Journal of Cleaner Production 233 (2019) 546–560

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Data-driven strategic planning of building energy retrofitting: The case of Stockholm

Oleksii Pasichnyi [a, *], Fabian Levihn [b, c], Hossein Shahrokni [a], Jörgen Wallin [d], Olga Kordas [a]

[a] Research Group for Urban Analytics and Transitions (UrbanT), Department of Sustainable Development, Environmental Science and Engineering (SEED), KTH Royal Institute of Technology, Teknikringen 10b, 100 44, Stockholm, Sweden

[b] Research Group for Urban Analytics and Transitions (UrbanT), Department of Industrial Economics (INDEK), KTH Royal Institute of Technology, Lindstedtsvägen 30, 100 44, Stockholm, Sweden

[c] AB Stockholm Exergi, Jägmästargatan 2, 115 41, Stockholm, Sweden

[d] Research Group for Urban Analytics and Transitions (UrbanT), Department of Energy Technology (ETT), KTH Royal Institute of Technology, Brinellvägen 68, 101 44, Sweden

ARTICLE INFO

Article history:
Received 30 June 2018
Received in revised form
8 May 2019
Accepted 30 May 2019
Available online 4 June 2019

Keywords:
Urban energy planning
Building energy retrofitting
Urban building energy modelling
High-resolution metered data
Urban energy efficiency
Stockholm

ABSTRACT

Limiting global warming to 1.5 °C requires a substantial decrease in the average carbon intensity of buildings, which implies a need for decision-support systems to enable large-scale energy efficiency improvements in existing building stock. This paper presents a novel data-driven approach to strategic planning of building energy retrofitting. The approach is based on the urban building energy model (UBEM), using data about actual building heat energy consumption, energy performance certificates and reference databases. Aggregated projections of the energy performance of each building are used for holistic city-level analysis of retrofitting strategies considering multiple objectives, such as energy saving, emissions reduction and required social investment. The approach is illustrated by the case of Stockholm, where three retrofitting packages (heat recovery ventilation; energy-efficient windows; and a combination of these) were considered for multi-family residential buildings constructed 1946–1975. This identified potential for decreasing heat demand by 334 GWh (18%) and consequent emissions reduction by 19.6 kt-CO_2 per year. The proposed method allows the change in total energy demand from large-scale retrofitting to be assessed and explores its impact on the supply side. It thus enables more precisely targeted and better coordinated energy efficiency programmes. The case of Stockholm demonstrates the potential of rich urban energy datasets and data science techniques for better decision making and strategic planning.

1. Introduction

Reaching the Paris Agreement goal of limiting climate change well below 2 °C requires transformation of energy systems globally. The world is becoming increasingly urbanised, with more than 50% of the global population currently living in cities (X. Wang et al., 2017). Thus, sustainable and viable cities play an important role in energy system transformation.

In 2014, buildings accounted for 31% of final energy use and 8% of energy-derived CO_2 emissions globally. Coal and gas are commonly used for supplying heating and cooling. The more stringent target for global warming, of 1.5 °C, implies that the carbon intensity of buildings must be limited to on average 36 g/kWh. This can be achieved through a combination of reduced heating and cooling demand, and/or transforming supplying energy to buildings (Rogelj et al., 2018).

In buildings, about 50% of the energy demand comes from space heating and cooling and 16% from water heating (IEA, 2017a). While conserving energy through the use of more efficient appliances could play an important role (Huebner et al., 2016), heating and cooling account for almost 80% of direct CO_2 emissions from buildings (IEA, 2017a). Therefore, reducing the space heating and cooling demand, combined with decarbonisation of district heating and electricity generation, is recognised as an essential strategy in realising a vision of 'decarbonised buildings' (EU Commission, 2016). Setting stricter requirements on new buildings is one part of the strategy, supported by energy standards for new buildings or

* Corresponding author.
E-mail address: oleksii.pasichnyi@abe.kth.se (O. Pasichnyi).

https://doi.org/10.1016/j.jclepro.2019.05.373

Nomenclature

Acronyms

CAPEX	capital expenditure
DH	district heating
EPC	energy performance certificate
ES	energy signature
GHG	greenhouse gas
IDE	integrated development environment
IRR	internal rate of return
KPI	key performance indicator
NPV	net present value
PVQ	present value quota
RDBMS	relational database management system
UBEM	urban building energy model

Variables and parameters

A_{temp}	heated area [m^2]
$q(t)$	specific heat power [W/m^2]
t	ambient temperature [°C]
θ	balance point temperature [°C]
c	base load (domestic hot water consumption) [W/m^2]
β	energy performance coefficient [$W/m^2.K$]
r	discount rate [%]
P	number of time periods
τ	time period [year]
R_τ	net cash flow in period τ [Swedish krona, SEK]

building energy codes established in around 60 countries for residential and non-residential buildings (IEA, 2017b). However, the vast majority of building stock in the European Union (EU), North America and China already exists and requires energy efficiency improvements (IEA, 2017a). In particular, more than 60% of existing buildings in the EU were built under limited or non-existent energy efficiency requirements and most of these will still be part of the building stock in 2050 (EU Commission, 2016). Thus, energy retrofitting of the existing building stock must constitute an important part of energy efficiency strategies (Zhou et al., 2016).

Although sizeable energy savings can be achieved through ‘conventional’ retrofitting (i.e. simple measures such as improving insulation or installing energy-efficient windows), there are multiple barriers to retrofit uptake by energy consumers (Bertone et al., 2018). These barriers include lack of concern, lack of awareness of possibilities and benefits, limited access to trusted advice and reliable information on technical details and financial models, split incentives between property owners and tenants (e.g. when improvements in energy efficiency are used to justify a rent increase) and financing constraints (Webber et al., 2015). These barriers can partly explain the reported information gap between city authorities who introduce retrofitting schemes and property owners or other individual energy consumers who must finance these schemes (Rysanek and Choudhary, 2013). Therefore, city authorities and large property owners need methods and tools that allow faster identification and evaluation of energy efficiency potential on a large scale (Cerezo Davila et al., 2016).

Large-scale identification of the potential of energy efficiency measures would enable mapping of the building stock, revealing cases where economically driven retrofitting is viable (Moschetti et al., 2018). This might stimulate the retrofitting market and increase the pace of retrofitting. It could also provide information to decision makers and legislators, facilitating introduction of new policies supporting the retrofitting market (Persson and Grönkvist, 2015). There are many models available for identification of optimum retrofitting packages for individual buildings based on their geographical, architectural and operational characteristics (Rysanek and Choudhary, 2013). In contrast, models targeting large-scale retrofitting strategies are rarely addressed in research, possibly because of a lack of high-resolution data on actual heat energy use by individual buildings (Gupta and Gregg, 2018). However, the cities of today are becoming smarter and more information-intense, as reflected in a number of recent studies on big-meter data within energy contexts (Deakin and Reid, 2018). Many of these studies focus on smart meters and electricity consumption, e.g. Gouveia et al. (2017) showed, for a limited sample of 19 households, how air temperature and electricity consumption could be understood through using smart meters. Other studies report on retrofitting and optimisation of electrical equipment (Ye and Xia, 2016). An Irish study of 4232 households showed how metering data could be used for targeted energy efficiency programmes addressing electric power consumption (Beckel et al., 2014). Another Irish study using metering data from more than 5000 households examined how social factors influence electricity consumption (Tong et al., 2016). Use of metering data has been extended beyond electric power, e.g. a Brazilian study included water consumption in a smart metering project (Fróes Lima and Portillo Navas, 2012), while a wider study analysed data from more than 8 000 000 gas and electricity meters in London and Chicago, to identify improvements for energy-related practices (Mohammadi and Taylor, 2017). Systematic collection and transparent availability of data are crucial in monitoring and improvement of retrofitting activities, in terms of both uptake and energy efficiency benefits (Brooks et al., 2014). Using rich datasets allow for more advanced data analysis techniques for developing retrofitting strategies and assessing conducted retrofits (Geyer et al., 2017).

Urban retrofit transitions are complex, co-evolutionary, nonlinear processes that draw upon a range of actors and focus on different levels and dimensions over time (Dixon et al., 2018). They can be seen as intersections between diverse infrastructures, non-aligned policy objectives and generally diverging interests and perspectives of a variety of social groups (Späth and Rohracher, 2015). Hence urban retrofitting requires an integrated approach combining environmental, economic and social sustainability (Eames et al., 2014). Large-scale implementation of energy efficiency measures can have an effect on total heat load duration and, consequently on operation of heat supply, and should be further studied (Truong et al., 2018). In a quest for even more holistic views, system boundaries for analysis of building retrofits can be expanded to the multi-system nexus (Engström et al., 2018), life cycle (Seo et al., 2018) and socio-economic (Mangold et al., 2016) perspectives.

The aim of this study was to develop, demonstrate and evaluate a data-driven approach to planning city-wide building retrofitting. The significant advantage of the proposed novel data-driven urban building energy model (UBEM)-based approach is the use of high-resolution metered data in fact-based modelling of the energy performance of the building stock. This paper contributes to the literature on city-scale building energy modelling by introducing the use of large datasets that have become available recently, particularly high-resolution metered data on heat energy use. The contributions of this study are two-fold. First, a novel method for strategic planning of building energy retrofitting based on the city-wide building energy modelling framework is presented. This method makes it possible to assess changes in the total energy demand from large-scale retrofitting and explore the impact on the supply side. The level of detail provided by the method makes it suitable for the development of customised strategies by city authorities and large housing institutions targeting retrofitting investment options that meet certain criteria. Second, the method was demonstrated for the case of retrofitting 'Million Programme' buildings in Stockholm, which have been regularly highlighted for their high energy-retrofitting potential (Johansson et al., 2017). Heat recovery ventilation and energy-efficient windows were chosen as retrofitting options. Analysis of three retrofitting scenarios from multiple perspectives (energy, environment, economic) was used to reveal the benefits and drawbacks of each retrofitting measure and to make strategic decisions based on information about related trade-offs. The results of the study served as input to the City of Stockholm's ongoing work on its climate strategy. Finally, this paper demonstrated the outcomes of digitalisation ongoing in the urban energy, pointing out the importance of urban energy data for urban strategic planning.

The remainder of the paper is organised as follows: section 2 presents the theoretical and empirical background to the study; section 3 describes the methodology in 10 sub-sections that outline the approach; section 4 describes the results of applying the strategic planning framework developed in section 3 to the case of large-scale retrofitting of 'Million Programme' buildings in Stockholm; section 5 discusses use of the framework and the results obtained; and finally section 6 presents some conclusions.

2. Background

2.1. Urban building energy modelling (UBEM)

New information and communication technologies (ICT), including internet of things (IoT) and applications, and digitalisation in all sectors of society are driving an exponential increase in the volume of information and the rate of information sharing at an ever-decreasing cost (Barahona and Pentland, 2007). This development has created a 'big data' challenge for conventional approaches to building energy simulations, creating a need for new ways of handling and utilising the data (Sanyal and New, 2014). The recently developed urban building energy model (UBEM) (W. Li et al., 2017), a hybrid of top-down statistical and bottom-up engineering approaches (Kavgic et al., 2010), can resolve this challenge.

