5 show that in general, large buildings with a steady base energy use, such as shopping centres and hotels, did not show a remarkable effect in terms of occupancy-induced energy consumption (mostly plug loads but also HVAC). On the other hand, the educational sector and private houses K4 seem to have been largely affected by the abrupt change in working habits. The entertainment sector K8 (theatres, clubs, discos, etc.) also displays a sizable reduction in energy use. However, sports halls K9 show the opposite, which is counterintuitive.

Some deeper analysis is indeed required: let us recall that Table 5 reports estimates computed by fitting the ETA and KEK certificates aggregated together, as EPCs. If ETAs (calculated) and KEKs (measured consumption) are instead fitted separately, one obtains a clearer picture. The breakdown reveals a critical role of regulations: as illustrated in Table 6, the KEK or metered consumption did not influence the result as critically as the calculated consumption ETA. Fitting only according to ETA provided differences of up to one hundred years.

As the simulation parameters are set by the country legislation, the ETAs directly result from the changes in such regulations. This explains the apparently counterintuitive estimate for sports halls K9 and perfectly reflects the result for dwellings D4. After 2018, class A must be reached only by the largest $A > 220\ \mathrm{m}^2$ houses, while those with $A < 220\ \mathrm{m}^2$ (i.e., 68% of the total) must only meet class B requirements.

One can accordingly infer that the COVID-19 emergency has not critically affected the Estonian building stock; rather, the role of governmental policies is largely dominant. This additional finding hence constitutes more evidence that to accomplish the ZEB and carbon neutrality goals of the European Green Deal, particular attention should be devoted to the remodulation of national renovation campaigns.

V. CONCLUSION

Energy-efficient districts are a key objective of municipalities and national agencies, as well as the European Union and other transnational agencies. Benchmarking is a procedure that allows the quantification of the energy efficiency of the present building stock.

In the present study, we have approached an Estonian EPC database consisting of approximately 35 000 buildings, which was subdivided into 11 clusters according to typology and energy label thresholds. A statistical analysis based on data mining has unveiled some distinctive features of the different building clusters and of the general database as a whole.

The readiness level of the Estonian building stock is estimated to be very good. Newly constructed or renovated buildings for most of the 11 categories will accomplish the ZEB status before 2050 if current trends continue. In particular, secondary schools might reach it even in 2023, while commercial and services should do so in 2027.

Lacking overall coordination in privately funded buildings determines a larger spread of EPCs; this ultimately pushes the ZEB accomplishment to occur later, if compared to public buildings. Stricter energy label requirements have positively affected some selected building clusters, with a gradual increase in classes A and B over time, although not as strongly or uniformly as one might expect. Additionally, the incentives pre/post-renovations proved to have been moderately successful, as the energy consumption was reduced overall, despite a high degree of variance. In general, our procedure allows a building to be identified according to renovation incentives. Our analysis of the COVID-19 period 2020–2022 provided a perhaps unexpected result that is directly related to the above: it was found that the pandemic affected the building stock only marginally. Conversely, the national regulations, which directly affect the ETAs (or calculated EPCs), have been the main driving factor in improving the readiness of some building typologies, typically the public sector, as well as in worsening that of others (mostly the private sector).

Fitting the EPC distribution for each cluster provided a benchmarking table that allows the comparison of a given building with the existing stock. This procedure can be automated and implemented into spreadsheet tools aimed at selection for digital audit or energy performance assessments. Finally, in contrast to earlier investigations, we found no relevant correlation between energy consumption per unit surface and building heated area.

Our study can be improved in several respects, first by examining the large cluster of detached houses (dwellings K4), which the Estonian energy labelling subdivides into three groups according to heated area. Further improvements could include the implementation of EPC time trends and benchmarking tables into computational tools for municipalities and energy agencies or by more statistically refined methods to determine the ZEB year with higher precision.

TABLE 7. Benchmarking table for educational buildings (gamma).

