Energy & Buildings 238 (2021) 110786

Contents lists available at ScienceDirect

Energy & Buildings

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

Top-down spatially-explicit probabilistic estimation of building energy performance at a scale

Ruslan Zhuravchak [a,*], Raquel Alonso Pedrero [b], Pedro Crespo del Granado [b], Natasa Nord [a], Helge Brattebø [a]

[a] Department of Energy and Process Engineering, NTNU. Kolbjørn Hejes v. 1B, NO-7491 Trondheim, Norway

[b] Department of Industrial Economics and Technology Management, NTNU. Alfred Getz v. 3, NO-7491 Trondheim, Norway

ARTICLE INFO

Article history:
Received 7 August 2020
Revised 14 December 2020
Accepted 23 January 2021
Available online 4 February 2021

Keywords:
Built stock
Energy performance
Geographical information systems (GIS)
Urban energy planning
Top-down modelling
Probabilistic sampling
Parametric density estimation
Frequentist inference

ABSTRACT

Achieving the energy-related and environmental targets for nations and municipalities is largely dependent on the existing built stock. It plays a pivotal role in the accomplishment of these targets through the implementation of energy efficiency and flexibility programs, involving the deployment of distributed energy resource management technologies, refurbishment of building envelopes and upgrading of indoor environmental control equipment. Spatial awareness about urban energy use enables to prioritise the areas where these solutions will be most effective and balanced with the plans for new constructions. Large-scale building energy mapping, however, must cope with heterogeneity of buildings within the built stock, absence of detailed information and multiple sources of uncertainty that stem from the complex and dynamic properties of the phenomenon at a building level. One of the key challenges in the discipline is to account for these uncertainties while maintaining the rational model complexities and data needs. This study, therefore, suggests a parsimonious top-down probabilistic modelling recipe to enable geospatial energy mapping and analysis. Under such modelling principles, an inverse propagation of uncertainties is carried out from the status quo of the built stock. The proposed framework is based on probabilistic sampling with prior parametric univariate density estimation and statistical hypothesis testing. Consolidation with the exogenous influencing factors is facilitated through the measure of statistically significant difference. This approach is exemplified with the data from two sources: the cadastral system and the energy performance certificates registry. A case study developed for Trondheim (Norway) quantified the central tendency and dispersion in the distributions of the simulated bulk total annual energy use by buildings per 1×1 km grid cell over the urban territory. The results suggest that best estimates of these values vary between 11 $MWh \cdot y^{-1}$ and 141 $GWh \cdot y^{-1}$ depending on the grid cell. A measure of dispersion in the simulated results is highly correlated with these estimates. Robust handling of uncertainties and the possibility to accommodate a variety of modelling objectives make this approach practical for energy mapping with a flexible spatial resolution that may facilitate numerous applications in energy planning. A collection of methods for univariate density estimation discussed in this study together with the empirical data are accessible through Built Stock Explorer:https://builtstockexplorer.indecol.ntnu.no. This open web application for knowledge discovery in building energy data enables to reproduce some of the results presented in the article.

1. Introduction

Built stock is perceived as holding a large potential for mitigating the environmental impacts directly or indirectly associated with its final energy use [1], which reached 128 EJ globally in 2019 [2]. Improving the energy performance of buildings, therefore, is being supported through regulatory mechanisms at various levels of governance. These mechanisms, usually initiated at a national or municipality levels, are targeting the solutions at districts or neighbourhoods [3] and focused primarily on well-reasoned infrastructural transformations [4], retrofitting and upgrading programs [5] and more sustainable energy management technologies [6].

An effective strategic energy planning of these and the related solutions relies on geospatial information in several ways. Spatial awareness enables to priortise the areas of high energy use, where the technical, economic, and environmental feasibility of relevant measures may be justified. Such solutions could lead to multiple benefits, e.g. decrease the total energy use and reduce the costs

* Corresponding author.
E-mail address: ruslan.zhuravchak@ntnu.no (R. Zhuravchak).

https://doi.org/10.1016/j.enbuild.2021.110786
0378-7788/

Nomenclature

CDF	Cumulative distribution function, page 12	r.v.	Random variable, page 7
i.i.d.	Independent identically distributed (sample), page 24	SD	Standard deviation, page 9
KS	Kolmogorov–Smirnov (test), page 14	SS	Sample size, page 14
MC	Monte-Carlo (simulation), page 7		
MLE	Maximum likelihood estimation, page 13		
PDF	Probability density function, page 5		

of energy by avoiding transmission losses, support the integration of non-dispatchable renewables, enhance the reliability and resilience of power grid through peak load smoothing and frequency control. Given the anticipated growth of the electric vehicles fleet, a more favourable deployment of charging stations can also be accommodated by spatially-informed energy planning [7]. Another important reason for mapping the energy bottlenecks is studying the contribution of buildings to the atmospheric heating – a phenomenon known as urban heat island [8,9].

The need for spatial analysis of urban energy use prompted numerous attempts to complement built stock energy information with geospatial data [10,11]. The availability of multiple studies indicates a widespread interest in developing the means to enable such analysis. Building level [12,13], block area [14], square grid cell [8,15,9] and local authority [16] are the most common choices of spatial or administrative resolution that such analysis is focused on.

Energy performance of buildings, however, is characterised as highly complex, complicated and dynamic phenomena, which is attributed to numerous factors of physical and occupancy-related origins [17]. At a larger scale, the associated uncertainties are amplified by the diversity of buildings, variations in their exposure to the outdoor conditions, ageing processes, use/maintenance practices and other. Determining the energy performance of buildings at the large scale given these explanatory factors is the subject of built stock energy modelling [18]. It is, amongst other important applications, an essential component of geospatial energy mapping. Following both, the original classification system proposed by Swan and Ugursal [19] and its recent revision [20], the prevailing practices for spatially-explicit built stock energy modelling correspond to the bottom-up approach. Bottom-up reasoning enables to infer energy use at the aggregated level based on the information available at a lower spatial level. Several studies suggested engineering-based ("white-box") models as the means for building energy mapping [15,9,12,21]. Other authors make use of "black-box" methods [22] instead [14,23,13]. As opposed to bottom-up methods, top-down approach implies spatial downscaling procedures from a broader aggregated scope to the city- district- or building-level. Studies [8,9,24–26] suggested top-down methods as suitable for examining spatial variations of energy use for numerous purposes.

Regardless of which, bottom-up or top-down, approach is used, the attempts are made to model the phenomenon that lacks either detailed and complete knowledge or order or pattern or coherence or a combination of these. Booth et al. [27], systematized the uncertainties behind bottom-up engineering-based built stock models by their origins: a) variability of energy use due to chance within the identical buildings (aleatory uncertainties); b) heterogeneity of buildings within groups or typologies; and c) epistemic uncertainties which accommodate lack of knowledge about the phenomenon, the choice of inadequate model parameters and/or the risk of obtaining a biased model. These are also applicable to "black-box" methods that seek to approximate the uncertainties and assume a likewise deterministic relationship between the variables.

Probabilistic modelling enables to account for uncertainties and to address the limitations of approximating them in a deterministic alternative [27]. At a building level, the available studies quantify the uncertainties probabilistically in either forward or inverse manner [28]. Forward uncertainty propagation is dominated by sampling methods, where the inputs of the model are intentionally varied to obtain the likely variations of model outputs. Amongst the built stock energy studies, forward propagation principles were used to account for the epistemic uncertainties [27,29,30]. Inverse uncertainty propagation methods aim to relate the observed empirical data to both known and unknown model parameters and/or built stock properties. Given the underlying statistical inference approach, Tian et al. [28] categorised the inverse uncertainty analysis practices in the discipline as either frequentist or Bayesian. Whereas the former consists of methods for operating solely on empirical data, the latter also accommodates the prior knowledge and beliefs to aid the inferential statistics. In the context of built stock energy modelling, inverse Bayesian-based inference is advocated in studies [16,31,32]. Built stock energy studies with frequentist inference were not found in the domain-specific literature.

