Developing a Data-driven school building stock energy and indoor environmental quality modelling method

Y. Schwartz *, D. Godoy-Shimizu, I. Korolija, J. Dong, S.M. Hong, A. Mavrogianni, D. Mumovic

UCL IEDE – University College London, Institute for Environmental Design and Engineering, United Kingdom

ABSTRACT

The school building sector has a pivotal role to play in the transition to a low carbon UK economy. School buildings are responsible for 15% of the country's public sector carbon emissions, with space heating currently making up the largest proportion of energy use and associated costs in schools. Children spend a large part of their waking life in school buildings. There is substantial evidence that poor indoor air quality and thermal discomfort can have detrimental impacts on the performance, wellbeing and health of schoolchildren and school staff. Maintaining high indoor environmental quality whilst reducing energy demand and carbon emissions in schools is challenging due to the unique operational characteristics of school environments, e.g. high and intermittent occupancy densities or changes in occupancy patterns throughout the year. Furthermore, existing data show that 81% of the school building stock in England was constructed before 1976. Challenges facing the ageing school building stock may be exacerbated in the context of ongoing and future climate change.

In recent decades, building stock modelling has been widely used to quantify and evaluate the current and future energy and indoor environmental quality performance of large numbers of buildings at the neighbourhood, city, regional or national level. Building stock models commonly use building archetypes, which aim to represent the diversity of building stocks through frequently occurring building typologies.

The aim of this paper is to introduce the Data dRiven Engine for Archetype Models of Schools (DREAMS), a novel, data-driven, archetype-based school building stock modelling framework. DREAMS enables the detailed representation of the school building stock in England through the statistical analysis of two large scale and highly detailed databases provided by the UK Government: (i) the Property Data Survey Programme (PDSP) from the Department for Education (DfE), and (ii) Display Energy Certificates (DEC). In this paper, the development of 168 building archetypes representing 9,551 primary schools in England is presented. The energy consumption of the English primary school building stock was modelled for a typical year under the current climate using the widely tested and applied building performance software EnergyPlus. For the purposes of modelling validation, the DREAMS space heating demand predictions were compared against average measured energy consumption of the schools that were represented by each archetype. It was demonstrated that the simulated fossil-thermal energy consumption of a typical primary school in England was only 7% higher than measured energy consumption (139 kWh/m²/y simulated, compared to 130 kWh/m²/y measured). The building stock model performs better at predicting the energy performance of naturally ventilated buildings, which constitute 97% of the stock, than that of mechanically ventilated ones. The framework has also shown capabilities in predicting energy consumption on a more localised scale. The London primary school building stock was examined as a case study.

School building stock modelling frameworks such as DREAMS can be powerful tools that aid decision-makers to quantify and evaluate the impact of a wide range of building stock-level policies, energy efficiency interventions and climate change scenarios on school energy and indoor environmental performance.

* Corresponding author.
  E-mail address: yair.schwartz@ucl.ac.uk (Y. Schwartz).

1. Introduction

It is estimated that people in Europe spend more than 90% of their time indoors on average [1]. Children, in particular, spend a large part of their waking life in school buildings (approximately 30% of their life at school, around 70% of which inside a classroom) [2]. School buildings typically have high and intermittent occupancy densities, which can result in high internal heat gains and irregular heating demand patterns) [3]. Furthermore, the way school classrooms are used changes throughout the day and around the year. Special attention is, therefore, required when designing school environments, due to the unique challenges they face.

Maintaining indoor environmental comfort in such spaces is, thus, particularly challenging; especially in the context of climate change and associated increases in ambient temperatures. According to current climate change projections, the UK is expected to experience warmer, wetter and windier winters, and hotter and drier summers [4]. As the UK is a predominantly heating dominated climate, its school buildings were originally designed to be primarily naturally ventilated and may not be prepared to cope with high levels of indoor overheating risk [5] Furthermore, data suggest that up to two-thirds of the total English school floor area was built before 1976 [6]. Therefore, research on current and future school building stock performance has become increasingly important in recent decades [7,8].