According to Reinhart and Cerezo Davila (2016), UBEM will become an important planning tool for urban planners, energy utilities and other policy makers. Sola et al. (2018) highlight UBEM as an important component of urban-scale energy models (USEM) and provide a comprehensive review of modelling tools currently available for UBEM. Nageler et al. (2018a) cite physical and data-driven approaches as the two main paradigms for UBEM. Physical UBEMs are based on building energy simulations, and therefore have better potential for co-simulation. However, they require more input data, modelling efforts and computational power (Nageler et al., 2018b), which can be partly addressed through automation in the case of large-scale applications (Dogan and Reinhart, 2013). Data-driven UBEMs are based on statistical representation of the building energy performance obtained from measured data, and therefore are usually less complex. For instance, Swan and Ugursal (2009) argue that energy signatures (defined as reduced-order physics-based regression models) are a useful tool because they are simple, requiring only energy use data, and ensure comparability across large numbers of buildings. However, data-driven models alone can hardly be used to predict the effect of retrofitting measures (Y. Chen et al., 2017).

Building archetyping is the most widespread technique used in various UBEMs and provides a satisfactory compromise between accuracy and speed of simulation (Q. Li et al., 2015). This approach has been employed to analyse the current state of building stock and the aggregated impact of new energy efficiency policies and measures using regional and national bottom-up building stock models (Mata et al., 2014). Archetyping can also be used as a test-bed for retrofit scenarios in small urban districts (Sokol et al., 2017).

In this study, we developed a UBEM that combines energy signatures and building energy simulations. Calibrated engineering models combined with building energy simulation tools enable reliable simulations of various measures in order to evaluate their potential outcomes in terms of energy savings and emissions reductions (Fumo, 2014). As noted previously (Foucquier et al., 2013), this hybrid approach is "a nice trade-off between physical and machine-learning based methods". In model development we used DesignBuilder, an interactive interface for the energy simulation programme EnergyPlus that has been widely applied for modelling building heating, cooling, ventilation and other energy flows (Clément, 2012). Frayssinet et al. (2018) stress that the individual power demand of urban buildings depends on the diversity of occupant behaviours, rapid micro-meteorological phenomena and specific building characteristics. Therefore, the time resolution of the measured data should be at least as high as the time resolution of any analysis based on the UBEM (Sokol et al., 2017).

2.2. Stockholm case

Sweden has always been a pioneer in setting and achieving climate goals: reducing greenhouse gas (GHG) emissions by 40%, exceeding 50% share of renewables and improving the efficiency of energy use by 20% by 2020 (compared with 1990). It has set the overall goals of having net zero emissions of GHG by 2050 and becoming "one of the first fossil-free welfare states in the world" (Ministry of the Environment and Energy, 2015).

The City of Stockholm is even more ambitious, recently shifting its fossil-free goal from 2050 to 2040 (Stockholms stad, 2015). Heating and cooling of the building stock produces approximately 40% of total GHG emissions in Stockholm and thus this is a fundamental target for the City of Stockholm to decrease its energy use. Setting stricter requirements on new buildings is part of the strategy. However, since most of the city's building stock (64 million m2; Stockholms stad, 2019) already exists, in order to achieve its environmental goals the city intends to introduce energy efficiency measures (varying for different categories of buildings) that will achieve an average decrease of 30% in energy use for heating and cooling (Stockholms stad, 2014).

The Greater Stockholm area has an extensive district heating (DH) system that supplies half the overall heating demand in the region (see Fig. 1 for the demand structure in the City of Stockholm), providing around 12 TWh with a dimensioning load of 4.8 GW annually. AB Stockholm Exergi (co-owned by Fortum Group) is the largest actor, producing around 8 TWh annually (Levihn, 2017). The system is one of the most advanced multi-energy systems in the world and largely corresponds to "fourth-generation DH" (Lund et al., 2014).

The case of the City of Stockholm was chosen to test and refine the method developed in the present study, based on availability of data and interest among local stakeholders.

3. Methods

A city-wide building energy retrofitting modelling framework was developed based on the principle of the hybrid ('grey-box') bottom-up model, combining physical and statistical approaches to construct an integrated view of building heat energy consumption

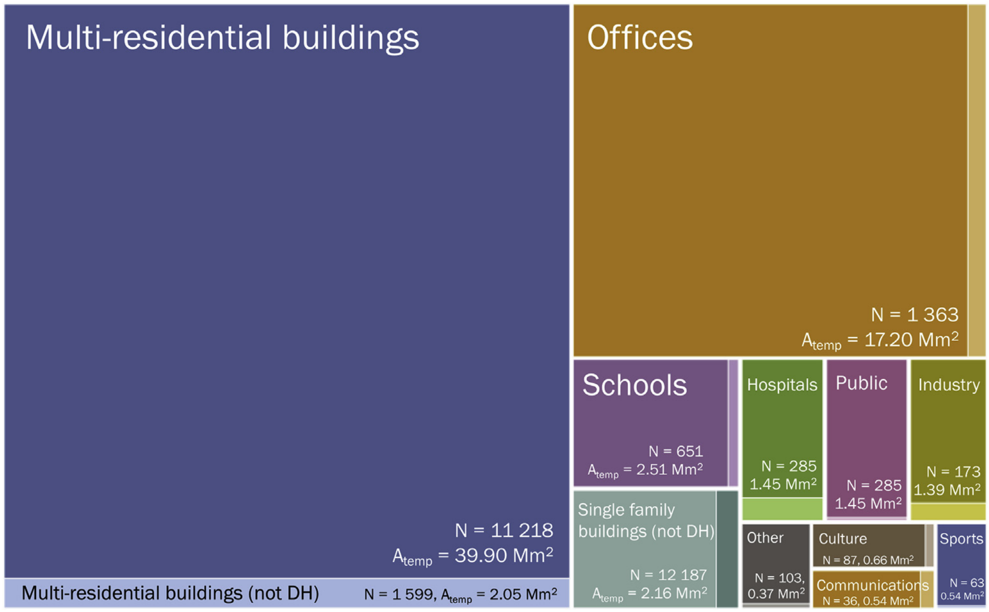

Fig. 1. Heat energy consumption by the City of Stockholm's building stock according to energy performance certificates (energy declarations). Block size corresponds to total heated area,[1] A_{temp}. Lighter-coloured bands beside/below each block indicate the heat energy demand met not by district heating (DH). Supporting data for this diagram are provided in Appendix A (Table A1).

(total heated area)[1] on city scale (Table 1). Coupled with emissions accounting and investment analysis modules, this approach enabled us to perform a comprehensive evaluation of the city-wide building energy retrofitting plans from both an exploration ('what if') and optimisation ('which') perspective.

Two key principles were formulated to guide framework development:

1. The workflow had to be arranged according to reproducible research principles (Peng, 2011), i.e. with all primary data kept unchanged, no side data transformations and all calculations being independently repeatable.
2. A modular structure (Pereverza et al., 2019) had to be used, to decrease interdependencies throughout the modelling framework and reduce the complexity of further upgrades in different stages.

3.1. Data import

The proposed framework utilises data about measured building heat energy use and data from energy performance certificates and reference and climate databases.

Measured data on heat energy use for the case of Stockholm were obtained from the main heat generation utility, AB Stockholm Exergi, and are thus limited to the building stock connected to the city's DH network, which accounts for approximately 85% of the total heated area. For initial framework development, data for one calendar year (2012), representing 15 068 district heating delivery points in the city of Stockholm with hourly precision, were provided.

Energy performance certificate (EPC) data were obtained from the national building energy declarations database maintained by the Swedish Board of Housing, Building and Planning (*Boverket*). The imported dataset represented energy declarations for 30 472 buildings in Stockholm, with 58.45 million m^2 total heated area, and contained information on building type, construction year, floor area, envelope form, energy use per source, energy system installations etc.

Reference data on standardised use and building envelope information were obtained from the Sveby project (Svebyprogrammet, 2012) and the book "*Så byggdes husen*" (Björk et al., 2013).

Climate data were extracted from the Swedish Meteorological and Hydrological Institute (SMHI) Open Data Catalogue (SMHI 2017). Hourly and daily ambient temperature measurements from SMHI's Observatorielunden station (climate number 98 210/98 230) were used.

Most of the data used were imported into PostgreSQL (PostgreSQL 2019), a widely used free, open-source RDBMS[2] well supported in various platforms, research and office software.

3.2. Data pre-processing

Most of the data pre-processing and analysis were performed in RStudio (RStudio Team, 2015), a free open-source IDE[3] for R, a programming language for statistical computing and graphics (R Core Team, 2013). R is one of the most popular languages in data science, as reflected in its well-developed ecosystem, i.e. various third-party extensions and a strong user community. In particular, it has good scalability potential for elastic and distributed cloud services using Hadoop/Spark, which is important for future development of the proposed framework.

To make the energy performance of buildings comparable, most of the calculations were made for specific units, so power demand was calculated for all measured values of actual energy use. Data tables were re-arranged according to the 'tidy data' paradigm (Wickham, 2014) to meet the objectives of the study. Data quality

[1] In the remainder of the paper we use the industry standard A_{temp}, which represents the area heated to above 10 °C.

[2] Relational database management system.

[3] Integrated development environment.

Table 1
The nine stages in the proposed modelling framework.

Stage	Summary
1. Data import	Data are imported from external data sources (data dumps from partner data warehouses, open data sources etc.) to a single database engine, which serves as a unified data source for the computation environment. General compliance of formats, scales and measurement units is established at this stage.
2. Data pre-processing	The collected data are curated, treating missing data and outliers[a]. Auxiliary parameters (e.g. specific energy consumption, aggregates per various grouping factors) are calculated. The dataset is restructured in accordance with the 'tidy data' paradigm (Wickham, 2014).
3. Segmentation	The entire building stock is divided into different building clusters corresponding to several archetypes. Stages 4–5 are performed for each building cluster separately.
4. Characterisation	Parametric models of building energy performance based on the energy signatures method are identified for each building through regression from the measured data.
5. Building energy simulations	An archetype building energy model is created based on the statistics for the whole building cluster. The model is then calibrated through matching energy signatures.
6. Scenarios	Retrofitting packages, target key performance indicators (KPIs) and available investment mechanisms are identified and translated into the scenarios to be analysed in the model. Scenarios can differ for different building clusters.
7. Aggregation	Building energy performance for each building in the entire building stock is projected from the results of the archetype building simulations for baseline and with possible retrofitting packages. Energy savings, emissions reductions and economic performance indicators for the proposed retrofitting measures are calculated and then aggregated for the city level.
8. Testing and optimisation	Sensitivity analysis is performed for the solutions identified. The optimisation problem is solved to meet the targets identified in Stage 6 with regard to the various optimisation objectives (e.g. capital investment required).
9. Communication	The potential of proposed retrofitting packages identified for the whole city building stock is communicated to the research community and project stakeholders through various visualisation tools. Feedback is also used to improve the workflow.