If combined with the inverse uncertainty propagation principles, top-down modelling reveals numerous useful properties. Both, aleatory uncertainties and heterogeneity of buildings are reflected in the built stock energy data at the aggregated spatial level. A step-wise disaggregation of this data with exogenous factors naturally conveys the associated uncertainties in an inverse manner. An empirically inferred probability density function (PDF) is a proxy for central tendency and variability due to yet unexplained uncertainty at each of these steps. Consequently, every subsequent disaggregation may lead to better estimates of uncertainties and thus, higher modelling accuracy. If the rational disaggregation steps are reflected in the structure of the model, numerous advantages directly or indirectly stem from the following:

- Uncertainties can be quantified at each level of the step-wise disaggregation, which enables a modeller to make judgements about the quality of the model and to control the trade-off between the expected accuracy gains versus additional data feed;
- The levels of sensitivity to adding the exogenous factors can be quantified through statistical hypothesis testing. This can prevent from using the redundant or insignificant model inputs and thus, to address the overfitting;
- Data requirements to achieve the necessary level of modelling details can be calculated beforehand. This enables to set up and efficiently manage the data collection process.

Despite these substantial advantages, top-down probabilistic modelling remains poorly explored within the discipline. This article, therefore, is motivated by the need to elaborate on the workflow, methods, and procedures that such modelling may involve. It is shown that this modelling approach may facilitate spatially-explicit energy mapping with a flexible spatial resolution and the varying levels of details. Exemplification is made through the case study developed for Trondheim, Norway. Section 2 Method explains the non-parametric probabilistic model, data resources that the model relies on and a collection of methods which are illustrated using the available data. All these components are further synthesised into a coherent computational procedure used to obtain the estimates of bulk total annual energy use over 1×1 km geospatial grid. Section 3 Results provides both, intermediate and final outputs obtained through this procedure. Section 4 Discussion evaluates the strengths, weaknesses and potential further developments for such modelling. Extra care is taken to elaborate on the capabilities of top-down models to account for an increasingly detailed architectural and technical information needed for built stock energy research. And to discuss the role of statistical inference in further shaping the available domain knowledge. The Conclusions (Section 5) summarise the findings made in this study, reflect upon potential opportunities and barriers for further developments.

2. Method

The proposed procedural framework facilitates estimation of bulk (for all buildings) total (for all energy sources) annual energy use in geospatial zones. This can be achieved with a non-parametric model described in Section 2.1. Two data sources, namely the National Cadastral System (Section 2.2.1) and the Energy Performance Certificates (EPC) dataset (Section 2.2.2) provide the inputs into the model. Whereas accurate information on geospatial positioning, size and type of buildings and dwellings is available, the data on real energy performance are relatively scarce and available only at a city level. Therefore, the computational procedure involves Monte-Carlo (MC) simulation and considers the energy use intensity of buildings as a continuous random variable (r.v.). Inferring the properties of this r.v. from the available sample is the subject to density estimation procedure described in Section 2.3. Section 2.4 provides the description of a comprehensive computational procedure used to achieve the desired results – the estimates of central tendency and dispersion of bulk total annual energy use by buildings per geospatial zone.

The smallest element of built stock accounted for in the model (Section 2.1) is a building unit,[1] which enables to harmonise the data and to preserve the consistency across all steps of the study. Also, this allows to explicitly account for the energy performance of buildings that have a mixed use purpose, e.g. offices and apartments across multiple floors of the same building.

2.1. Probabilistic model

The simplest non-parametric model for estimating the bulk total annual energy use $E_{zone\ tot}$ of a geospatial zone consisting of $j \in [0, m]$ units can be defined as:

$$E_{zone\ tot} = \sum_{j=0}^{m} (a_j \cdot r_j) \qquad (kWh \cdot y^{-1}) \tag{1}$$

where:

a_j – heated floor area (m^2) of j^{th} unit;
r_j – energy use intensity $(kWh \cdot m^{-2} \cdot y^{-1})$ of j^{th} unit.

[1] The smallest element registered (in the cadastral system) or certified (in the EPC system): dwelling if residential use purpose; the section or the whole building otherwise.

The need to account for exogenous variables (disaggregation), in order to reflect the properties specific to the group of units, entails modifying the Eq. (1). A generalised form of the top-down model in Eq. (1) accepts t categorical variables, each of which has $k_t \in [0, l_t]$ categories:

$$E_{zone\ tot} = \sum_{k_1=0}^{l_1} \sum_{k_2=0}^{l_2} \dots \sum_{k_t=0}^{l_t} \sum_{j=0}^{m} (a_{k_1,k_2,\dots,k_t,j} \cdot r_{k_1,k_2,\dots,k_t,j}) \qquad (kWh \cdot y^{-1}) \tag{2}$$

For this study, the generalised model in Eq. (2) is adapted to accommodate the typology-specific information. Thus, a model for estimating bulk total annual energy use $E_{zone\ tot}$ of a geospatial zone consisting of n building types with m units is defined as:

$$E_{zone\ tot} = \sum_{i=0}^{n} \sum_{j=0}^{m} (a_{i,j} \cdot r_{i,j}) = \sum_{i=0}^{n} A_i \cdot R_i^T \qquad (kWh \cdot y^{-1}) \tag{3}$$

where:

$a_{i,j}$ – heated floor area (m^2) of j^{th} unit of the i^{th} type;
$r_{i,j}$ – energy use intensity $(kWh \cdot m^{-2} \cdot y^{-1})$ of j^{th} unit of the i^{th} type;
$A_i = [\, a_{i,0} \dots a_{i,m} \,]$ – a row matrix containing the known values of heated floor area (m^2) of all m units of the i^{th} type;
$R_i^T = [\, r_{i,0} \dots r_{i,m} \,]^T = \begin{bmatrix} r_{i,0} \\ \dots \\ r_{i,m} \end{bmatrix}$ – a column matrix containing the unknown values of energy use intensity $(kWh \cdot m^{-2} \cdot y^{-1})$ of all m units of i^{th} type.

Obtaining $E_{zone\ tot}$ without knowing the exact values of R_i^T represents an inverse probability estimation problem: "given the known univariate distribution of the uncertain model input, estimate the distribution of uncertain model output by repetitive random sampling of these inputs". This procedure is referred to as Monte Carlo simulation. With top-down reasoning, the distribution of R_i^T is inferred from higher spatial level, i.e. from the data available for the city.

Since the model in Eq. (3) applies in-sample summation, the simulated output tends towards normal distribution (Fig. 1) as the number of simulations gets larger, according to the Central Limit Theorem. Therefore, the output across simulation trials is well described by two parameters: a measure of central tendency given by the mean value μ and the dispersion properties quantified by standard deviation (SD) σ.

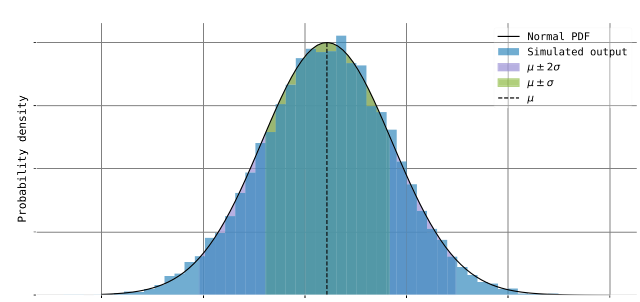

Fig. 1. Mean and standard deviation in the univariate distribution of simulated results.

2.2. Data sources

2.2.1. The size and structure of built stock

The known inputs of the Eq. (3) are the values of heated floor area and building type for each unit within the geospatial zone. This information is being collected, updated and made available through the Norwegian Cadastral system which is managed by the Norwegian Mapping Authority. The Norwegian Cadastral system was established in 2010 following the guidelines and the requirements of the Land Act 2005, offering a reliable, transparent and updated registry of all land users [33]. Currently, the registry contains almost 5 millions of registered properties nationwide, classified according to the Standard for building types [34].

Concerning the Trondheim municipality, the cadastral system's registry contains more than 92000 units covering 12 km^2 of total constructed floor area, 83% of which is residential. A spatial join of the attributes of these units enables the analysis and simulation per geospatial zones with a flexible spatial resolution, ranging from the individual building to the city level. A square grid of 1×1 km is an arbitrary choice of spatial resolution made to exemplify the computational procedures proposed in this study. Fig. 2 illustrates this geospatial grid over the urban territory and some attributes of the built stock per grid cell analysed: total constructed floor area (colour intensity) and the share of residential in the constructed floor area (marker size). The figure suggests that high construction density is present in the historical centre of the city and in the more industrialised southern part. These areas are often associated with a higher rate of non-residential units. The majority of the urban territory, however, is represented by sparse construction density and is dominated by residential buildings.