Score	Cumul. %	EPC $\geq$	EPC $<$
100	0	0	83.74
99	1	83.74	92.28
98	2	92.28	106.74
95	5	106.74	121.48
90	10	121.48	132.56
85	15	132.56	142.08
80	20	142.08	150.79
75	25	150.79	159.07
70	30	159.07	175.18
60	40	175.18	191.72
50	50	191.72	209.82
40	60	209.82	231.08
30	70	231.08	258.71
20	80	258.71	302.58
10	90	302.58	∞

TABLE 8. Benchmarking table for dwellings (trimodal).

Score	Cumul. %	EPC $\geq$	EPC $<$
100	0	0	36.85
99	1	36.85	70.54
98	2	70.54	95.08
95	5	95.08	105.09
90	10	105.09	112.45
85	15	112.45	116.11
80	20	116.11	118.23
75	25	118.23	119.63
70	30	119.63	129.75
60	40	129.75	137.79
50	50	137.79	142.74
40	60	142.74	154.27
30	70	154.27	159.31
20	80	159.31	174.66
10	90	174.66	∞

TABLE 9. Benchmarking table for apartment buildings (trimodal).

Score	Cumul. %	EPC $\geq$	EPC $<$
100	0	0	96.49
99	1	96.49	102.41
98	2	102.41	104.94
95	5	104.94	112.37
90	10	112.37	117.93
85	15	117.93	119.32
80	20	119.32	121.15
75	25	121.15	126.24
70	30	126.24	140.95
60	40	140.95	149.65
50	50	149.65	172.9
40	60	172.9	197.73
30	70	197.73	225.27
20	80	225.27	258.09
10	90	258.09	∞

Several other future work perspectives are possible, such as studying the impact of specific incentives and clustering buildings into subgroups for more meaningful benchmarking, namely, measuring individual buildings against those that should be comparably more energy efficient. Furthermore, new or corrected EPC class values can be obtained by virtue of the distributions computed here, e.g., according to the ISO 52003-1 standard [34] and the new European directive [2]. In other words, the study at hand also constitutes the groundwork for drawing suggestions/recommendations towards a new EPC class scaling.

TABLE 10. Benchmarking table for office buildings (lognormal).

Score	Cumul. %	EPC $\geq$	EPC $<$
100	0	0	63.33
99	1	63.33	71.50
98	2	71.50	85.78
95	5	85.78	100.85
90	10	100.85	112.48
85	15	112.48	122.67
80	20	122.67	132.15
75	25	132.15	141.29
70	30	141.29	159.42
60	40	159.42	178.46
50	50	178.46	199.77
40	60	199.77	225.40
30	70	225.40	259.61
20	80	259.61	315.79
10	90	315.79	∞

TABLE 11. Benchmarking table for commercial and service buildings (lognormal).

Score	Cumul. %	EPC $\geq$	EPC $<$
100	0	0	104.26
99	1	104.26	111.74
98	2	111.74	123.96
95	5	123.96	135.94
90	10	135.94	144.67
85	15	144.67	152.01
80	20	152.01	158.60
75	25	158.60	164.76
70	30	164.76	176.51
60	40	176.51	188.23
50	50	188.23	200.74
40	60	200.74	215.05
30	70	215.05	233.09
20	80	233.09	260.64
10	90	260.64	∞

TABLE 12. Benchmarking table for the entertainment sector (bimodal).

Score	Cumul. %	EPC $\geq$	EPC $<$
100	0	0	59.74
99	1	59.74	74.92
98	2	74.92	97.76
95	5	97.76	118.17
90	10	118.17	132.02
85	15	132.02	143.1
80	20	143.1	152.67
75	25	152.67	161.32
70	30	161.32	177.15
60	40	177.15	192.29
50	50	192.29	207.94
40	60	207.94	225.66
30	70	225.66	248.93
20	80	248.93	295.87
10	90	295.87	∞

APPENDIX

VI. BENCHMARKING TABLES FOR SELECTED CATEGORIES

This section displays the benchmarking tables for all of the examined building clusters, computed as per Section II-B and corresponding to the analysis reported in Section III-C. The EPC values are given in kWh/(m^{2}a).