[a] Data points very different from most of the remaining data (Aggarwal, 2015).

assessment was an essential part of this stage, using the approach we describe in a previous paper (Pasichnyi et al., 2019b). Building-related information and building categorisation information were extracted from the specific energy use data as separate tables, while temperature data were joined to it. A normal year[4] for the period 1981–2010 was constructed to make normal-year correction according to Schulz (2003).

3.3. Segmentation

The building archetyping (segmentation and characterisation) was performed as described in our previous paper (Pasichnyi et al., 2019a). The analysed building stock was split into building clusters, i.e. subsets of buildings of the same kind, through filtering by nomenclature from the imported datasets. The clusters obtained were summarised through creation of 'archetype buildings', i.e. virtual buildings corresponding to the cluster centroids weighted by heated area A_{temp}. Virtual archetype buildings were then used for validation of the clusters through comparison with reference buildings (those in Björk et al., 2013; Svebyprogrammet, 2012).

3.4. Characterisation

Each building cluster was characterised from the measured heat energy use data for the ensemble of heating delivery points used to supply heat to the buildings in that cluster. All buildings at a heating delivery point i were characterised with an energy signature (ES) (see Fig. 2):

$$ES_i : MeteredHeatConsumption_i \rightarrow (\theta, c, \beta)_i \tag{1}$$

obtained from fitting the specific heat power $q(t)$ calculated from the *measured* data with the following quasilinear regression model:

[4] 'Normal year' is a Swedish analogue for typical meteorological year (TMY).

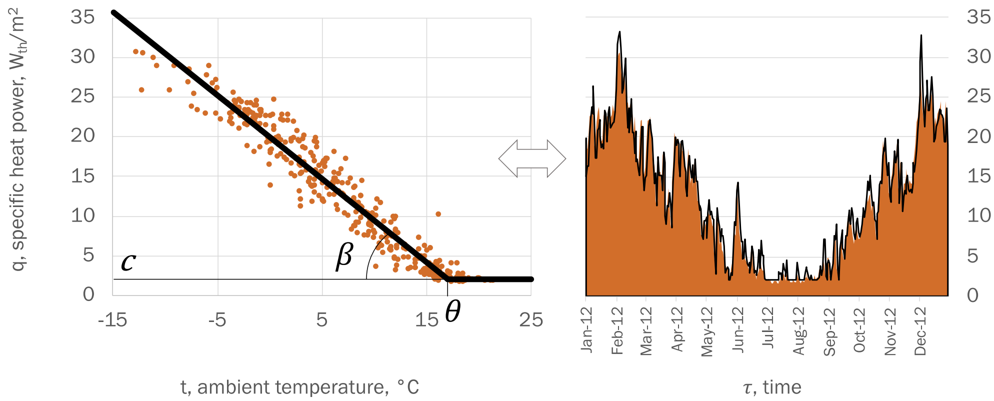

Fig. 2. Example of (left) the energy signature of a building $\boldsymbol{q}(\boldsymbol{t})$ and (right) specific heat energy use by time $\boldsymbol{q}(\boldsymbol{\tau})$. Measured and modelled values are coloured orange and black, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

$$q(t) = \begin{cases} c + \beta\,(\theta - t), t < \theta \\ c, t \geq \theta \end{cases} \tag{2}$$

where t is ambient temperature, θ is balance point temperature, c is base load (domestic hot water consumption) and β is energy performance coefficient.

Energy signature was selected for its simplicity and comparability across large numbers of dwellings (Swan and Ugursal, 2009). This model and its applications have been addressed in a number of studies, being first described 30 years ago (Hammarsten, 1987). It is still relevant for large-scale applications (e.g. Nageler et al., 2018a) and is being used as a European standard (EN 156036:2008).

The ES value for each archetype building was then obtained through weighted averaging of all energy signatures by heated area A_{temp} for the corresponding building cluster:

$$feature_{archetype_i} = \sum_{\forall building \in archetype_i} \frac{feature_{building}}{A_{temp_{building}}}, \quad i = 1..N \tag{3}$$

3.5. Building energy simulations

To obtain the baseline for archetype buildings used in the analysis, building energy models were constructed in the computer simulation tool DesignBuilder.[5] It was chosen because it is flexible and easy to use, due to a user-friendly graphical user interface. Moreover, DesignBuilder is based on the EnergyPlus[6] simulation engine, which allows for automated simulation, a process planned for future investigations.

Models of buildings allow evaluation of the energy performance for the current state and analysis of the effects of energy-saving measures on performance. Simulation provides the opportunity to get a picture of the building's energy performance and the influence of changes in any of the underlying systems, in a fast and cost-efficient way. The baseline models for specific archetype buildings were constructed using standardised occupant input data from the Sveby project (Svebyprogrammet, 2012) and building envelope information (Björk et al., 2013). The models were further validated and calibrated using measured data for the specific archetype analysed.

The information from the simulations was then used to investigate the impact of the energy efficiency measures analysed on all buildings in the specific building cluster. An individual energy signature for each of the building types considered was obtained through scaling back, as exemplified for the balance point temperature parameter:

$$\theta^*_{building} = \frac{\theta^*_{archetype_i}}{\theta_{archetype_i}} \cdot \theta_{building} \qquad \forall\, building \in archetype_i,\ i = 1..N, \tag{4}$$

where $\theta_{archetype_i}$ and $\theta_{building}$ are, respectively, baseline balance point temperature of the virtual archetype building and the modelled building; and $\theta^*_{archetype_i}$ and $\theta^*_{building}$ are, respectively, estimated balance point temperature of the virtual archetype building and the modelled building after retrofitting measures.

3.6. Scenarios

A scenario approach is widely applied to inform and support decision-making processes (Amer et al., 2013). In the present study, internal and external scenarios were considered. Internal scenarios represent the space of strategies (combinations of decisions) available to decision makers. External scenarios represent the space of futures (combinations of key uncertainties accompanied by existing trends) and were used for robustness analysis of the selected strategies. Depending on the needs of a particular case, scenario analysis can be conducted from explorative ('what if'), normative ('given the limitations') and/or optimisation ('the most efficient') perspectives. Thus, the required input for scenario construction can include: a) possible alternatives for exploration, b) normative constraints and c) targets for optimisation. The scenario space was generated here following the morphological approach (Pereverza et al., 2017).

Input alternatives represent possible internal changes (e.g. implementation of energy-saving measures in the particular subset of the building stock) and external options (e.g. discount rates applied for financing approved investment projects). Energy-saving measures were selected and tested one-by-one and were further combined into energy-saving packages.

National building regulations, accompanied by the thresholds representing the available financial mechanisms, constituted the set of constraints applied to the scenario space generated. As most of energy-saving measures analysed are not universal for different building clusters, auxiliary consistency constraints were set up.

Cost optimisation with respect to defined energy and climate

[5] https://www.designbuilder.co.uk.
[6] https://energyplus.net.

targets is a widespread practice for selection of the most optimal scenario (e.g. Yazdanie et al., 2017). However, our workflow allows different optimisation target functions to be applied for the scenarios analysed, as described further in subsection 3.8 'Testing and optimisation'.

3.7. Aggregation

Overall changes on the city level were determined by adding together the projections for individual building energy performance for the normal year.

3.7.1. Energy savings

Changes in total energy consumption for heating purposes in the baseline and in selected retrofitting scenarios were calculated through accounting for total annual energy consumption at each heating delivery point included. In building energy retrofitting, heat energy consumption generally undergoes a reduction, but for some retrofitting solutions (those that require additional use of electrical energy), a certain amount of additional electricity consumption may arise. Projections for total heat load demand were also used to investigate the effect of energy efficiency measures on DH supply (as proposed by Truong et al., 2018).

3.7.2. Emissions reduction

Estimated changes in total energy consumption were then used to account for the changes in emissions associated with the retrofitting scenarios. Emissions for projected energy demand were accounted for in two ways: a) with a physical process-based load duration optimisation model (Minerva) owned and operated by the DH utility; and b) by direct calculation using the emissions factors from the reference dataset (Gode et al., 2011). The Minerva model is logically similar to the EUROSPOT and Optima models used by Egeskog et al. (2009) and Levihn (2014), but essentially has higher detail and time resolution and allows all heating plants to be covered, accounting for limitations in distribution, maintenance shut-downs and boundary conditions for interactions between district cooling and district heating production in heat pumps. Using Minerva, changes in emissions of carbon dioxide (CO_2), particulate matter (PM_{10}), sulphur (S), nitrous oxide (NO_x) and mercury (Hg) were calculated. For comparative purposes, CO_2 emissions were also calculated using statistical emissions factors from local, residual and marginal perspectives (as in Levihn, 2014).

3.7.3. Investment analysis

Evaluating a potential investment decision is not easy, as different economic actors show different economic behaviour and preferences (Simon, 1959). One basic assumption is that profit generated tomorrow has less value than profit generated today (Levihn, 2016). This rests on the assumption that profit generated today might be used for generating more profit in the future or attract additional capital that might be used for generating future profit (Stiglitz, 1993). In this paper, the economic analysis and corresponding discussion were limited to the following four evaluation criteria:

i) Capital expenditure (CAPEX), or simply the sum of money invested I (equation (5)). It is important to consider CAPEX, as there may be various reasons to limit or expand the amount of fixed capital held by a firm.

$$CAPEX = \sum I \tag{5}$$

ii) Net present value (NPV), which takes into account the time-dependent aspects of cash flow. NPV is calculated by adding together the present value of all CAPEX, incomes and costs over the lifetime of the investment. NPV depends on the number of periods P (set at 25 years corresponding to Swedish accounting practices for building retrofits), discount rate r (discussed below), time period τ, and net cash flow at each period R_τ:

$$NPV(r, P) = \sum_{\tau=0}^{P} \frac{R_\tau}{(1+r)^\tau} \tag{6}$$

iii) Internal rate of return (IRR). While NPV calculates the total profitability of an investment, some actors prefer near-term generation of profit, or a profit less sensitive to risk. Optimising IRR meets this requirement and is calculated by solving for the value of r at which NPV equals 0:

$$IRR(r),\ NPV = 0 \tag{7}$$

iv) Present value quota (PVQ). By dividing equation (5) by equation (6), the PVQ is generated. It provides insights into how much profit is generated per unit capital invested. This metric bears some similarity to return index, which is often used when evaluating the profitability of public corporations, by basically being return index minus 1. PVQ is defined as:

$$PVQ = \frac{NPV}{CAPEX} \tag{8}$$

These four criteria provide different answers, but investment options that are better according to all four should be prioritised. These options would provide the largest and fastest generated profits with the least investment.