2.2.2. Energy performance of built stock

Inferring the statistical properties of the r.v. energy use intensity ($kWh \cdot m^{-2} \cdot y^{-1}$) per building type in Trondheim is based on the Norwegian EPC dataset. EPC dataset is the component of the Norwegian Energy Labelling System for Houses and Dwellings – a mechanism established to support the progress towards low energy use in communities and nationwide. The Norwegian EPC Scheme follows the implementation of the Energy Performance of Buildings Directive (EPBD), similarly to the other EU's Member States [35,36].

Hence, the Norwegian EPC scheme has been in place since 2010 intended to ensure Norway's compliance with the EPBD 2002/91/EC, to improve building energy awareness and to promote high energy performance. By 2016, more than 670 000 certificates were issued. The background for certification, legislative and practical framework in the Norwegian context was discussed in source [37].

The total annual energy use ($kWh \cdot y^{-1}$) per certified unit is voluntarily specified and registered in approx. 10% of all certificates. These values, normalised per unit of heated floor area (m^2), constitute an empirical sample of 4660 records representing dwelling/building units registered in Trondheim. Fig. 3 illustrates the univariate distribution of energy use intensity in this sample for both, non-residential (NR) and residential (RE) units. The figure demonstrates that the statistical properties of energy use intensity, accommodated by the shape of the density histogram, vary significantly per building type. This applies to such parameters as dispersion (range of values and variance), central tendency (mean, median, and mode), skewness and kurtosis. Accounting for these distinct properties is expected to positively contribute to the accuracy of best estimates and the margins of error provided by the built stock model.

2.3. Density estimation

To simulate R_i^T for the arbitrary number of units, one must know the relative likelihood of its values to occur. This information is communicated with two properties of a parameterised theoretical r.v: PDF and cumulative distribution function (CDF). Deciding which distribution type and parameters characterise the theoreti-

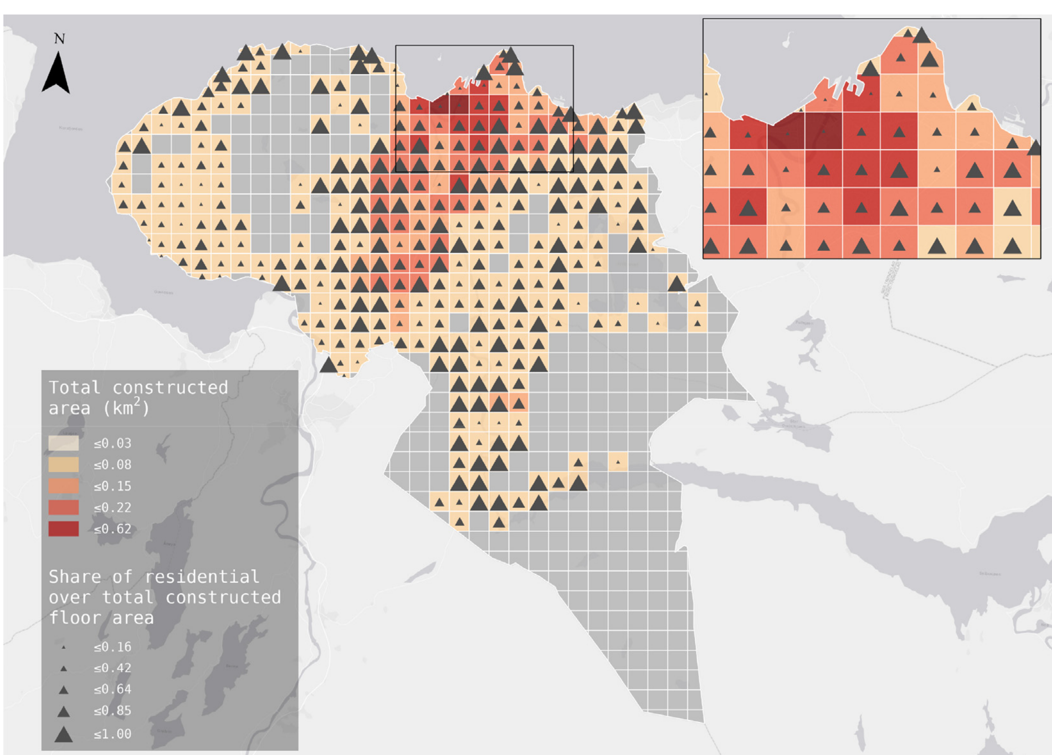

Fig. 2. Spatial variabilities of total constructed floor area and the share of residential buildings per geospatial zone in Trondheim.

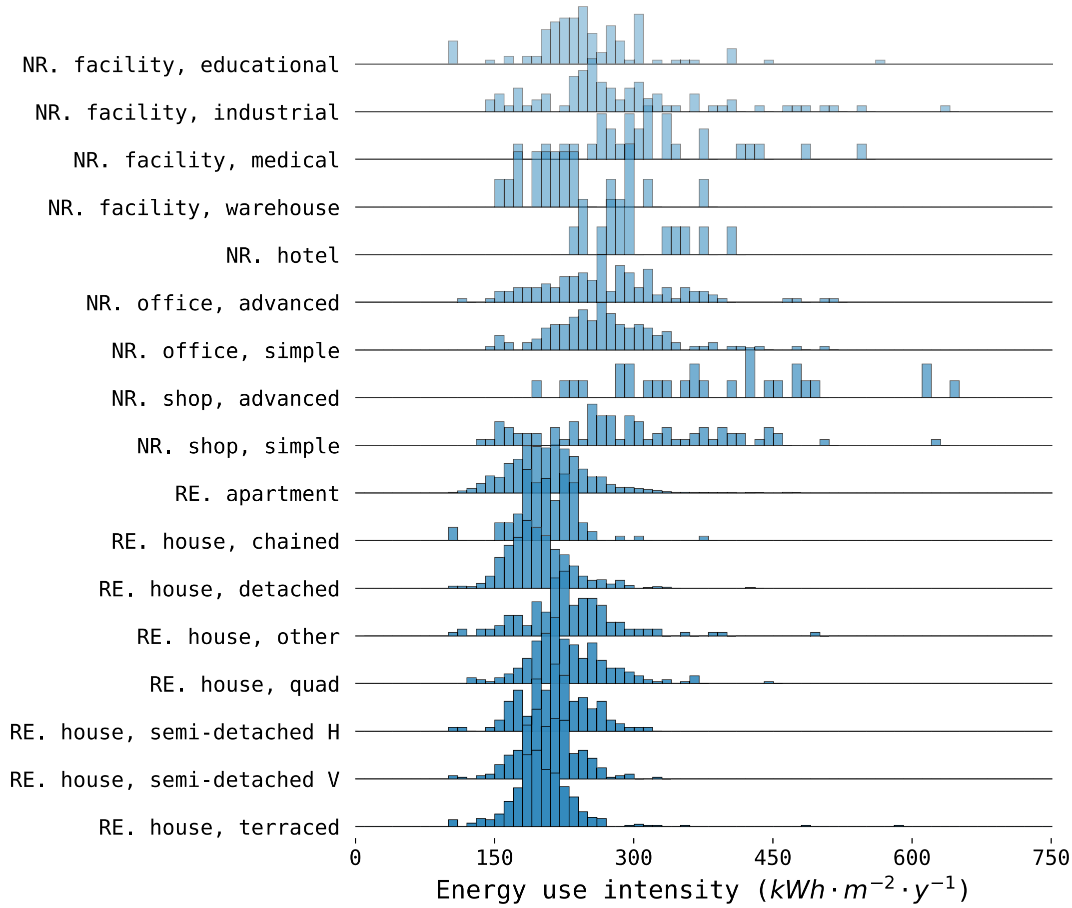

Fig. 3. Univariate distribution of energy use intensity per building type in Trondheim.

cal continuous r.v. can be carried out within the density estimation procedure consisting of:

1. Fitting (parameterising) each available distribution type individually based on the empirical sample;
2. Quantifying the goodness-of-fit;
3. Finding the theoretical parameterised distribution that characterises the sample best.