TABLE 13. Benchmarking table for sports buildings (bimodal).

Score	Cumul. %	EPC ≥	EPC <
100	0	0	68.54
99	1	68.54	82.63
98	2	82.63	103.9
95	5	103.9	123.01
90	10	123.01	136.07
85	15	136.07	146.58
80	20	146.58	155.72
75	25	155.72	164.05
70	30	164.05	179.53
60	40	179.53	194.73
50	50	194.73	211.17
40	60	211.17	231.53
30	70	231.53	266.57
20	80	266.57	364.15
10	90	364.15	∞

TABLE 14. Benchmarking table for the welfare sector (lognormal).

Score	Cumul. %	EPC ≥	EPC <
100	0	0	61.92
99	1	61.92	70.74
98	2	70.74	86.39
95	5	86.39	103.17
90	10	103.17	116.29
85	15	116.29	127.91
80	20	127.91	138.79
75	25	138.79	149.36
70	30	149.36	170.51
60	40	170.51	192.98
50	50	192.98	218.41
40	60	218.41	249.34
30	70	249.34	291.15
20	80	291.15	360.98
10	90	360.98	∞

TABLE 15. Benchmarking table for hotels and dormitories (lognormal).

Score	Cumul. %	EPC ≥	EPC <
100	0	0	97.69
99	1	97.69	105.19
98	2	105.19	117.53
95	5	117.53	129.71
90	10	129.71	138.63
85	15	138.63	146.16
80	20	146.16	152.94
75	25	152.94	159.29
70	30	159.29	171.45
60	40	171.45	183.65
50	50	183.65	196.72
40	60	196.72	211.73
30	70	211.73	230.76
20	80	230.76	260.02
10	90	260.02	∞

ACKNOWLEDGMENT
The authors would like to thank Kalle Kuusk for providing data from the Estonian Building Stock Database.

REFERENCES
[1] J.-M. Glachant, P. Linares, A. Loeschel, O. Edenhofer, S. Tagliapietra, and G. Zachmann, "The European union energy transition: Key priorities for the next five years," *Eur. Energy Climate J.*, vol. 9, no. 1, pp. 5–14, Jan. 2020.

[2] A. Wilson, "Revision of the energy performance of buildings directive: Fit for 55 package," Eur. Parliamentary Res. Service, Bruxelles, Belgium, Briefing, Tech. Rep., Feb. 2022.

[3] J. Roth, B. Lim, R. K. Jain, 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, 2020, Art. no. 111327.

[4] 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 Proc.*, vol. 101, pp. 169–175, Nov. 2016.

[5] V. Corrado, I. Ballarini, S. Paduos, and L. Tulipano, "A new procedure of energy audit and cost analysis for the transformation of a school into a nearly zero-energy building," *Energy Proc.*, vol. 140, pp. 325–338, Dec. 2017.

[6] A. Kubule, K. Ločmelis, and D. Blumberga, "Analysis of the results of national energy audit program in Latvia," *Energy*, vol. 202, Jul. 2020, Art. no. 117679.

[7] O. Pasichnyi, J. Wallin, F. Levihn, H. Shahrokni, and O. Kordas, "Energy performance certificates—New opportunities for data-enabled urban energy policy instruments?" *Energy Policy*, vol. 127, pp. 486–499, Apr. 2019.

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

[9] H. G. Kim and S. S. Kim, "Development of energy benchmarks for office buildings using the national energy consumption database," *Energies*, vol. 13, no. 4, p. 950, 2020.

[10] J. Liu, J. Wang, G. Li, H. Chen, L. Shen, and L. Xing, "Evaluation of the energy performance of variable refrigerant flow systems using dynamic energy benchmarks based on data mining techniques," *Appl. Energy*, vol. 208, pp. 522–539, Dec. 2017.

[11] D. Syed, H. Abu-Rub, A. Ghrayeb, and S. S. Refaat, "Household-level energy forecasting in smart buildings using a novel hybrid deep learning model," *IEEE Access*, vol. 9, pp. 33498–33511, 2021.