Thus, the first objective in city-scale analysis was screening for these economically preferable investments. However, most options can be expected to excel in one criterion, but not in the others. Thus, a discussion of how different investments correspond to different investor priorities is included in this paper.

The discount rate is always a matter of debate. Some building sector actors in Stockholm value formal demand on dividend on equity provided by shareholders and interest rate paid on debt. Other actors demand very fast payback, while some old housing cooperatives (BRF[7]) in central Stockholm with low or no debt, and with excess income from rented office spaces or stores, value capital as free. In the present case, rates of 4%, 8% and 20% were selected to represent different investment behaviours. Actors corresponding to these rates are explained and exemplified further:

- 4% — actors with easy access to capital and low demand for the profitability of investment, e.g. old housing associations with low debt and incomes from businesses on the ground floor of their buildings;
- 8% — actors with a demand for dividend on equity, some risk exposure and interest paid on debt, e.g. real estate companies;
- 20% — private actors seeking very fast payback on their investment, e.g. owners of individual houses who want quick generation of positive cash flow. This is a less normative discount rate, describing behaviours rather than intention and thus implicit discount rates. Previous research has shown that actual behaviours of individuals correspond to implicit discount rates sometimes above 60% (Meier and Whittier, 1983). Jaffe et al.

[7] BRF stands for '*bostadsrättsforening*' (housing association).

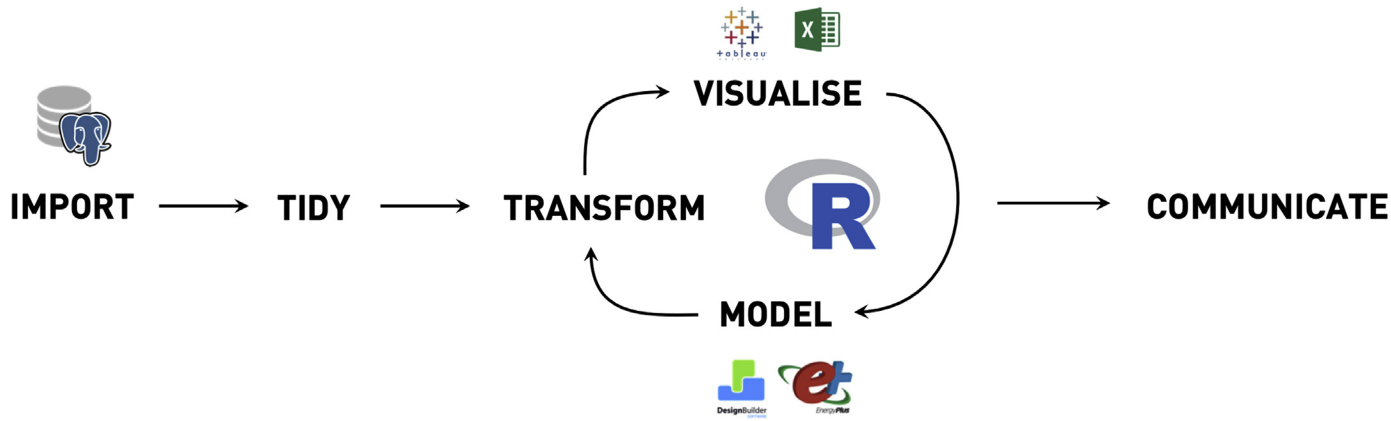

Fig. 3. Data analysis workflow. (Adapted from Wickham and Grolemund, 2017).

(2004) provide an empirical interval of 0% to over 100% for implicit discount rates and use 20% in their examples.

3.8. Testing and optimisation

The retrofitting scenarios were then analysed using the following two approaches:

1. Sensitivity analysis. Sensitivity analysis techniques were applied to improve understanding of the uncertainty embodied in the model and to perform robustness testing of the scenarios considered.

2. Mathematical optimisation. The set of retrofitting scenarios was used as a domain for optimisation of the target function. An optimisation problem can accordingly be treated either as (a) a linear programming (LP) problem with one of the initial targets (energy saving, emission reduction, social investment costs) as an objective function and the remaining targets as constraints; or (b) a multi-objective optimisation problem for all target functions considered.

3.9. Communication

Various data visualisation techniques and reference indicators were used to: a) gain a better understanding of imported datasets and guide the data curation process; b) test various hypotheses and techniques; c) check the accuracy of calculations at each stage; and d) communicate results to external actors. Communication and discussion of the preliminary results within the research team and with project partners were used to tune the workflow set-up and validate results from early stages of the study.

Data visualisation was conducted partly in R for quick hands-on exploration by the research team, while Tableau (Tableau Software, 2019) or MS Excel chart engine was used for further communication to external collaborators. Tableau was found to be more relevant for visualising larger datasets and for live demonstrations, through additionally created interactive data products.

3.10. Modelling process

The interconnection of all tools used in the modelling framework during the data analysis workflow is illustrated in Fig. 3. While Table 1 explained the logic of the framework, which is quite linear and modular, Fig. 3 illustrates the process, which is largely iterative.

Two core stage sequences in Table 1 correspond to the iterative part of Fig. 3: a) computing in stages 3–6 (segmentation, characterisation, building energy simulations and scenarios), where the energy hybrid statistical-physical model is set up and calibrated using the tidied data input; and b) analysis in stages 7–8 (aggregation, testing and optimisation), where the input scenarios are investigated using the city-level aggregates for energy, emissions and costs.

4. Results

This section presents the main results from development and application of the framework to the case of Stockholm.

1. **Data import.** Measured heat energy use data were provided in the form of the MS SQL Server database dump. The data were imported into PostgreSQL and then normalised, splitting data on meterings, heat delivery points and heat delivery points categories into separate tables. The EPC data were provided in the form of a text data file that was read in directly from the RStudio. The reference data were imported manually at Stage 5 ("Building energy simulations"). The climate data were imported into PostgreSQL.
2. **Data pre-processing.** Missing values were treated in two ways. In cases where imputation was easy to apply and did not pose a risk of distribution distortion (e.g. a single missing point in temperature time series), mean/linear imputation was applied and the sample was retained for further analysis. In all other cases the complete sample approach was applied, excluding corresponding objects from further investigation.

Outliers were treated in a similar way as missing values. Extreme values and anomalies were first identified with basic physical (e.g. *HeatEnergy* $\geq$ 0 for DH consumers) or statistical (e.g. $x \in [-3\sigma; 3\sigma]$) thresholds. Some additional errors were identified through checking the energy balances for each building. Errors were then corrected to the closest marginal values where applicable, or otherwise discarded.

3. **Segmentation.** Three building archetypes that we developed and described in a previous study (Pasichnyi et al., 2019a) were used to analyse building retrofitting potential. To exemplify the strategic planning framework developed in this study, a single homogeneous cluster of multi-family residential buildings constructed in the period 1946–1975 was used for further analysis. These buildings were largely constructed within the

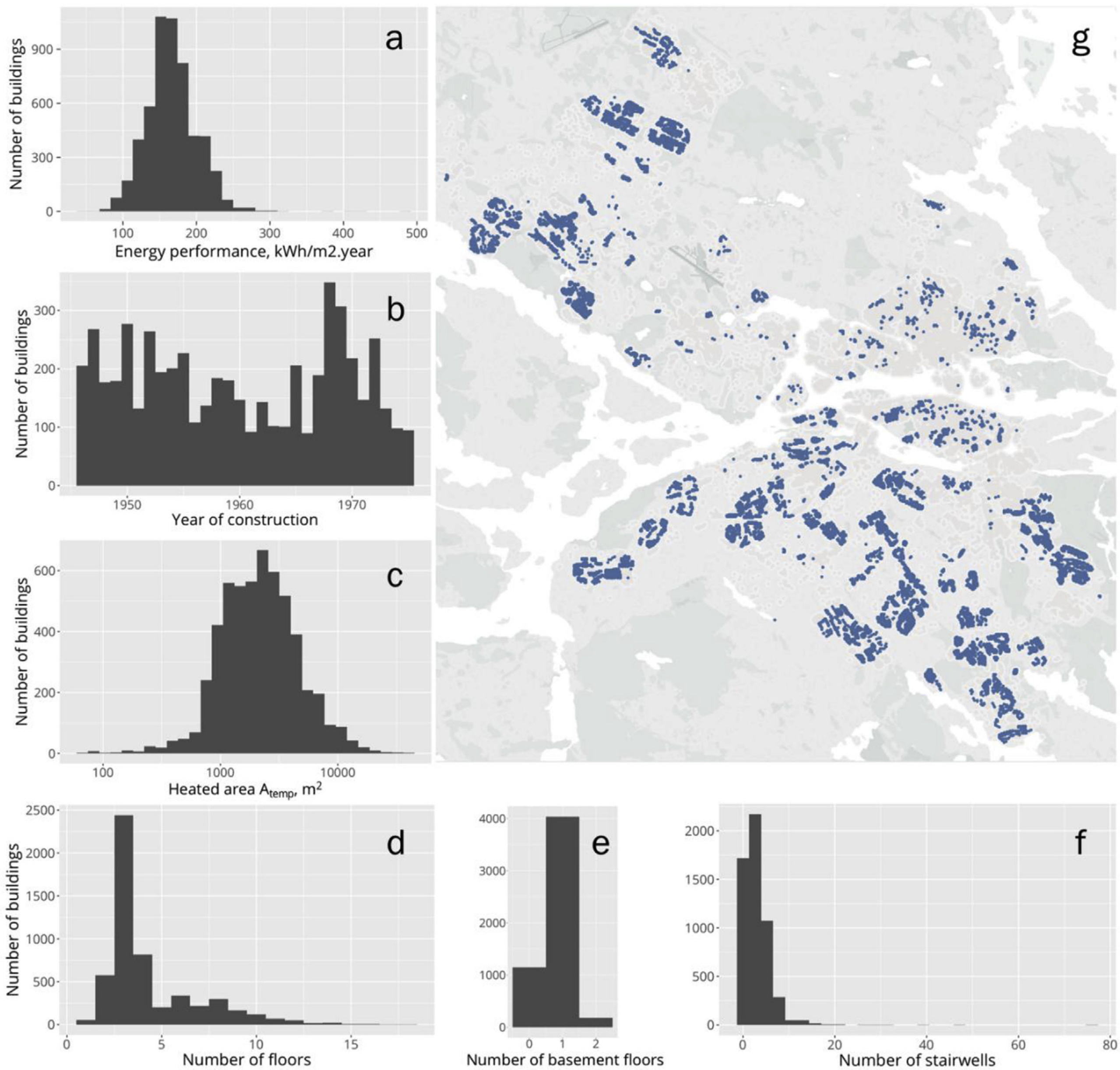

Fig. 4. Distribution of features for the archetype 1 (multi-family residential, 1946–1975) building cluster: a) energy performance, kWh/m^2. year; b) year of construction; c) heated area $\boldsymbol{A_{temp}}$; d) number of floors; e) number of basement floors; f) number of stairwells; g) spatial distribution.