Maximum likelihood estimation (MLE) [38,39] is one way to fit the parameters of the theoretical distribution to the sample. In MLE, the objective is set as: given the observed sample $x : [x_1, x_2, x_3, \dots, x_n]$ and the theoretical continuous PDF $p_X(x|\theta)$, find the vector θ of parameters that are most likely to generate such sample. This is achieved through maximising the log-likelihood function:

$$f(\theta, x) = max_\theta \left\{ ln \left[\prod_i^n p_X(x_i|\theta) \right] \right\} = max_\theta \left\{ \sum_i^n ln[p_X(x_i|\theta)] \right\} \quad (4)$$

Finding the objective function in Eq. (4), represents a multivariate unconstrained optimisation problem with potentially noisy (non-smooth) functions. An effective search method for the problems of this kind is downhill simplex (Nelder–Mead) method [40–42] which is also known as a generalisation of dichotomic search to higher dimensions. Depending on the distribution type, vector θ may have between two and five parameters, meaning that the simplex takes a form of a triangle, tetrahedron, pentachoron or 5-simplex accordingly. Convergence to the optima is carried out through stepwise improvement of the initial guess without computing the gradients. The exit condition is either achieving the desired error tolerance or lack of progress in objective function compared to previous iterations.

Fig. 4 illustrates the results of the MLE-based fitting of some theoretical distributions to the empirical sample (also shown in Fig. 3) of detached houses in Trondheim. It is shown that the PDFs (Fig. 4 [A]) and CDFs (Fig. 4 [B]) follow the shape of sample distribution with various precision. This entails deciding which distribution describes the sample best and requires quantitative metrics to facilitate the decision.

The goodness-of-fit between the continuous theoretical distribution and the empirical sample may be studied with a non-parametric Kolmogorov–Smirnov (KS) test [43,44]. The KS test quantifies the difference between the empirical CDF represented by the step-function and the CDF of the theoretical distribution (as shown in Fig. 4 [B]). The test returns two values of interest: the D_n statistic (Eq. 5) and the measure of statistical significance (p-value).

D_n for the sample with sample size (SS) n is the supremum (the maximum or the bound) of the absolute difference between the CDF of a theoretical distribution $P_0(x)$ and the empirical CDF $\hat{P_n}(x)$ [45]:

$$D_n = \sup_x |P_0(x) - \hat{P_n}(x)| \quad (5)$$

The p-value corresponds to the survival function $(1 - CDF)$ in the asymptotic distribution of D_n at $\sqrt{n} \cdot D_n$. In statistical hypothesis testing, p-value serves as the basis for accepting or rejecting the hypotheses about the conformity between distributions. Low p-value suggests statistically significant evidence against the asserted

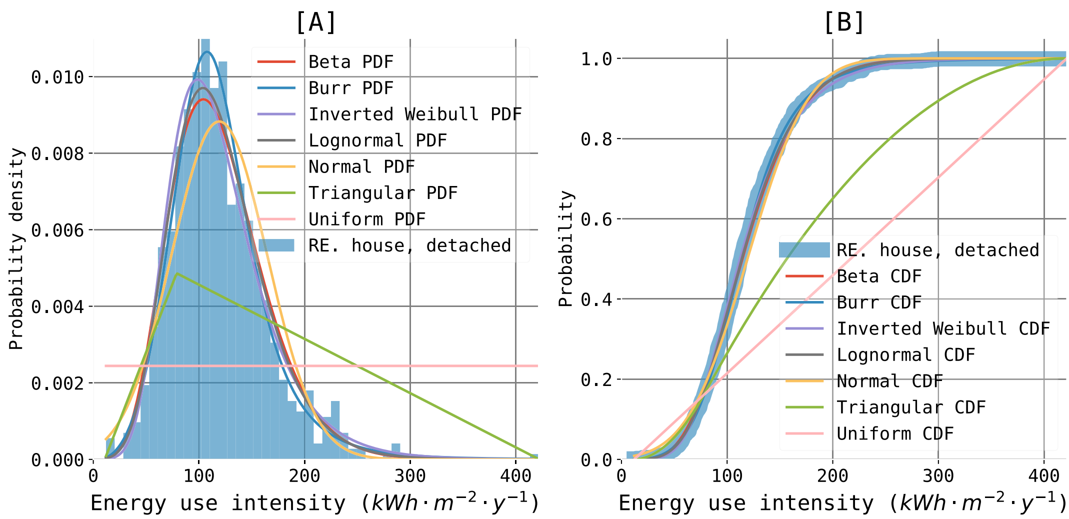

Fig. 4. Sample density [A] and the empirical CDF [B] with fitted PDFs and CDFs accordingly for some theoretical distributions.

null hypothesis: "sample x is generated by the r.v. X with the PDF $p_X(x|\theta)$". The threshold value α for statistical significance has to be chosen prior to the experiment. The null hypothesis, therefore, is rejected under the condition $p < \alpha$ and is failed to be rejected otherwise.

2.4. Computational framework

The components of the probabilistic model (Section 2.1) together with methods and procedures discussed in Section 2.3 are organised in a computational framework (Fig. 5).

Density estimation component is designed to find the parameterised theoretical distributions that describe the energy use intensity of distinct building types. The process starts with obtaining a subset of the sample with energy use intensity corresponding to particular building type in the city. For each of the available theoretical distributions, their parameters are fitted with MLE using the downhill simplex method. MLE is terminated either if the objective function is found with the absolute error tolerance $\epsilon \leqslant 1 \cdot 10^{-10}$ or if the maximum number of iterations $N \times 200$ (N – the number of simplex's dimensions) is achieved. The KS test is then carried out with the CDFs of an empirical sample and of a fitted distribution. The null hypothesis is rejected under the condition $p < 0.05$. At the end of the loop, the most suitable distribution amongst those passing the test is selected. This choice is based on comparing the associated D statistic and selecting the smallest. The procedure is then repeated for all building types in Trondheim.

Within the simulation component, the primary loop carries out iterations over the grid cells. In each cell, a secondary loop iterates over the building types that are present which is followed by retrieving a row matrix A_i. A series of 10000 Monte-Carlo trials are then carried out using the Mersenne Twister [46,47] pseudo-random number generator. At each trial, a column matrix R_i^T is simulated as a r.v. using the previously found parameterised distribution that characterises this building type. The dot product $A_i \cdot R_i^T$ is computed per trial and stored as one likely value of total energy use by the typology in the cell. When the iterations over the building types are complete, the total energy use across simulation trials per typology are aggregated to the grid cell level. This output forms a normal distribution, the mean value μ and the standard deviation σ for which are computed.

3. Results

The output of Density Estimation component, as discussed in Sections 2.3 and 2.4, are the parameterised distributions that are found to represent the data generation processes for individual building types in Trondheim. This information is summarised, together with the metrics for goodness-of-fit and sample statistic (minimum/maximum values and the sample size) in Table 1.

Vector θ of distribution's parameters in Table 1 is structured as $\theta : [\theta_1, \ldots, \theta_{k-1}, \theta_k]$. Two last elements in the list θ_{k-1} and θ_k are location (l) and scale (s) parameters accordingly. Any additional shape parameters, if applicable, are at the beginning of this list.

An example of interpreting the information provided in Table 1 is the following: energy use intensity of "RE. house, terraced" in Trondheim conforms to Johnson SU distribution parameterised by vector $[-0.392,\ 1.309,\ 108.848,\ 40.094]$. The difference between the empirical CDF of a sample with the size 407 and the CDF of this theoretical distribution is found to be 0.02. This difference is insignificant ($p > \alpha :\ 0.96 > 0.05$), thus implying a failure to reject the asserted null-hypothesis "the empirical sample is generated by Johnson SU r.v. with these parameters". The empirically evident range of values taken by the r.v. is $[25,\ 623]\ kWh \cdot m^{-2} \cdot y^{-1}$. Energy use intensity of "RE. house, terraced" in Trondheim within this range, therefore, can be simulated as the Johnson SU r.v. that has the PDF of a form:

$$f(x,a,b,l,s) = \frac{b}{s\sqrt{\left(\frac{x-l}{s}\right)^2+1}}\phi\left(a + b\log\left(\frac{x-l}{s} + \sqrt{\left(\frac{x-l}{s}\right)^2+1}\right)\right) \quad (6)$$

where:

x – energy use intensity ($kWh \cdot m^{-2} \cdot y^{-1}$);
ϕ – normal PDF;
a, b, l, s – a list of parameters identified with MLE:
- $[a,\ b] = [-0.392,\ 1.309]$ – distribution-specific shape parameters;
- $[l,\ s] = [108.848,\ 40.094]$ – location and scale parameters accordingly.