School buildings are responsible for around 15% of the UK's public sector carbon emissions. Energy expenditure is often the largest non-staff-related cost for schools [9]. This makes school buildings an important element in the transition to a low energy and low carbon economy: It is estimated that the UK school building stock has the potential to reduce its energy bills by £44 million and prevent 625,000 tonnes of CO_2 from entering the atmosphere, annually [9]. The school building stock offers significant opportunities for reduction in energy use, as the factors that determine energy performance, such as activity patterns and equipment use, are fairly similar across the stock. For this, existing school buildings will need to be retrofitted to higher energy standards. While energy efficient design techniques, such as high thermal insulation and airtightness levels, may potentially lead to reduced heating demands, it is important to note that such design strategies could impair indoor environmental quality and lead to overheating, if they lack a whole system approach. A detailed analysis of the performance of the stock can assist policymakers and stakeholders in improving energy performance and indoor air quality in the school building stock.

Building stock modelling is widely used to examine the current and future energy and indoor environmental quality performance of large numbers of buildings at the neighbourhood, city, regional or national level. These models often adopt an archetype approach that uses a number of 'typical' buildings to represent the diversity of the building sector. This approach enables decision-makers and other stakeholders to investigate the performance of the entire building stock under a range of different scenarios (e.g., climate change, refurbishment packages, design strategies etc.). Such an approach can help policymakers predict building stock performance under different policy scenarios and help inform the development of appropriate policies and regulations. This study forms part of the UK Engineering and Physical Sciences Research Council (EPSRC) funded project 'Advancing School Performance: Indoor environmental quality, Resilience and Educational outcomes' ('ASPIRE'), which aims to understand how energy efficient building design strategies might affect the indoor environmental quality of schools in the UK, and provide recommendations for optimum low carbon and healthy school building design.

This paper presents a novel stock-modelling framework - Data dRiven Engine for Archetype Models of Schools (DREAMS) - that is capable of offering a detailed representation of the English primary school building stock. DREAMS uses data from two large-scale databases (i) Property Data Survey Programme (PDSP [6]), a survey of school buildings in England, and (ii) Display Energy Certificates (DEC), which contains data on a range of public building thermal properties. The objectives of this study are:

- To accurately characterise, model and simulate the entire English primary school building stock based on the statistical analysis of the PDSP and DEC databases.
- To present the development of the archetype-based English primary school building stock model. This approach can enable a detailed exploration of energy efficiency levels (heating and cooling load) and a range of other building performance metrics (e.g., indoor air quality, thermal comfort etc.) under different climate change scenarios.
- To present the results of the model's validation by comparing simulated performance against measured energy use data.

2. Background

2.1. Building stock modelling approaches

In recent decades, building stock modelling has been widely employed to predict energy performance of building stocks at different scales (neighbourhood, city, regional, national and cross-national), and evaluate the potential impact of energy conservation measures and climate change scenarios. Approaches to modelling building stocks can broadly be broken down into two main categories: 'top-down' and 'bottom-up' approaches [10–13].

2.1.1. The top-down approach works at an aggregated level

Energy consumption is modelled by establishing statistical relationships between building energy use and macroeconomic or other variables (often climatic ones). This approach is, therefore, often used to estimate the aggregated impacts of building energy consumption at regional and national levels. The top-down approach relies on historical data to define those statistical relationships. This means that predictions of future performance are based on past performance trends, which might not be applicable for examining future scenarios and unpredicted events, such as climate change, accurately [10]. The top-down approach is easy to apply as it requires a limited set of inputs for model development [12], while it considers the building stock as an energy sink and does not provide detailed energy results of individual buildings or end-uses [11]. This may lead to limitations in identifying specific policy measures for improving energy efficiency of different building types.