'Million Programme' and are the most widespread type in Stockholm, representing 26% of the total and 33% of the residential building stock in the city (Fig. 4). The building subset considered can be split into a number of subsets, e.g. by construction age or number of floors (as proposed by (Q. Wang and Holmberg, 2015). The total number of buildings was 5400, the number of DH heat delivery points was 2402 and the total heated area A_{temp} was 14.18 million m^2.

The characteristics of the virtual archetype building for the constructed cluster were obtained through weighted averaging of the features of all buildings in the cluster by the heated area, while its heated area was set to mean $A_{temp} = 2876\,\text{m}^2$.

4. **Characterisation.** Each heat delivery point was characterised with the energy signature (ES) model derived from fitting measurements of the actual heat energy consumption (Fig. 5). The energy signature method proved useful in this case, as it described the energy performance of the investigated building type reflecting the characteristics of the building set-up, with or without imposed energy-saving measures.
5. **Building energy simulations.** Based on the descriptive statistics from the segmentation stage and the input reference data, a base case building energy simulation model for the virtual archetype building was created. It was then calibrated to correspond to the aggregated ES from the characterisation stage (Fig. 6).
6. **Scenarios.** Three retrofitting packages were selected for analysis in this study. The first comprised installation of heat recovery ventilation, the second installation of energy-efficient windows and the third installation of both heat recovery ventilation and energy-efficient windows. According to Brown et al. (2014), 83%

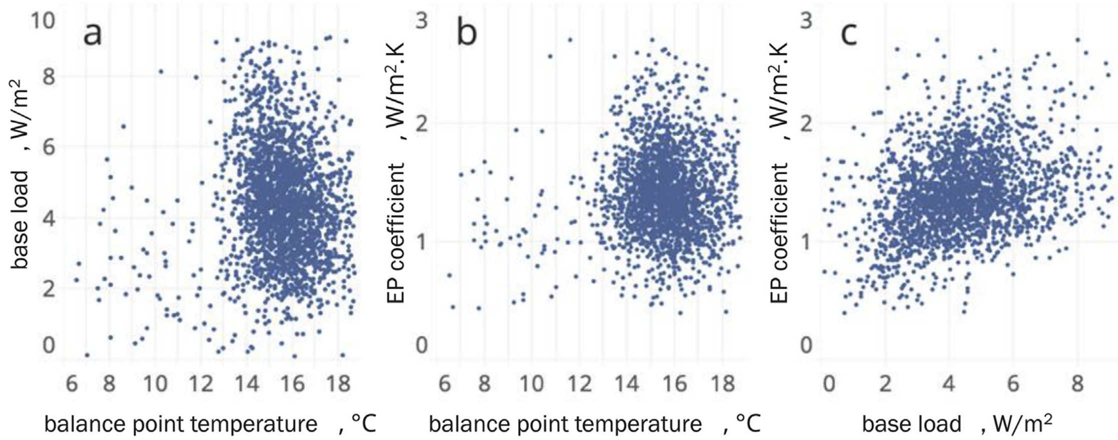

Fig. 5. Distribution of parameters in the characterisation models (each dot represents a model): a) balance point temperature θ vs base load c; b) balance point temperature θ vs energy performance (EP) coefficient β; c) base load c vs energy performance coefficient β.

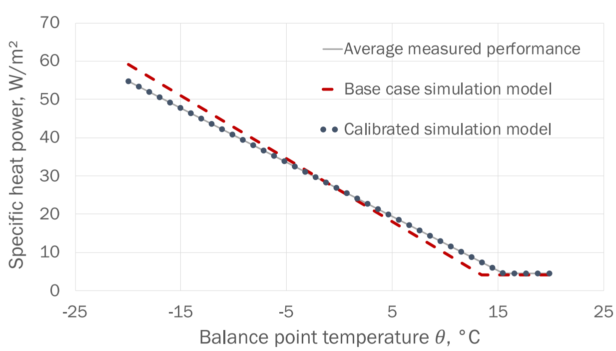

Fig. 6. Energy signatures for the virtual archetype 1 (multi-family residential, 1946–1975) building.

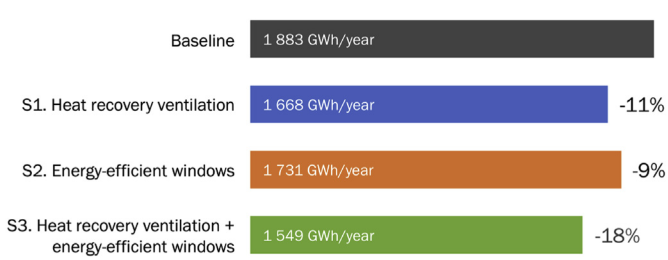

Fig. 7. Energy-saving potential of the archetype 1 (multi-family residential, 1946–1975) buildings for the baseline and three selected retrofitting packages.

of the calculated stockwide embodied global warming potential for refurbishment measures arises from only two types of measures – ventilation and windows.

The first selected measure chosen was exhaust air heat recovery (scenario S1). The reason for choosing this measure was that 93% of the buildings in this particular subset are fitted with exhaust air-type ventilation without heat recovery and heat losses from ventilation are usually significant. Besides, this measure is often implemented by professional building owners and is considered profitable. The heat recovery system implemented in the simulation model was designed to have a temperature efficiency of 50%.

The second measure selected was changing windows (S2). It is known to be an energy-saving measure that is not economically justified based on energy savings. In the model, the thermal transmittance (U-value) of the windows was changed from 2.7 to 0.8 W/m^2.K.

The third package (S3) was intended to show how combining measures affected the cumulative energy performance and influenced the economics of building retrofitting.

Information regarding costs for implementation of the different energy-saving measures was taken from the renovation catalogue Repab (Incit AB, 2014) and adapted to different building sizes.

7. **Aggregation.** Energy. The results of building energy simulations for the virtual archetype building with implementation of selected retrofitting packages were scaled out to the entire archetype building cluster, to obtain individual projections of the energy performance for all metering points considered. A summary of the heat energy consumption by the building cluster analysed (multi-family residential buildings constructed in 1946–1975) in different selected scenarios is provided in Fig. 7. Installation of heat recovery ventilation (S1) had greater energy-saving potential than switching to energy-efficient windows (S2). Implementation of the combined retrofitting package (S3) reduced the heat energy demand by 18%, or 334 GWh per year.

The effect of large-scale retrofitting on heat supply was examined using the Minerva model to calculate the resulting reduction in supplied energy corresponding to the changed demand aggregated per hour of the normal year. Fig. 8 exemplifies the change in demand before and after retrofitting for the combined package (S3). The resulting reduction in supplied energy due to the changed demand is the energy depicted between the two dotted lines. As can be seen from Fig. 8, retrofits do not affect the performance of buildings uniformly over the year, e.g. no or less heat is required outside the heating season combined with the different effects of different retrofitting options. Hence different heat production processes are affected differently, with no effect on waste incineration combined heat and power (CHP) plants and a larger effect on heat pumps and bio-oil utilisation.

Emissions. Hourly projections on changes in the energy demand were then used to account for associated emissions changes for each of the retrofitting strategies (S1–S3). Using the Minerva model, local emissions changes for the year 2025, with forthcoming changes in DH production structure included, were computed (Fig. 9a). The estimates for CO_2 emissions changes obtained using the Minerva model (2025) and current (2016) emissions factors for

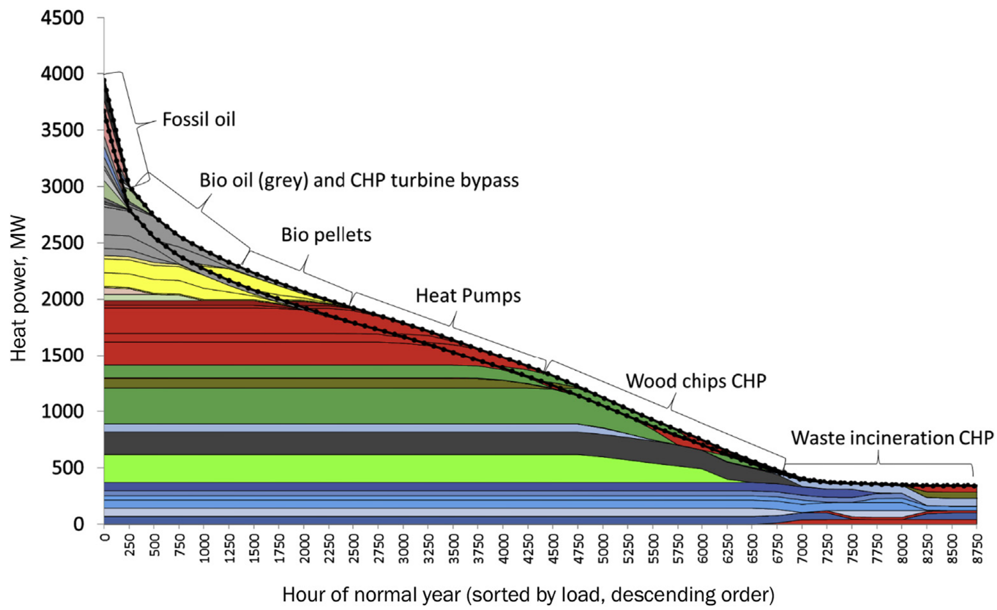

Fig. 8. Annual load duration for Stockholm district heating (DH) production before and after the implementation of the combined retrofitting package (S3). The change in demand before and after retrofitting is the area between the two dotted lines.

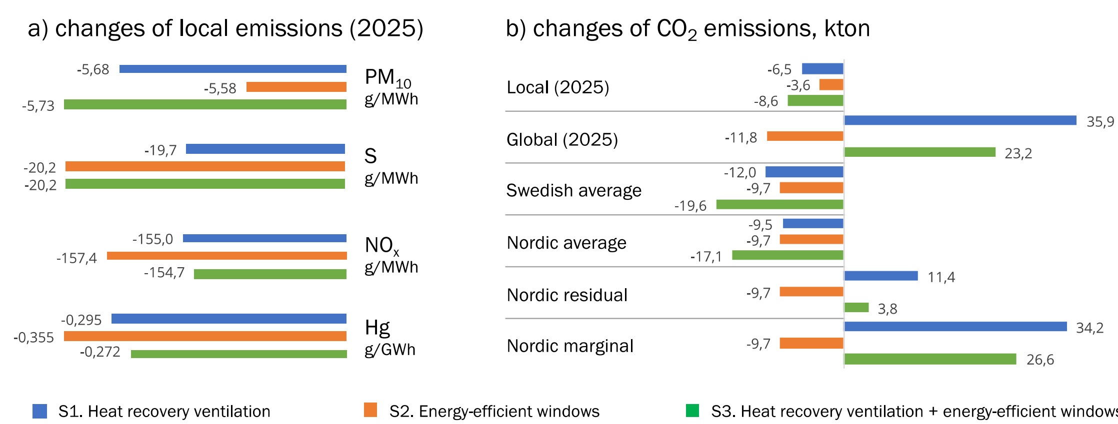

Fig. 9. Emissions changes for the three retrofitting packages (S1–S3). a) Projections for 2025 obtained using Minerva, a physical model of district heating (DH) production after phase-out of coal-based combined heat and power in 2022 and introduction of new production facilities, and b) comparison of estimated changes in CO_2 emissions in 2025 from Minerva (local and global) and the current system (2016), obtained using emissions factors for the various allocation schemes.

various allocation schemes are compared in Fig. 9b. For heat production, energy mixes for DH Stockholm and Sweden were considered, while for electricity production Swedish average, Nordic average, Nordic residual and Nordic marginal data were used.