The outputs of a probabilistic model (Eq. (3) in Section 2.1), produced within the Simulation component (Section 2.4) using the parameterised distributions listed in Table 1 are the estimates of the mean (μ) and the SD (σ) of the bulk total annual energy use

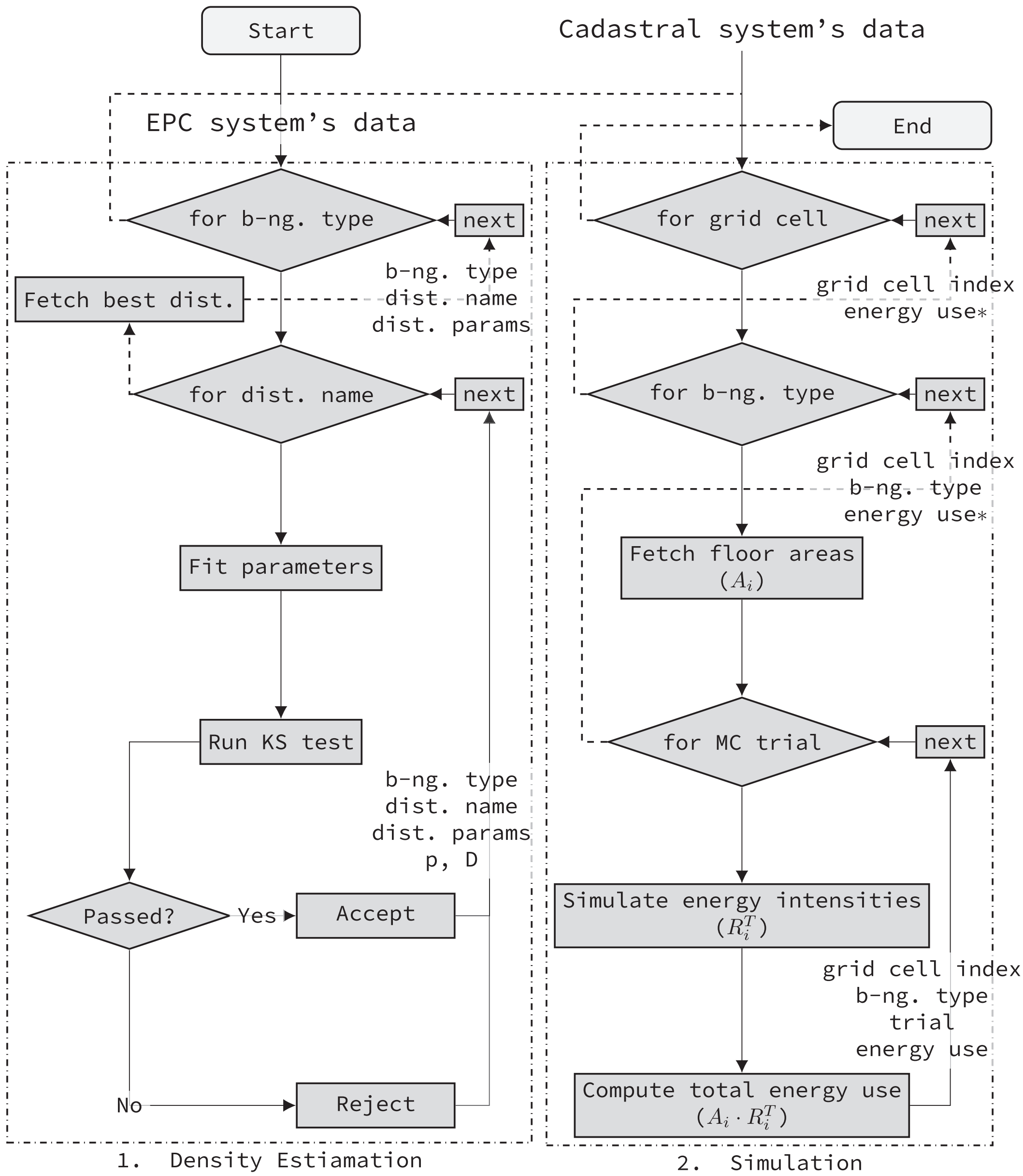

Fig. 5. The flowchart of computational procedures (Column or row matrices are denoted by asterisk ∗. Otherwise, a single categorical or numerical value is returned.).

Table 1
Sample information and parameterised distributions identified per building type in Trondheim.

Building type	Distribution	Parameters	D	*p*-value	Min	Max	SS
NR. facility, educational	Mielke Beta-Kappa	[6.444, 5.037, −0.446, 172.923]	0.06	0.85	55	602	117
NR. facility, industrial	Folded Cauchy	[2.612, 47.979, 57.731]	0.06	0.87	48	698	80
NR. facility, medical	Log-laplace	[3.662, −0.615, 264.554]	0.06	1.00	90	573	31
NR. facility, warehouse	Alpha	[3.785, −68.437, 771.015]	0.11	0.98	65	353	17
NR. hotel	Inverse Gaussian	[0.365, 136.117, 324.874]	0.12	0.93	167	390	17
NR. office, advanced	Logistic	[210.129, 51.094]	0.03	1.00	15	535	128
NR. office, simple	Tukey-Lambda	[−0.087, 205.013, 38.051]	0.04	0.99	51	524	129
NR. shop, advanced	Maxwell	[20.864, 219.505]	0.08	0.99	115	703	28
NR. shop, simple	Exponentially modified normal	[1.036, 154.399, 88.784]	0.06	0.93	33	680	80
RE. apartment	Mielke Beta-Kappa	[2.641, 5.967, −0.348, 163.23]	0.01	0.91	12	467	1844
RE. house, chained	Vonmises (non-circular)	[3.808, 131.935, 69.196]	0.07	0.67	66	349	103
RE. house, detached	Exponentially modified normal	[1.374, 82.776, 26.539]	0.01	1.00	12	422	881
RE. house, other	Tukey-Lambda	[−0.156, 157.477, 27.537]	0.04	0.98	17	516	136
RE. house, quad	Johnson SU	[−0.726, 1.561, 121.128, 67.087]	0.04	0.83	22	438	243
RE. house, semi-detached H	Rice	[1.477, 44.969, 53.336]	0.04	0.99	51	276	124
RE. house, semi-detached V	Exponentially modified normal	[0.645, 111.087, 34.003]	0.03	0.98	16	292	295
RE. house, terraced	Johnson SU	[−0.392, 1.309, 108.848, 40.094]	0.02	0.96	25	623	407

per grid cell. These attributes are displayed in Fig. 6. Colour intensities in the figure illustrate spatial variations of μ, whereas the diameter of markers is proportional to σ.

Fig. 6 facilitates the analysis of the city's energy hotspots (areas with high μ) and where the additional information may be needed (high σ). Examining these results against the data on the built stock in Fig. 2 demonstrates that the mean of the simulated bulk total annual energy use in Fig. 6 is correlated with the built area density and the share of energy-intensive non-residential buildings. The energy hotspots are located in the centre of the city and the industrial southern suburbs. Remote, mostly residential areas, which are known to have low unit density, are associated with relatively low energy use. Standard deviation correlates with mean of simulated bulk total annual energy use in the geospatial zones. Further analysis suggests a roughly linear relationship (Fig. 7) between μ and σ.

The scatter plot in Fig. 7 presents the results separated into two groups based on the arbitrary condition $c \geqslant 0.1$ and $c < 0.1$ where c is the coefficient of variation ($c = \sigma/\mu$). More detailed analysis suggested that c exceeds 0.1 for those grid cells where constructed units density is sparse or, alternatively, dense but with a high share of non-residential build area (above 60%). For most of the areas where energy use is high, c remains below 0.1. According to the empirical interpretation of normal distribution, $c < 0.1$ suggests at least 67% of confidence that the true value of bulk total annual energy use in the spatial zone is within the range $\mu \pm 10\%$. Similarly, $\mu \pm 20\%$ is the 95% confidence range.

4. Discussion

Through the case study developed for Trondheim, this article demonstrates a top-down modelling approach with the inverse uncertainty propagation for urban energy mapping purposes. Methodically, it implies spatial downscaling of the energy use intensity values from the city-level to the finer resolution. As a

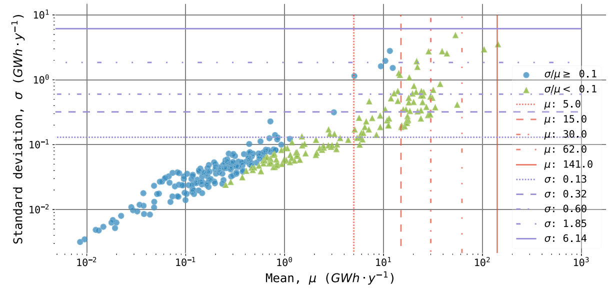

Fig. 7. The relationship between mean and standard deviation of simulated bulk total annual energy use per geospatial zone.

result, the probabilistic estimates of bulk total annual energy use per geospatial zone may be obtained. Random sampling used to compute these estimates enables to address aleatory uncertainties and heterogeneity discussed by Booth et al. [27]. The model in Eq. (3) does not contain any parameters and does not assume any rigorous knowledge about the factors that drive the phenomenon, thus eliminating the associated epistemic uncertainties.