[12] M. H. Mohammadi and K. Saleh, "Synthetic benchmarks for power systems," *IEEE Access*, vol. 9, pp. 162706–162730, 2021.

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

[14] S.-M. Hong, G. Paterson, E. Burman, P. Steadman, and D. Mumovic, "A comparative study of benchmarking approaches for non-domestic buildings: Part 1—Top-down approach," *Int. J. Sustain. Built Environ.*, vol. 2, no. 2, pp. 119–130, 2013.

[15] D. Calì, T. Osterhage, R. Streblow, and D. Müller, "Energy performance gap in refurbished German dwellings: Lesson learned from a field test," *Energy Buildings*, vol. 127, pp. 1146–1158, 2016.

[16] E. Delzendeh, S. Wu, A. Lee, and Y. Zhou, "The impact of occupants' behaviours on building energy analysis: A research review," *Renew. Sustain. Energy Rev.*, vol. 80, pp. 1061–1071, Dec. 2017.

[17] M. K. Annaqeeb, R. Markovic, V. Novakovic, and E. Azar, "Non-intrusive data monitoring and analysis of occupant energy-use behaviors in shared office spaces," *IEEE Access*, vol. 8, pp. 141246–141257, 2020.

[18] S. Gilani, W. O'Brien, and H. B. Gunay, "Simulating occupants' impact on building energy performance at different spatial scales," *Building Environ.*, vol. 132, pp. 327–337, Mar. 2018.

[19] K. Ahmed, M. Carlier, C. Feldmann, and J. Kurnitski, "A new method for contrasting energy performance and near-zero energy building requirements in different climates and countries," *Energies*, vol. 11, no. 6, p. 1334, 2018.

[20] J. Zhao, B. Lasternas, K. P. Lam, R. Yun, and V. Loftness, "Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining," *Energy Buildings*, vol. 82, pp. 341–355, Oct. 2014.

[21] I. E. Bennet and W. O'Brien, "Office building plug and light loads: Comparison of a multi-tenant office tower to conventional assumptions," *Energy Buildings*, vol. 153, pp. 461–475, Oct. 2017.

[22] A. Ferrantelli, H. Kuivjõgi, J. Kurnitski, and M. Thalfeldt, "Office building tenants' electricity use model for building performance simulations," *Energies*, vol. 13, no. 21, p. 5541, 2020.

[23] T. Cholewa, A. Siuta-Olcha, and R. Anasiewicz, "On the possibilities to increase energy efficiency of domestic hot water preparation systems in existing buildings—Long term field research," *J. Cleaner Prod.*, vol. 217, pp. 194–203, Apr. 2019.

[24] A. Ferrantelli, K. Ahmed, P. Pylsy, and J. Kurnitski, "Analytical modelling and prediction formulas for domestic hot water consumption in residential Finnish apartments," *Energy Buildings*, vol. 143, pp. 53–60, May 2017.

[25] Q. Zhou, M. Wang, J. Ni, Y. Ye, and W. Zuo, "DdBB: A data-driven framework for building energy use benchmarking," in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Aug. 2018, pp. 716–721.
[26] H. S. Park, M. Lee, H. Kang, T. Hong, and J. Jeong, "Development of a new energy benchmark for improving the operational rating system of office buildings using various data-mining techniques," *Appl. Energy*, vol. 173, pp. 225–237, Jul. 2016.
[27] R Core Team, R Foundation for Statistical Computing, Vienna, Austria. (2020). *R: A Language and Environment for Statistical Computing*. [Online]. Available: https://www.R-project.org/
[28] M. L. Delignette-Müller and C. Dutang, "fitdistrplus: An R package for fitting distributions," *J. Statist. Softw.*, vol. 64, no. 4, pp. 1–34, 2015.
[29] A. C. Cullen and H. C. Frey, *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. New York, NY, USA: Plenum, 1999.
[30] P. Macdonald and P. Green, *User's Guide to Program MIX: An Interactive Program for Fitting Mixtures of Distributions*. Hamilton, ON, Canada: Ichthus Data Systems, 1988.
[31] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1. New York, NY, USA: Wiley, 1994.
[32] T. Lee, "Lognormality in turbulence energy spectra," *Entropy*, vol. 22, no. 6, p. 669, Jun. 2020.
[33] M. S. Geraldi and E. Ghisi, "Building-level and stock-level in contrast: A literature review of the energy performance of buildings during the operational stage," *Energy Buildings*, vol. 211, Mar. 2020, Art. no. 109810.
[34] *Energy Performance of Buildings—Indicators, Requirements, Ratings and Certificates—Part 1: General Aspects and Application to the Overall Energy Performance*, Standard ISO 52003-1:2017, International Organization for Standardization, 2017.