It was found that while the installation of heat recovery ventilation had a positive environmental effect from an 'accounting' perspective (local energy mixes), the effect from a 'long-term' marginal perspective was negative for both 2016 and 2025. In other words, if the electrical system is viewed as more CO_2-intensive, adopting this option could actually result in increased CO_2 emissions. This can be explained by the substitution effect of this measure when the reduction in heat energy use is partly compensated for by increased electric energy use and by the fact that the environmental profile of heat production by local DH is better than that of Nordic electricity production. At the same time, the environmental effect of energy-efficient windows was found to be positive in all cases.

Economics. A general picture of the investments associated with the different retrofitting options (S1–S3) is provided in Fig. 10. Overall, installation of heat recovery ventilation (S1) proved to be economically viable (NPV>0, IRR>15%, $r = 4\%$). This resulted in a competitive reduction of 2.85 MW through retrofitting buildings from 493 (21%) metering points with heat recovery.

For energy-efficient windows (S2) and for combining this option with heat recovery (S3), it was difficult to achieve a positive economic outcome for the vast majority of the building stock. As an example, even at a reduced discount rate and demand for IRR in S3 (NPV>0, IRR>0%, $r = 0\%$), buildings from only 7 (0.3%) metering points met the investment criteria, resulting in only 86.3 kW saved.

For investors prioritising multiple investments, the generally higher PVQ for heat recovery ventilation measures (S1) shows that multiple investments in ventilation should be preferred over investments in energy-efficient windows (S2) or a combined package (S3). It should be noted, however, that the effect on property values, which could tilt the result, was not included in the analysis.

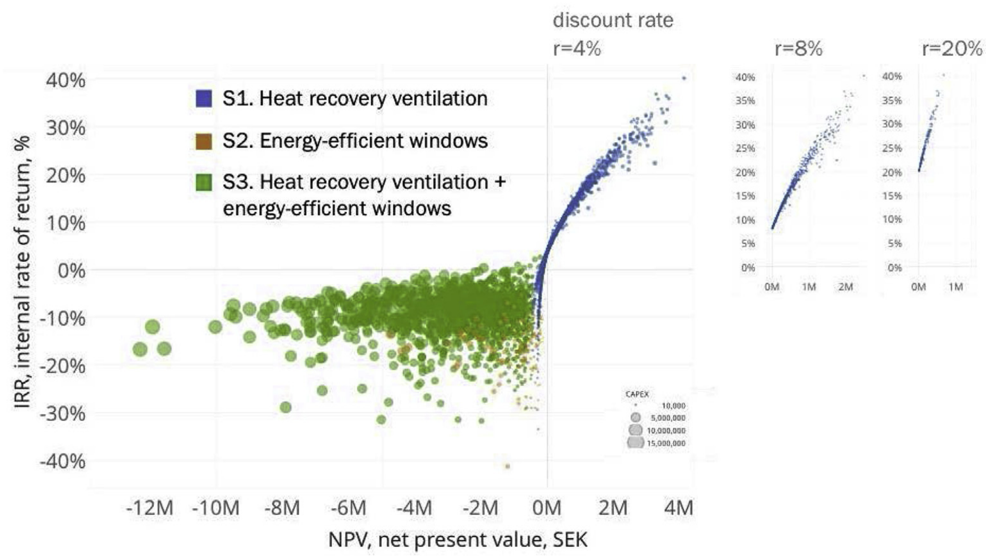

Fig. 10. Investment indicators for the three retrofitting packages considered (S1–S3). Sensitivity analysis of the investments for different discount rates (r = 4%, 8% and 20%) is provided for NPV>0 and IRR>0.

8. **Testing and optimisation.** Sensitivity analysis was performed for the discount rates used for making economic projections (Fig. 10). As the main goal of the example case study was to provide the accounting function of the framework to analyse different 'what-if' scenarios, development of an optimisation module that allows fitting portfolios of buildings and measures to various targets is planned for future work.
9. **Communication.** Communication of the modelling results was organised in three tiers: i) project team of researchers; ii) project representatives of non-academic stakeholders and associated researchers; and iii) a broad academic audience. Tier (i) was characterised by high intensity, short frequency, large amount and low adaptation of exported information. Tiers (ii) and (iii) were characterised by low intensity, long frequency, small amount and high adaptation of exported information.

5. Discussion

The framework developed in this study allows the effects of applying various retrofitting options to a large part of the urban building stock to be analysed. Considering the available scenarios from multiple perspectives (energy, environment, economic), as done here for the empirical case of Stockholm, makes it possible to identify the benefits and drawbacks of each retrofitting measure and to make strategic decisions with information about related trade-offs, therefore reducing the 'silo effect' identified previously (Brooks et al., 2014). For example, installation of energy-efficient windows (S2) had good energy reduction potential and a positive environmental effect, but could not compete with heat recovery ventilation (S1) in terms of investment attractiveness and economic viability in general.

The level of detail provided by the framework means that it can be applied for development of customised strategies for large management institutions targeting retrofitting investment options that meet certain criteria, e.g. meeting particular targets for reduction of final heat energy use and CO_2 emissions with a fixed amount of capital investment.

The framework's architecture and the underlying data analysis methods ensure good scalability. This allows for analysis of multiple building archetypes and scenarios in the case example of Stockholm. After setting the core framework, building energy simulation is the only effort-demanding stage for such an extension. Furthermore, the modularity principle applied allows for flexibility in further development of the framework, as adding or modifying particular modules is associated with low efforts required to enable their compliance with the rest of the framework.

It is worth pointing out that, in development of modelling frameworks, there is a permanent trade-off between further model improvement and associated gains in requirements and applicability. Moreover, the complexity related to refinement of such frameworks is largely related to organisational issues of data availability, quality and consistency, rather than the methodology per se.

The use of well-known and validated simulation software for building energy simulations limits the uncertainty factors to the input parameters. The input parameters should be transparent and well established. In the present study the input parameters were taken from the Sveby framework for building energy simulation (Svebyprogrammet, 2012). The output data from the simulation model were then validated against measured data to ensure reliable output from the model.

The possibility of large-scale energy retrofitting of 'Million Programme' buildings in Sweden has been frequently discussed (Johansson et al., 2017; Mangold et al., 2016). In the context of Stockholm, this study showed that wide-scale implementation of only two measures (heat recovery ventilation and energy-efficient windows) would reduce final energy use by 334 GWh, which is only ~3% of total city heat energy demand. However, while the positive environmental effects and investment attractiveness of this combined package were not very impressive in the case of Stockholm, the outcome would depend on local contexts, e.g. the environmental effect could be greater in European cities with less efficient, fossil-based heating systems. This finding aligns with SR15 conclusions (IPCC, 2018) that a combination of energy efficiency measures and less carbon-intensive supply is needed to meet stringent climate mitigation goals. We demonstrated the possibility of applying different perspectives in accounting of GHG emissions and showed how this can significantly affect the key environmental performance indicator of CO_2 emissions for a particular set of measures. The results also illustrated the short- and long-term effects of large-scale retrofitting on the heat supply, information that can be used to analyse the interplay of supply and demand in urban energy strategies, as proposed elsewhere (Lundström and Wallin, 2016).

The proposed method can also be applied in other geographical locations and contexts, although at the moment the availability and quality of input data might be considered a bottleneck for its wider utilisation, which is a frequently cited limitation with data-driven models (e.g. Nageler et al., 2018a). However, the recent trend for rapid urban digitalisation indicates positive prospects for modelling frameworks of this type (Liu et al., 2014).

A significant advantage in this study was the great availability of metered data of high resolution, which allowed us to construct a fact-based model of the energy performance for the building stock analysed. However, using datasets of high volume and veracity required advanced analytical approaches for data cleaning, multithread computations etc. Use of multiple data sources also imposed a number of consistency problems to be solved (linking metered and EPC datasets, matching similar features etc.), but it allowed us to perform a number of cross-validation checks during the whole modelling process, as described in our previous paper (Pasichnyi et al., 2019b).

The study also confirmed the growing value of urban energy data and consequent emergence of new functions of urban energy utilities such as data collection and provision. The current extensive digitalisation initiatives provide new possibilities to combine information from multiple big data sources in large-scale identification of tailored measures to increase the energy performance of individual buildings. Hence cities with more centralised energy systems and homogeneous energy datasets can be expected to have an additional advantage in developing a fact-based understanding of urban building energy use and more precisely targeting policies related to building energy efficiency. At the same time, possible conflicts of interest between data providers (energy utilities) and data users (city authorities and building owners) pose a higher risk of lock-ins that would require additional efforts in finding paths for 'win-win' data collaboration between all stakeholders involved. Hence, following the definitions from Späth and Rohracher (2015), energy data can be treated as another urban junction emerging at the moment.

Limitations: The method presented has some limitations that should be acknowledged. Due to the availability of high-resolution metering data, it was possible to obtain a precise individual characterisation model for each building analysed. However, the resulting models were affected by the behaviour of the single virtual archetype building, an effect which was only partly compensated for through proportional scaling of effects for the initial energy signature parameters of each building. As the fundamental object for metering factual heat energy use by the DH utility is not stand-alone physical buildings, but metering points, this resulted in a lower level of resolution for the metered dataset and consequently led to the necessity for uniform distribution of results obtained for separate metering points to corresponding buildings. However, in the vast majority of cases each metering point had connections from buildings of the same type, so this did not have a significant influence on the accuracy of the model obtained. While residents' behaviour was not explicitly distinguished as one of the factors affecting cumulative building heat energy demand, it was embedded in the actual energy use meter data. Hence residents' behaviour was accounted for through calibration of the physical building energy simulation model with statistical characterisation models and standardised occupant input data. However, issues such as social acceptance of retrofitting and possible rebound effects for energy saved from retrofitting were beyond the scope of this study.