Typology-specific density estimation of energy use intensity carried out at the city-level enabled to downscale the analysis to 1×1 km square grid. This choice of spatial resolution was arbitrary and can be substituted in the model with any other spatial or administrative boundaries. The parsimonious model design options enable the upper and lower boundaries to be anywhere between national and the building levels accordingly.

Within the modelling framework, disaggregation by exogenous influencing factors is supported and may lead to more accurate estimates. The simplest top-down model (Eq. (1)), for example, would contain a single parameterised r.v. that simulates energy intensity of all units with no regards to any other exogenous fac-

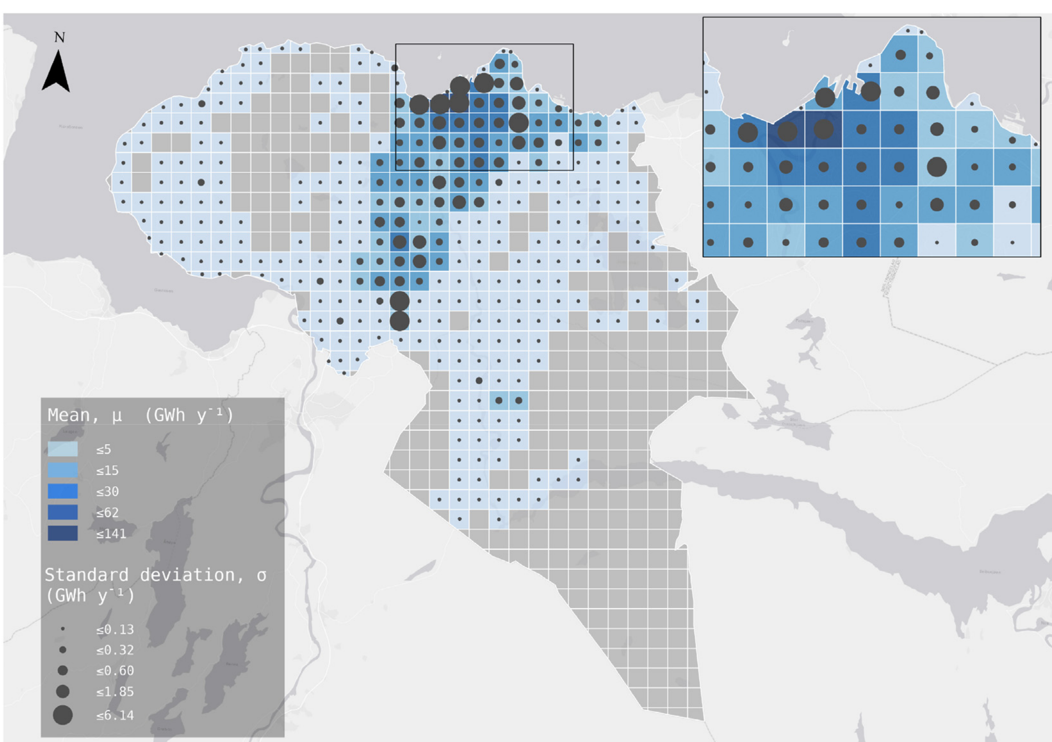

Fig. 6. Spatial variation of the mean and the standard deviation in simulated bulk total annual energy use per grid cell in Trondheim.

tors in the computational framework (Section 2.4). In this study, disaggregation by building types is beneficial because typology-specific r.v. evidently conveys a more detailed empirical information about the unique data generation process, i.e. relative likelihood of energy use intensity values to occur. Such disaggregation in a top-down manner may be done further to reach the necessary levels of technical details. Fig. 8 illustrates one of the plausible next steps in the disaggregation procedure to reflect distinct energy performance in "NR. apartment" given the Construction Year (CY).

Fig. 8 illustrates that the originally bimodal distribution at a higher (Fig. 8 [A]) level may be successfully disaggregated into at least two unimodal groups (Fig. 8 [B]). The basis for disaggregation in this case is arbitrary set to the year 1955, which is the first threshold between construction year classes defined for the Norwegian built stock within the TABULA [48] system. The empirical evidence supports such archetypes definition, since the two samples are characterised by substantially distinct statistical properties. Stricter energy efficiency standards for buildings, together with the other contributing factors, directly or indirectly led to the observable shift of central tendency between the two distributions. Dispersion is also affected by regulatory, technological, and socio-economic transformations.

A measure of dispersion quantifies yet unexplained uncertainties. By examining this parameter, a modeller may decide whether the remaining level of uncertainty is acceptable to address the purpose of the modelling or if further disaggregation is needed. If the latter, additional exogenous factors may be tested. Two distributions in Fig. 8 [B] are associated with the dispersion smaller than the composite (Fig. 8 [A]), meaning that the construction period explains a portion of the original uncertainties.

Statistical significance of the difference between the disaggregated distributions suggests the level of sensitivity to disaggregation by the exogenous factors. Since the resulting distributions in Fig. 8 [B] are distinct, it is plausible to disaggregate by construction period. The reverse statement also holds true – mutually conforming distributions exhibited by disaggregated groups suggest a little or no benefit from disaggregation. The previously mentioned KS test may be used to quantify the difference between the two empirical samples. A null hypothesis for testing is formulated as: "two samples are drawn from the same continuous distribution", and high p-value ($p > \alpha$) implies a failure to reject this null hypothesis. An example of such pairwise testing of samples with the significance threshold $\alpha = 0.05$ is presented in Table 2.

Test 1 in Table 2 resulted in a large D-statistic. If the null hypothesis is true, obtaining such a large value of D by chance is unlikely given the samples sizes. This likelihood is reflected by the negligibly low p-value which suggests to reject the asserted null-hypothesis. This indicates high sensitivity to the construction period. The null hypothesis cannot be rejected in Test 2 and therefore, these two samples are found to conform even though they

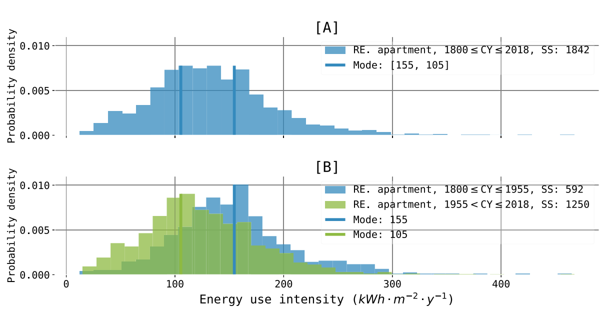

Fig. 8. Univariate distribution of energy use intensity of apartments in Trondheim: [A] composite; [B] disaggregated by construction period.

represent distinct archetypes in TABULA. Higher likelihood of apartments from Period 1 to be in their renovated state may explain the absence of statistically significant difference between energy use intensities of samples in Test 2.

The disaggregation procedure discussed above reveals the source of both, advantages and limitations of the proposed approach: if exogenous factors of influence on energy use intensity are not represented in the model, they are assumed consistent between the upper and the lower spatial levels. Practically, it suggests that the model needs to reflect only those factors that are known to lead to spatial variations of energy use. Otherwise, the unexplained uncertainty of energy use intensity entails larger dispersion in the simulated results. Although the most essential operations for density estimation and probabilistic simulation can be automated, the choices behind disaggregation procedures remain manual. This entails that the choice of the acceptable unexplained uncertainty level and the number of categories for disaggregation must involve domain knowledge even if these judgements are supported by quantitative metrics.