2.1.2. The Bottom-up approach works at a disaggregated level

Energy consumption is usually calculated for individual end-uses, buildings or groups of buildings, and then the modelling outputs are aggregated at stock level. The bottom-up approach can be further divided into two sub-categories:

a. Bottom-up statistical approach – This approach estimates building energy consumption based on empirical and measured data that expands further from the data used in the 'top-down' approach. As it is based on the use of historic data, this statistical approach is capable of incorporating the effects of occupancy behaviour and other measured or observed data on energy consumption [11]. However, similarly to the top-down approach, the reliance on historical data limits its capability to explore the impacts of future technologies on building energy performance.

b. Bottom-up engineering approach – The engineering approach estimates building performance through thermodynamic calculations using inputs that are related to the physical characteristics of buildings. It is possible, by using this approach, to explore the impact of various thermal performance measures (e.g., building fabric or systems improvements). In contrast to the statistical approach, however, one of its main limitations is that describing and modelling human behaviour may not be complete [10].

For the purpose of guiding energy policy and decision making, applying building simulations through the bottom-up engineering approach may require significant resources, especially when a large number of buildings is involved. This approach can be time consuming, complex and costly, in particular in terms of data gathering, computing time and power.

2.2. Archetype stock modelling approach

To address these challenges, an archetype approach is often adopted for building stock modelling, a modelling technique whereby a small number of buildings are defined as broadly representative of the entire building stock. A key step in developing an archetype-based building stock model is the selection and definition of archetypes: too many unique buildings will increase the resource requirements and complexity of the model; too few may mean that the model fails to accurately represent the complexity and diversity of the stock being examined [14]. Therefore, the archetype approach relies on the identification of similar building properties across the overall stock and the identification of the optimum number of categories into which buildings are classified.

To date, archetype building stock models have been mostly developed for the residential sector, ranging from the regional to the national level [15–19]. Archetype-based building stock approaches vary as a function of data availability, specific study aims and regional/national stock characteristics. However, common patterns emerge across methodological approaches. General procedures of archetype-based building stock modelling approaches include:

1) Data collection and processing;
2) Classification of buildings;
3) Determination of representative parameters for each building category;
4) Development of archetype models.

With the collected data providing information of given buildings, the archetypes are classified based on a set of shared energy-related properties within each building category. This is followed by determining representative parameters depending on their occurrence frequency at stock levels, which will finally be used to develop archetypes models.

Archetype based building stock models are typically further split into two different categories: *virtual archetype models* and *sample archetype models*.

2.2.1. Virtual archetype models

Use notional averaged archetypes that represent different building categories. For example, in studies in some European countries, such as Germany and Denmark, the residential building stock has been classified to enable the development of archetypes for energy performance analysis [18,19]. In most cases, however, input information requires access to different building databases or even databases that were not specifically developed for building energy assessment processes (e.g. census data or other local authority owned data) [17,20].

Since the data that are used to establish archetypes at different studies come from a diverse, non-homogeneous range of data sources, the classfication methods applied to building stocks vary widely. Some building parameters (e.g. floor areas) used to represent each category and characterise archetype models are defined by statistical analysis, and often derived from average values of the entire building stock data [16]. Other parameters such as thermal properties, building systems and occupant behaviours) are often determined through referring to building codes and standards, and research literature [17,21]. Building properties such as building type and construction age are often recognised by many studies as the most impactful variables for energy use, however, in cases where detailed buildings thermal properties are missing, the specific criteria are often decided based on the modeller's expertise [22].

2.2.2. Sample archetype models

Use the information of real buildings with similar characteristics to the mean features of the concepts that were introduced in section 2.2.1. Mata et al. [23] used the sample approach to represent the Swedish residential stock by a sample of 1,400 buildings. The individual energy performance predictions were assigned weighting coefficients to represent the fraction of each building category in the stock, and results of building stock were later aggregated. Another study [24] selected 12 sample buildings as archetype models to generate detailed thermal energy demand analysis at an urban district in Turin, Italy. A main limitation of this approach would be the assurance that the selected sample accurately represent the entire stock, however, using a large number of sample archetype models will increase the reliability of the stock's representation, especially for large-scale building stock with a variety of building types [11,24].