Future work: In future work, we plan to explore the following refinements to the method. First, there is potential for reducing exclusion of buildings at the data pre-processing stage by applying more advanced techniques for treating missing values (Graham, 2009). Second, a shift from nomenclature-based to data-driven segmentation should allow more relevant virtual archetype buildings to be constructed and should provide statistical quality control on the information reduction happening in the model transition from the 'single building' to the 'archetype/set of buildings' modes and back. Third, the building energy simulation stage has good potential for automation and improvement of model resolution through accounting for individual building geometry for each processed building, as proposed recently (Cerezo Davila et al., 2016). Fourth, for such sparse urban areas as Greater Stockholm, more advanced urban climate modelling (as presented e.g. by (Yang and L. Chen, 2016)) could benefit the overall accuracy of the model. Fifth, developing an even more holistic view on the retrofitting process from an industrial ecology perspective could be achieved as proposed by Brown et al. (2014) by applying a life cycle perspective and including in the analysis the building renovation process itself, with its embodied energy use and related GHG emissions.

6. Conclusions

The significant advantage of the proposed novel data-driven urban building energy model (UBEM)-based approach is the use of high-resolution metered data in fact-based modelling of the energy performance of the building stock. This paper contributes to the literature on city-scale building energy modelling by introducing the use of large datasets that have become available recently, particularly high-resolution metered data on heat energy use.

A new method for strategic planning of building energy retrofitting based on a city-wide building energy modelling framework was developed. This method makes it possible to assess the change in total energy demand from large-scale retrofitting and explore its impact on the supply side. The method uses high-resolution actual energy use datasets and advanced analytics techniques to construct an UBEM of high detail and to identify buildings and retrofitting measures that have the highest potential. This can improve the quality of decision making through analysis of city energy strategies from multiple perspectives.

The method was empirically tested in the case of retrofitting 'Million Programme' buildings in Stockholm with heat recovery ventilation and energy-efficient windows. The results revealed that large-scale introduction of these two retrofitting measures would: a) reduce the annual heat energy demand of the building stock in question by 18%; b) have variable impacts on the environmental performance of urban energy systems; and c) be only partly economically viable. These outcomes are partly explained by availability of an advanced DH system in Greater Stockholm, where the supply of energy has largely been decarbonised. The results were further used in the City of Stockholm's ongoing work on its climate strategy. The proposed method can be applied in different geographical locations and contexts, depending on sufficient availability and quality of input data. The study exemplified the effects of ongoing digitalisation of the energy sector in cities and demonstrated the growing value of urban energy data, particularly for urban strategic planning.

Acknowledgements

This research was supported by the E2B2 research programme run by IQ Samhällsbyggnad and funded by the Swedish Energy Agency (projects No. 40846-1 and STEM 2018-006282). The study was performed through collaboration between the authors, the City of Stockholm, Hammar by Sjöstad Organisation of Housing Associations (Sjöstadsforeningen), Lassila & Tikanoja, the Association of Property Owners (Fastighetsägarna) and AB Stockholm Exergi.

Appendix A

Table A1
Structure of Stockholm building stock by building usage types and connection to district heating (DH) according to energy performance certificates data

	Total		Connected to DH		
Building usage type	N, number of buildings	Heated area A_{temp}, m^2	N, number of buildings	Heated area A_{temp}, m^2	% of total area
Multi-residential buildings	12 817	41 951 339	11 218	39 898 962	95%
Offices	1645	17 984 850	1363	17 201 730	96%
Schools	867	2 674 287	651	2 514 371	94%
Single family houses	14 452	2 484 998	2265	323 943	13%
Hospitals	638	1 692 365	285	1 454 459	86%
Public	177	1 677 425	124	1 644 927	98%
Industry	231	1 562 831	173	1 391 272	89%
Other	140	759 025	103	722 429	95%
Culture	122	709 433	87	660 366	93%
Communications	72	611 748	36	543 211	89%
Sports	91	551 927	63	540 070	98%

References

Aggarwal, C.C., 2015. Outlier analysis. In: Data Mining. Springer, Cham, pp. 237–263. https://doi.org/10.1007/978-3-319-47578-3.

Amer, M., Daim, T.U., Jetter, A., 2013. A review of scenario planning. Futures 46, 23–40.

Barahona, J.C., Pentland, A.S., 2007. Advice networks and local diffusion of technological innovations. In: Steinfield, C., Pentland, B.T., Ackerman, M., Contractor, N. (Eds.), Communities and Technologies 2007. Springer London, London, pp. 509–529. https://doi.org/10.1007/978-1-84628-905-7_25.

Beckel, C., Sadamori, L., Staake, T., Santini, S., 2014. Revealing household characteristics from smart meter data. Energy 78, 397–410. https://doi.org/10.1016/j.energy.2014.10.025.

Bertone, E., Stewart, R.A., Sahin, O., Alam, M., Zou, P.X.W., Buntine, C., Marshall, C., 2018. Guidelines, barriers and strategies for energy and water retrofits of public buildings. J. Clean. Prod. 174, 1064–1078. https://doi.org/10.1016/j.jclepro.2017.11.065.

Björk, C., Kallstenius, P., Reppen, L., 2013. Så Byggdes Husen 1880-2000. Svensk Byggtjänst.

Brooks, E., Law, A., Huang, L., 2014. A comparative analysis of retrofitting historic buildings for energy efficiency in the UK and China. disP - The Planning Review 50, 66–75. https://doi.org/10.1080/02513625.2014.979044.

Brown, N.W.O., Olsson, S., Malmqvist, T., 2014. Embodied Greenhouse Gas Emissions from Refurbishment of Residential Building Stock to Achieve a 50% Operational Energy Reduction, vol. 79, pp. 46–56. https://doi.org/10.1016/j.buildenv.2014.04.018.

Cerezo Davila, C., Reinhart, C.F., Bemis, J.L., 2016. Modeling Boston: a workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets. Energy 117, 237–250. https://doi.org/10.1016/j.energy.2016.10.057.

Chen, Y., Hong, T., Piette, M.A., 2017. Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. Appl. Energy 205, 323–335. https://doi.org/10.1016/j.apenergy.2017.07.128.

Clément, P.F., 2012. Building Energy Retrofitting: from Energy Audit to Renovation Proposals: the Case of an Office Building in France.

Deakin, M., Reid, A., 2018. Smart cities: under-gridding the sustainability of city-districts as energy efficient-low carbon zones. J. Clean. Prod. 173, 39–48. https://doi.org/10.1016/j.jclepro.2016.12.054.

Dixon, T., Lannon, S., Eames, M., 2018. Reflections on disruptive energy innovation in urban retrofitting: methodology, practice and policy. Energy Research & Social Science 37, 255–259. https://doi.org/10.1016/j.erss.2017.10.009.

Dogan, T., Reinhart, C., 2013. Automated conversion of architectural massing models into thermal "shoebox"models. In: Presented at the Proceedings of BS2013.

Eames, M., Dixon, T., Lannon, S., Hunt, M., De Laurentis, C., Marvin, S., Guthrie, P., Georgiadou, M.C., 2014. Retrofit 2050: Critical Challenges for Urban Transitions. Cardiff University.

Egeskog, A., Hansson, J., Berndes, G., Werner, S., 2009. Co-generation of biofuels for transportation and heat for district heating systems—an assessment of the national possibilities in the EU. Energy Policy 37, 5260–5272.

Engström, R., Howells, M., Mörtberg, U., Destouni, G., 2018. Multi-functionality of nature-based and other urban sustainability solutions: New York City study. Land Degrad. Dev. 29, 3653–3662. https://doi.org/10.1002/ldr.3113.

EU Commission, 2016. An EU strategy on heating and cooling [WWW document]. ec.europa.eu. https://ec.europa.eu/energy/sites/ener/files/documents/1_EN_ACT_part1_v14.pdf. accessed 2.6.18.

Foucquier, A., Robert, S., Suard, F., Stéphan, L., Jay, A., 2013. State of the art in building modelling and energy performances prediction. Review 23, 272–288. https://doi.org/10.1016/j.rser.2013.03.004.

Frayssinet, L., Merlier, L., Kuznik, F., Hubert, J.-L., Milliez, M., Roux, J.-J., 2018. Modeling the heating and cooling energy demand of urban buildings at city scale. Renew. Sustain. Energy Rev. 81, 2318–2327. https://doi.org/10.1016/j.rser.2017.06.040.

Fróes Lima, C.A., Portillo Navas, J.R., 2012. Smart metering and systems to support a conscious use of water and electricity. Energy 45, 528–540. https://doi.org/10.1016/j.energy.2012.02.033.

Fumo, N., 2014. A review on the basics of building energy estimation. Renew. Sustain. Energy Rev. 31, 53–60. https://doi.org/10.1016/j.rser.2013.11.040.

Geyer, P., Schlüter, A., Cisar, S., 2017. Application of clustering for the development of retrofit strategies for large building stocks. Adv. Eng. Inf. 31, 32–47. https://doi.org/10.1016/j.aei.2016.02.001.

Gode, J., Martinsson, F., Hagberg, L., Öman, A., Höglund, J., Palm, D., 2011. Miljöfaktaboken 2011. Uppskattade Emissionsfaktorer För Bränslen, El, Värme Och Transporter. Värmeforsk Service AB.

Gouveia, J.P., Seixas, J., Mestre, A., 2017. Daily electricity consumption profiles from smart meters - proxies of behavior for space heating and cooling. Energy 141, 108–122. https://doi.org/10.1016/j.energy.2017.09.049.

Graham, J.W., 2009. Missing data analysis: Making it work in the real world. Annu. Rev. Psychol. 60, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530.

Gupta, R., Gregg, M., 2018. Targeting and modelling urban energy retrofits using a city-scale energy mapping approach. J. Clean. Prod. 174, 401–412. https://doi.org/10.1016/j.jclepro.2017.10.262.

Hammarsten, S., 1987. A critical appraisal of energy-signature models 26, 97–110. https://doi.org/10.1016/0306-2619(87)90012-2.

Huebner, G., Shipworth, D., Hamilton, I., Chalabi, Z., Oreszczyn, T., 2016. Understanding Electricity Consumption: A Comparative Contribution of Building Factors, Socio-Demographics, Appliances, Behaviours and Attitudes. Elsevier, pp. 692–702. https://doi.org/10.1016/j.apenergy.2016.04.075.

IEA, 2017a. World Energy Outlook 2017. OECD/IEA, Paris. https://doi.org/10.1787/weo-2017-en.

IEA, 2017b. Buildings Energy Efficiency Policies Database.

Incit AB, 2014. Repab Underhållskostnadsbok 2015. Sweden.

IPCC, 2018. Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty (in press).

Jaffe, A.B., Newell, R.G., Stavins, R.N., 2004. Economics of Energy Efficiency. Encyclopedia of Energy, pp. 79–90.

Johansson, T., Olofsson, T., Mangold, M., 2017. Development of an energy atlas for renovation of the multifamily building stock in Sweden. Appl. Energy 203, 723–736. https://doi.org/10.1016/j.apenergy.2017.06.027.