ANDREA FERRANTELLI received the M.Sc. degree in theoretical physics from the University of Turin, Italy, in 2002, and the Ph.D. degree in theoretical particle cosmology from the University of Helsinki, Finland, in January 2010. In 2011, he joined the Civil Engineering Department, Aalto University, Finland, as a Postdoctoral Fellow in building physics. Since 2017, he has been a Postdoctoral Researcher with the Department of Civil Engineering and Architecture, Tallinn University of Technology, Estonia. His research interests include theoretical models for energy-efficient buildings, including applications of heat transfer, HVAC, thermal comfort and indoor air quality in buildings, statistical modeling of energy consumption and occupancy, thermomechanical phenomena in soft matter, supersymmetric dark matter, and nuclear medicine.

JURI BELIKOV (Senior Member, IEEE) received the B.Sc. degree in mathematics from Tallinn University and the M.Sc. and Ph.D. degrees in computer and systems engineering from the Tallinn University of Technology, in 2006, 2008, and 2012, respectively. In 2015 and 2017, he was a Postdoctoral Fellow with the Faculty of Mechanical Engineering and the Andrew & Erna Viterbi Faculty of Electrical Engineering, Technion—Israel Institute of Technology. He is currently an Assistant Professor with the Department of Software Science, Tallinn University of Technology. His research interests include the edge of nonlinear control theory and power systems.

EDUARD PETLENKOV received the B.Sc., M.Sc., and Ph.D. degrees in computer and systems engineering from the Tallinn University of Technology. He is currently a tenured Full Professor with the Department of Computer Systems, Tallinn University of Technology, and the Head of the Centre for Intelligent Systems. His research interests include the domain of nonlinear control, system analysis, and computational intelligence.

MARTIN THALFELDT received the M.Sc. degree in environmental engineering, heating and ventilation, in 2009, and the Ph.D. degree in civil and environmental engineering from the Tallinn University of Technology, Estonia, in 2016. He is currently a tenured Assistant Professor with the Tallinn University of Technology. From 2016 to 2018, he was a Postdoctoral Researcher with the Norwegian University of Science and Technology, Trondheim, Norway. His research interests include the energy performance and indoor climate of buildings and the design of building service systems. His awards and honours include the REHVA Young Researcher Award, in 2019. He was elected to become a member of the Estonian Academy of Young Sciences, in 2022.

JAREK KURNITSKI is currently a Professor with the Tallinn University of Technology, Estonia, and Aalto University, Finland. He is also the Chair of REHVA Technology and Research Committee and is currently leading the REHVA COVID-19 Task Force preparing European HVAC and ventilation guidance for the epidemic situation. He is also the Leader of the Estonian Center of Excellence in Research ZEBE, Zero Energy and Resource Efficient Smart Buildings and Districts. He is internationally known for the preparation of technical definitions for nearly zero-energy buildings through many activities with the REHVA Technology and Research Committee and for contributions to European standards. He has been deeply involved in the work to improve the energy efficiency and indoor climate of the built environment in Estonia and Finland, with major contributions in the development of dynamic simulation-based energy calculation frameworks for present energy performance regulations. He has been recognized with scientific awards from the professional associations SCANVAC, in 2000, REHVA, in 2005, and the Estonian State Research Award, in 2018. His list of publications includes more than 500 items.