The available sample size is an important aspect that represents a source of epistemic uncertainty associated with the proposed modelling approach. Density estimation with insufficient sample size may suggest the type of distribution or the parameters that poorly describes the data generation process and should be avoided whenever possible. A frequentist-based density estimation discussed in this article may inform empirically the choice of prior distribution in Bayesian inference, which may lead to the need for fewer samples and higher reliability of the latter. Therefore, complementing the two approaches may be beneficial for future studies. A rational threshold α of statistical significance for the p-value needs to be established in the discipline to support the coherence between the studies alike. Moreover, a rigorous recommendation on the domain-specific smallest sample size for density estimation is not yet available. Therefore, this study agrees with Booth et al. [27] on the need for adapting the existing practices from other domains, e.g. physics, medicine, and economy to assist overcoming these challenges in built stock energy modelling. The implications of data quality can be regarded as an additional source of epistemic uncertainty in modelling. High accuracy and soundness of conclusions made through statistical inference, similarly to other techniques that rely on data, require independent and identically distributed (i.i.d.) random samples. A practical way to obtain such data is through stratified (e.g. block) design of the experiment. Substituting such sample with potentially biased data may entail inaccuracies. It is, for example, debatable if any of the EPC system's designs is capable of providing randomised i.i.d. samples, because the sole phenomenon of certification is under the strong influence of numerous socio-techno-economic tendencies that may cause the bias.

Given the availability and high quality of empirical data, however, virtually any level of technical details and end-uses may be reached through such modelling. No obstacles are anticipated in evaluating the implications of altering building envelope, energy supply systems and/or indoor environmental quality at a large scale. Similar approaches can also accommodate energy use for source-specific space heating, hot water supply, plug loads and other through multivariate distributions, e.g. copulas. Currently, these capabilities of top-down modelling are underestimated and poorly explored in the domain. It is, however, shown through this study that handling aleatory uncertainties and heterogeneity of buildings yields numerous benefits, mitigating the "performance gap" being one of them. It is also evident that data-enabled knowledge discovery and modelling, facilitated by statistical inference and probabilistic programming, may complement already established architectural and engineering-based foundations of built stock energy research. Synthesising the methods from these

Table 2
Pairwise KS test results on conformity between samples from various construction periods for apartments in Trondheim

Test	Period 1 (SS)	Period 2 (SS)	D	p-value	Null-hypothesis
1	$1800 \leqslant CY \leqslant 1955(592)$	$1955 < CY \leqslant 2018(1250)$	0.256	$1.212 \cdot 10^{-23}$	$p < 0.05$: Reject
2	$1955 < CY \leqslant 1980(642)$	$1980 < CY \leqslant 2000$ (147)	0.086	0.323	$p > 0.05$: Failed to reject

domains is also the key objective set for the future developments of Built Stock Explorerhttps://builtstockexplorer.indecol.ntnu.no.

5. Conclusions

This study draws attention to the topic of urban energy mapping, where the uncertainties must be eliminated to the best possible extent while keeping data collection efforts rational. A proposed probabilistic top-down modelling approach is shown to have a high potential for addressing this trade-off. Probabilistic origins naturally account for aleatory uncertainties behind the phenomena and for the heterogeneity of buildings through parameterising the random variables. Disaggregation by exogenous factors conveys these uncertainties without the loss of information. Non-parametric model structure enables to address the epistemic uncertainties, associated with approximations and simplifications that are necessary otherwise. A suggested modelling approach offers adaptiveness to the purpose of the modelling and the associated level of details. The key benefits of the approach emerge from the ability to quantify and control the uncertainties while adding the explanatory (exogenous) factors to the model.

The results suggest that the typology-specific energy use intensity can be represented by parameterised random variables. With these random variables and the information on geospatial coordinates, size and type of buildings, bulk total annual energy use can be estimated at a spatially downscaled area, e.g. 1×1 km grid cell.

The coefficient of variation for most of Trondheim's energy hotspots remains below 0.1, which makes the results already suitable for many practical applications. Urban areas of high energy use, for example, can be prioritised for developing refurbishment strategies and/or deploying more efficient energy supply solutions. By resolving the energy-related bottlenecks first, higher energy and environmental performance of built stock may be achieved within a shorter time horizon. These results may also aid the planning of new construction and the energy-intensive units with a minimum intervention into the existing infrastructure for energy generation and distribution purposes.

Considering the additional factors of influence on building energy performance may further improve the accuracy of the modelling. The feasibility of using these factors can be guided by statistical hypothesis testing. Currently, a substantial barrier for such modelling is the absence of both, established practices for defining the levels of statistical significance and the recommendations on sample size for such tasks. These and the related challenges entail a yet unresolved epistemic uncertainty associated with the proposed modelling approach.

The instruments and the techniques discussed in this article may produce reliable insights into the spatial variabilities of the building energy use. They lay the foundations for the work ahead which will synthesise the probabilistic status quo with the probabilistic forecasting of future developments in the built stock. And hence, will assist with establishing the pathways towards higher efficiency of the built environment.

CRediT authorship contribution statement

Ruslan Zhuravchak: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Visualization. **Raquel Alonso Pedrero:** Conceptualization, Software, Formal analysis, Writing - original draft, Visualization. **Pedro Crespo del Granado:** Validation, Writing - review & editing. **Natasa Nord:** Validation, Writing - review & editing, Supervision. **Helge Brattebø:** Validation, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the support from the Research Council of Norway through the research projects:- Methods for transparent energy planning of urban building stocks - ExPOSe (project number 268248) under EnergiX program;- The value of end-use flexibility in the future Norwegian energy system - Flexbuild (project number 294920).

References

[1] F. Johari, G. Peronato, P. Sadeghian, X. Zhao, J. Widén, Urban building energy modeling: State of the art and future prospects, Renewable and Sustainable Energy Reviews 128 (2020), https://doi.org/10.1016/j.rser.2020.109902 109902.
[2] T. Abergel, C. Delmastro, K. Lane, Tracking buildings 2020, Tech. rep., International Energy Agency (IEA), 2020.
[3] J.-P. Petersen, E. Heurkens, Implementing energy policies in urban development projects: The role of public planning authorities in Denmark, Germany and the Netherlands, Land Use Policy 76 (2018) 275–289, https://doi.org/10.1016/j.landusepol.2018.05.004.
[4] M. Kljajic, A.S. Andelkovic, I. Mujan, Assessment of relevance of different effects in energy infrastructure revitalization in non-residential buildings, Energy and Buildings 116 (2016) 684–693, https://doi.org/10.1016/j.enbuild.2015.02.033.
[5] N.H. Sandberg, I. Sartori, O. Heidrich, R. Dawson, E. Dascalaki, S. Dimitriou, T. Vimm-r, F. Filippidou, G. Stegnar, M.Š. Zavrl, H. Brattebø, Dynamic building stock modelling: Application to 11 European countries to support the energy efficiency and retrofit ambitions of the EU, Energy and Buildings 132 (2016) 26–38, 10.1016/j.enbuild.2016.05.100.
[6] L. Tozer, Catalyzing political momentum for the effective implementation of decarbonization for urban buildings, Energy Policy 136 (2020), https://doi.org/10.1016/j.enpol.2019.111042 111042.
[7] M. Shepero, J. Munkhammar, Spatial Markov chain model for electric vehicle charging in cities using geographical information system (GIS) data, Applied Energy 231 (2018) 1089–1099, https://doi.org/10.1016/j.apenergy.2018.09.175.
[8] D.J. Sailor, L. Lu, A top–down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas, Atmospheric Environment 38 (17) (2004) 2737–2748, https://doi.org/10.1016/j.atmosenv.2004.01.034.
[9] S. Heiple, D.J. Sailor, Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles, Energy and Buildings 40 (8) (2008) 1426–1436, https://doi.org/10.1016/j.enbuild.2008.01.005.
[10] S.T. Moghadam, C. Delmastro, S.P. Corgnati, P. Lombardi, Urban energy planning procedure for sustainable development in the built environment: A review of available spatial approaches, Journal of Cleaner Production 165 (2017) 811–827, https://doi.org/10.1016/j.jclepro.2017.07.142.
[11] U. Ali, M.H. Shamsi, M. Bohacek, K. Purcell, C. Hoare, E. Mangina, J. O'Donnell, A data-driven approach for multi-scale gis-based building energy modeling for analysis, planning and support decision making, Applied Energy 279 (2020), https://doi.org/10.1016/j.apenergy.2020.115834 115834.
[12] M. Österbring, É. Mata, L. Thuvander, M. Mangold, F. Johnsson, H. Wallbaum, A differentiated description of building-stocks for a georeferenced urban bottom-up building-stock model, Energy and Buildings 120 (2016) 78–84, https://doi.org/10.1016/j.enbuild.2016.03.060.
[13] S.T. Moghadam, J. Toniolo, G. Mutani, P. Lombardi, A GIS-statistical approach for assessing built environment energy use at urban scale, Sustainable Cities and Society 37 (2018) 70–84, https://doi.org/10.1016/j.scs.2017.10.002.