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., Djurovic-Petrovic, M., 2010. A review of bottom-up building stock models for energy consumption in the residential sector. Build. Environ. 45, 1683–1697. https://

doi.org/10.1016/j.buildenv.2010.01.021.
Levihn, F., 2017. CHP and heat pumps to balance renewable power production: lessons from the district heating network in Stockholm. Energy. https://doi.org/10.1016/j.energy.2017.01.118.
Levihn, F., 2016. On the problem of optimizing through least cost per unit, when costs are negative: implications for cost curves and the definition of economic efficiency. Energy 114, 1155–1163. https://doi.org/10.1016/j.energy.2016.08.089.
Levihn, F., 2014. CO2 emissions accounting: whether, how, and when different allocation methods should be used. Energy 68, 811–818. https://doi.org/10.1016/j.energy.2014.01.098.
Li, Q., Quan, S.J., Augenbroe, G., Yang, P., Brown, J., 2015. Building energy modeling at urban scale: integration of reduced order energy model with geographical information. In: Presented at the 14th Conference of International Building Performance Simulation Association, pp. 190–199. Hyderabad, India.
Li, W., Zhou, Y., Cetin, K., Eom, J., Wang, Y., Chen, G., Zhang, X., 2017. Modeling urban building energy use: a review of modeling approaches and procedures. Energy 141, 2445–2457. https://doi.org/10.1016/j.energy.2017.11.071.
Liu, L., Moshfegh, B., Akander, J., Cehlin, M., 2014. Comprehensive investigation on energy retrofits in eleven multi-family buildings in Sweden. Energy Build. 84, 704–715. https://doi.org/10.1016/j.enbuild.2014.08.044.
Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J.E., Hvelplund, F., Mathiesen, B.V., 2014. 4th Generation District Heating (4GDH): integrating smart thermal grids into future sustainable energy systems. Energy 68, 1–11. https://doi.org/10.1016/j.energy.2014.02.089.
Lundström, L., Wallin, F., 2016. Heat demand profiles of energy conservation measures in buildings and their impact on a district heating system. Appl. Energy 161, 290–299. https://doi.org/10.1016/j.apenergy.2015.10.024.
Mangold, M., Österbring, M., Wallbaum, H., Thuvander, L., Femenias, P., 2016. Socio-economic impact of renovation and energy retrofitting of the Gothenburg building stock. Energy Build. 123, 41–49. https://doi.org/10.1016/j.enbuild.2016.04.033.
Mata, É., Sasic Kalagasidis, A., Johnsson, F., 2014. Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. Build. Environ. 81, 270–282. https://doi.org/10.1016/j.buildenv.2014.06.013.
Meier, A.K., Whittier, J., 1983. Consumer discount rates implied by purchases of energy-efficient refrigerators. Energy 8, 957–962. https://doi.org/10.1016/0360-5442(83)90094-4.
Ministry of the Environment and Energy, 2015. The Goal Is a Fossil-free Sweden.
Mohammadi, N., Taylor, J.E., 2017. Urban infrastructure-mobility energy flux. Energy 140, 716–728. https://doi.org/10.1016/j.energy.2017.05.189.
Moschetti, R., Brattebø, H., Skeie, K.S., Lien, A.G., 2018. Performing quantitative analyses towards sustainable business models in building energy renovation projects: analytic process and case study. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2018.06.091.
Nageler, P., Koch, A., Mauthner, F., Leusbrock, I., Mach, T., Hochenauer, C., Heimrath, R., 2018a. Comparison of dynamic urban building energy models (UBEM): sigmoid energy signature and physical modelling approach. Energy Build. 179, 333–343. https://doi.org/10.1016/j.enbuild.2018.09.034.
Nageler, P., Schweiger, G., Schranzhofer, H., Mach, T., Heimrath, R., Hochenauer, C., 2018b. Novel method to simulate large-scale thermal city models. Energy 157, 633–646. https://doi.org/10.1016/j.energy.2018.05.190.
Pasichnyi, O., Wallin, J., Kordas, O., 2019a. Data-driven building archetypes for urban building energy modelling. Energy 181, 360–377. https://doi.org/10.1016/j.energy.2019.04.197.
Pasichnyi, O., Wallin, J., Levihn, F., Shahrokni, H., Kordas, O., 2019b. Energy performance certificates — new opportunities for data-enabled urban energy policy instruments? Energy Policy 127, 486–499. https://doi.org/10.1016/j.enpol.2018.11.051.
Peng, R.D., 2011. Reproducible research in computational science. Science 334, 1226–1227. https://doi.org/10.1126/science.1213847.
Pereverza, K., Pasichnyi, O., Kordas, O., 2019. Modular participatory backcasting: a unifying framework for strategic planning in the heating sector. Energy Policy 124, 123–134. https://doi.org/10.1016/j.enpol.2018.09.027.
Pereverza, K., Pasichnyi, O., Lazarevic, D., Kordas, O., 2017. Strategic planning for sustainable heating in cities: a morphological method for scenario development and selection. Appl. Energy 186, 115–125. https://doi.org/10.1016/j.apenergy.2016.07.008.
Persson, J., Grönkvist, S., 2015. Drivers for and barriers to low-energy buildings in Sweden. J. Clean. Prod. 109, 296–304. https://doi.org/10.1016/j.jclepro.2014.09.094.
R Core Team, 2013. R: A Language and Environment for Statistical Computing.
Reinhart, C.F., Cerezo Davila, C., 2016. Urban building energy modeling – a review of a nascent field. Build. Environ. 97, 196–202. https://doi.org/10.1016/j.buildenv.2015.12.001.
Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., Kriegler, E., Mundaca, L., Seferian, R., Vilarino, M.V., 2018. Mitigation pathways compatible with 1.5°C in the context of sustainable development. In: Masson-Delmotte, V., Zhai, P., Portner, H.O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Pean, C., Pidcock, R., Connors, S., Matthews, J.B.R., Chen, Y., Zhou, X., Gomis, M.I., Lonnoy, E., Maycock, T., Tignor, M., Waterfield, T. (Eds.), Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty. Press, pp. 93–174.
RStudio Team, 2015. RStudio. Integrated Development for R.
Rysanek, A.M., Choudhary, R., 2013. Optimum building energy retrofits under technical and economic uncertainty. Energy Build. 57, 324–337. https://doi.org/10.1016/j.enbuild.2012.10.027.
Sanyal, J., New, J., 2014. Building Simulation Modelers–Are we big data ready?. In: Presented at the ASHRAE IBPSA-USA Building Simulation Conference, Atlanta, GA.
Schulz, L., 2003. Normalårskorrigering av energianvändningen i byggnader: en jämförelse av två metoder (No. 2003:01).
Seo, S., Foliente, G., Ren, Z., 2018. Energy and GHG reductions considering embodied impacts of retrofitting existing dwelling stock in Greater Melbourne. J. Clean. Prod. 170, 1288–1304. https://doi.org/10.1016/j.jclepro.2017.09.206.
Simon, H.A., 1959. Theories of decision-making in economics and behavioral science. Am. Econ. Rev. https://doi.org/10.2307/1809901.
Sokol, J., Cerezo Davila, C., Reinhart, C.F., 2017. Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. Energy Build. 134, 11–24. https://doi.org/10.1016/j.enbuild.2016.10.050.
Sola, A., Corchero, C., Salom, J., Sanmarti, M., 2018. Simulation tools to build urban-scale energy models: a review. Energies 2015 8. https://doi.org/10.3390/en11123269, 7261-7278 11, 3269–24.
Späth, P., Rohracher, H., 2015. Conflicting strategies towards sustainable heating at an urban junction of heat infrastructure and building standards. Energy Policy 78, 273–280. https://doi.org/10.1016/j.enpol.2014.12.019.
Stiglitz, J.E., 1993. Economics. W.W. Norton&Company Inc., USA.
Stockholms stad, 2019. Årsbokstabeller För Byggnadsverksamhet.
Stockholms stad, 2015. Vision 2040 – A Stockholm for Everyone. international, stockholm.se.
Stockholms stad, 2014. Roadmap for a Fossil Fuel-free Stockholm 2050.
Svebyprogrammet, 2012. Brukarindata Bostäder. Sveby Stockholm, Stockholm.
Swan, L.G., Ugursal, V.I., 2009. Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. Renew. Sustain. Energy Rev. 13, 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033.
Swedish Meteorological and Hydrological Institute, 2017. Open Data Catalogue.
Tableau Software, 2019. Tableau: Business Intelligence and Analytics Software.
The PostgreSQL Global Development Group, 2019. PostgreSQL.
Tong, X., Li, R., Li, F., Kang, C., 2016. Cross-domain feature selection and coding for household energy behavior. Energy 107, 9–16. https://doi.org/10.1016/j.energy.2016.03.135.
Truong, N.L., Dodoo, A., Gustavsson, L., 2018. Effects of energy efficiency measures in district-heated buildings on energy supply. Energy 142, 1114–1127. https://doi.org/10.1016/j.energy.2017.10.071.
Wang, Q., Holmberg, S., 2015. A methodology to assess energy-demand savings and cost effectiveness of retrofitting in existing Swedish residential buildings. Sustainable Cities and Society 14, 254–266. https://doi.org/10.1016/j.scs.2014.10.002.
Wang, X., Li, Z., Meng, H., Wu, J., 2017. Identification of key energy efficiency drivers through global city benchmarking: a data driven approach. Appl. Energy 18–28. https://doi.org/10.1016/j.apenergy.2016.12.111.
Webber, P., Gouldson, A., Kerr, N., 2015. The impacts of household retrofit and domestic energy efficiency schemes: a large scale, ex post evaluation. Energy Policy 84, 35–43. https://doi.org/10.1016/j.enpol.2015.04.020.
Wickham, H., 2014. Tidy data. J. Stat. Softw. 59, 1–23. https://doi.org/10.18637/jss.v059.i10.
Wickham, H., Grolemund, G., 2017. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media.
Yang, F., Chen, L., 2016. Developing a thermal atlas for climate-responsive urban design based on empirical modeling and urban morphological analysis. Energy Build. 111, 120–130. https://doi.org/10.1016/j.enbuild.2015.11.047.
Yazdanie, M., Densing, M., Wokaun, A., 2017. Cost optimal urban energy systems planning in the context of national energy policies: a case study for the city of Basel. Energy Policy 110, 176–190. https://doi.org/10.1016/j.enpol.2017.08.009.
Ye, X., Xia, X., 2016. Optimal metering plan for measurement and verification on a lighting case study. Energy 95, 580–592. https://doi.org/10.1016/j.energy.2015.11.077.
Zhou, Z., Zhang, S., Wang, C., Zuo, J., He, Q., Rameezdeen, R., 2016. Achieving energy efficient buildings via retrofitting of existing buildings: a case study. J. Clean. Prod. 112, 3605–3615. https://doi.org/10.1016/j.jclepro.2015.09.046.