[14] B. Howard, L. Parshall, J. Thompson, S. Hammer, J. Dickinson, V. Modi, Spatial distribution of urban building energy consumption by end use, Energy and Buildings 45 (2012) 141–151, https://doi.org/10.1016/j.enbuild.2011.10.061.
[15] Y. Yamaguchi, Y. Shimoda, M. Mizuno, Proposal of a modeling approach considering urban form for evaluation of city level energy management, Energy and Buildings 39 (5) (2007) 580–592, https://doi.org/10.1016/j.enbuild.2006.09.011.
[16] R. Choudhary, Energy analysis of the non-domestic building stock of Greater London, Building and Environment 51 (2012) 243–254, https://doi.org/10.1016/j.buildenv.2011.10.006.
[17] H. Yoshino, T. Hong, N. Nord, Iea ebc annex 53: Total energy use in buildings—analysis and evaluation methods, Energy and Buildings 152 (2017) 124–136, https://doi.org/10.1016/j.enbuild.2017.07.038.
[18] A.A.A. Gassar, S.H. Cha, Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review, Energy and Buildings 224 (2020), https://doi.org/10.1016/j.enbuild.2020.110238 110238.
[19] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews 13 (8) (2009) 1819–1835, https://doi.org/10.1016/j.rser.2008.09.033.
[20] J. Langevin, J. Reyna, S. Ebrahimigharehbaghi, N. Sandberg, P. Fennell, C. Nägeli, J. Laverge, M. Delghust, É. Mata, M. Van Hove, J. Webster, F. Federico, M. Jakob, C. Camarasa, Developing a common approach for classifying building stock energy models, Renewable and Sustainable Energy Reviews 133 (2020), https://doi.org/10.1016/j.rser.2020.110276 110276.
[21] C.F. Reinhart, C.C. Davila, Urban building energy modeling – a review of a nascent field, Building and Environment 97 (2016) 196–202, https://doi.org/10.1016/j.buildenv.2015.12.001.
[22] G. Tardioli, R. Kerrigan, M. Oates, J. O'Donnell, D. Finn, Data driven approaches for prediction of building energy consumption at urban level, Energy Procedia 78 (2015) 3378–3383, https://doi.org/10.1016/j.egypro.2015.11.754.
[23] A. Mastrucci, O. Baume, F. Stazi, U. Leopold, Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam, Energy and Buildings 75 (2014) 358–367, https://doi.org/10.1016/j.enbuild.2014.02.032.
[24] H.C. Gils, J. Cofala, F. Wagner, W. Schöpp, GIS-based assessment of the district heating potential in the USA, Energy 58 (2013) 318–329, https://doi.org/10.1016/j.energy.2013.06.028.
[25] M. Berger, J. Worlitschek, A novel approach for estimating residential space heating demand, Energy 159 (2018) 294–301, https://doi.org/10.1016/j.energy.2018.06.138.
[26] D. Meha, T. Novosel, N. Duić, Bottom-up and top-down heat demand mapping methods for small municipalities, case Gllogoc, Energy 199 (2020), https://doi.org/10.1016/j.energy.2020.117429 117429.
[27] A. Booth, R. Choudhary, D. Spiegelhalter, Handling uncertainty in housing stock models, Building and Environment 48 (2012) 35–47, https://doi.org/10.1016/j.buildenv.2011.08.016.
[28] W. Tian, Y. Heo, P. [de Wilde], Z. Li, D. Yan, C.S. Park, X. Feng, G. Augenbroe, A review of uncertainty analysis in building energy assessment, Renewable and Sustainable Energy Reviews 93 (2018) 285 – 301. doi:10.1016/j.rser.2018.05.029..
[29] M. Kavgic, A. Summerfield, D. Mumovic, Z. Stevanovic, Application of a Monte Carlo model to predict space heating energy use of Belgrade's housing stock, Journal of Building Performance Simulation 8 (6) (2015) 375–390, https://doi.org/10.1080/19401493.2014.961031.
[30] M. Hughes, J. Palmer, V. Cheng, D. Shipworth, Global sensitivity analysis of England's housing energy model, Journal of Building Performance Simulation 8 (5) (2015) 283–294, https://doi.org/10.1080/19401493.2014.925505.
[31] F. Zhao, S.H. Lee, G. Augenbroe, Reconstructing building stock to replicate energy consumption data, Energy and Buildings 117 (2016) 301–312, https://doi.org/10.1016/j.enbuild.2015.10.001.
[32] D. Yu, A two-step approach to forecasting city-wide building energy demand, Energy and Buildings 160 (2018) 1–9, https://doi.org/10.1016/j.enbuild.2017.11.063.
[33] Kommunal- og moderniseringsdepartementet, Lov om eigedomsregistrering (matrikkellova) LOV-2005-06-17-101 (2005). http://lovdata.no/lov/2005-06-17-101..
[34] STATISTISK SENTRALBYRÅ, Seksjon for eiendom-, areal- og primærnæringsstatistikk, Standard for bygningstype/matrikkelen (2000). https://www.ssb.no/klass/klassifikasjoner/31..
[35] M. Prieler, M. Leeb, T. Reiter, Characteristics of a database for energy performance certificates, Energy Procedia 132 (2017) 1000–1005, 11th Nordic Symposium on Building Physics, NSB2017, 11–14 June 2017, Trondheim, Norway. doi:10.1016/j.egypro.2017.09.704..
[36] Y. Li, S. Kubicki, A. Guerriero, Y. Rezgui, Review of building energy performance certification schemes towards future improvement, Renewable and Sustainable Energy Reviews 113 (2019), https://doi.org/10.1016/j.rser.2019.109244 109244.
[37] T. Brekke, O.K. Isachsen, M. Strand, EPBD implementation in Norway. Status in December 2016, Tech. rep., Enova, Norwegian Water Resources and Energy Directorate (NVE), Norwegian Building Authority (DIBK) (2018).
[38] C. Robert, G. Casella, Monte Carlo Statistical Methods, Springer Texts in Statistics, Springer New York, 2013. doi:10.1007/978-1-4757-3071-5..
[39] N.T. Thomopoulos, Statistical Distributions. Applications and Parameter Estimates, 1st Edition,., Springer International Publishing, 2017, 10.1007/978-3-319-65112-5.
[40] J.A. Nelder, R. Mead, A simplex method for function minimization, The Computer Journal 7 (4) (1965) 308–313, https://doi.org/10.1093/comjnl/7.4.308.
[41] F. Gao, L. Han, Implementing the Nelder-Mead simplex algorithm with adaptive parameters, Computational Optimization and Applications 51 (1) (2012) 259–277, https://doi.org/10.1007/s10589-010-9329-3.
[42] W. Härdle, O. Okhrin, Y. Okhrin, Basic elements of computational statistics, Springer International Publishing (2017), https://doi.org/10.1007/978-3-319-55336-8.
[43] R. Feldman, C. Valdez-Flores, Applied Probability and Stochastic Processes, 2nd ed., Springer-Verlag, Berlin Heidelberg, 2010, 10.1007/978-3-642-05158-6.
[44] R. Bhattacharya, L. Lin, V. Patrangenaru, A Course in Mathematical Statistics and Large Sample Theory, Springer Texts in Statistics, Springer New York, 2016. doi:10.1007/978-1-4939-4032-5..
[45] G. Marsaglia, W.W. Tsang, J. Wang, Evaluating Kolmogorov's distribution, Journal of Statistical Software, Articles 8 (18) (2003) 1–4, https://doi.org/10.18637/jss.v008.i18.
[46] M. Matsumoto, T. Nishimura, Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, ACM Transactions on Modeling and Computer Simulation 8 (1) (1998) 3–30, https://doi.org/10.1145/272991.272995.
[47] R. Kneusel, Random Numbers and Computers, Springer International Publishing, 2018. doi:10.1007/978- 3- 319- 77697- 2..
[48] T. Loga, B. Stein, N. Diefenbach, Tabula building typologies in 20 european countries—making energy-related features of residential building stocks comparable, Energy and Buildings 132 (2016) 4–12, towards an energy efficient European housing stock: monitoring, mapping and modelling retrofitting processes. doi:10.1016/j.enbuild.2016.06.094..