In some cases, archetype model development combines both the theoretical approach and the sample approach due to data availability variation in different regions. For instance, the TABULA project (Typology Approach for Building Stock Energy Assessment) offers a harmonized methodological framework for the representation of residential building stocks in 20 European countries using both theoretical and sample archetypes for energy performance prediction and assessment [25].

At the time of writing, there was only a limited number of archetype models for non-residential building stocks. This may partly be because non-residential buildings are more diverse than residential buildings and therefore, often more challenging to represent using a limited set of archetypes [26]. In order to tackle this challenge, Korolija et al. [27] developed a novel method to describe the UK office building stock by using parametric modelling to create office archetypes. The study parameterized the energy-related building characteristics (e.g., building form, glazing ratio, envelope construction etc.), and developed 3,840 virtual office building models overall. This process, however, did not account for the occurrence frequency of each archetype in the stock, which may not be suitable for providing aggregated stock-level performance figures.

2.3. Limitations of archetype stock modelling

Both archetype approaches outlined above (virtual and sample) are characterised by a number of limitations:

a. The ability to address uncertainties: The major issue affecting the robustness of archetype models is the uncertainty in modelling inputs. Booth et al. [22] summarized two types of uncertainties first-order (aleatory) uncertainty caused by the random variations of input parameters (e.g. occupant behaviour, HVAC systems efficiency levels etc.) and second-order (epistemic) uncertainty, due to lack of knowledge on certain modelled parameters. In both virtual and sample archetype models, input parameters are often defined by a deterministic approach (e.g., where a single value is assigned to a modelled parameter whereas stochastic models

(e.g., models that account for changes in certain parameters) may be capable of addressing some uncertainties [28]. These may include occupant behaviour [28,29], or climate and building-related parameters [30].

b. Levels of accuracy: The limited number of representative building models in the archetype model is likely to affect the accuracy of results, as archetype models are, by their very definition, approximations of actual buildings [31].

Whilst the availability of monitored building performance data at the building stock level is often limited [30], stock-level performance results are compared against measured data, whenever possible. Studies have shown promising agreement between simulated and measured performance using archetype models: 8% discrepancy in a Sicilian residential stock [21], and 4% discrepancy on a residential and commercial stock model in Milan, Italy [32]. The agreement between simulated results and measured data at stock level is explained by Reinhart and Cerezo Davila [31], who asserts that individual model inaccuracies due to uncertainties tend to average out when aggregated at stock level, so the apparent overall level of accuracy is higher than that for any individual archetype.

c. Lack of linking energy consumption evaluations to indoor environment quality: Energy performance of buildings is highly relevant to indoor environmental performance and occupant satisfaction [33]. However, the majority of existing archetype models to date have focused solely on energy performance. Although there are a few studies that focus on indoor environmental quality at stock levels [34–36], a more integrated approach is required, capable of considering both energy and indoor environmental conditions. Such approaches may be useful in informing integration of climate change mitigation and adaption strategies, in the context of international climate policies [37].

In summary, the archetype approach is widely adopted in building stock energy use and indoor environmental quality modelling due to its ability to describe and represent large groups of buildings in a relatively simple way. Under existing data and computational capacity, the archetype approach offers a good balance between detailed description of the whole building stock on the one hand, and modelling efforts on the other. To date, the archetype approach has been well developed for residential building sectors, but less so for more heterogeneous non-domestic building stocks. Nonetheless, as the operational energy use portion of non-domestic buildings is estimated to grow in the future [38], the performance evaluations of building performance for the non-domestic building stock is increasingly critical and a topic of interest in building stock modelling studies. Due to building stock models inherent constraints which could affect the result accuracy, a validation process is required to evaluate the representativeness of archetype models. The accuracy of modelling results can be improved if audits or measured energy data are available at the stock level.

3. Methodology

Building on existing archetype building stock modelling approaches outlined in the previous section, this paper presents a theoretical archetype stock model for primary school buildings in England. The Data dRiven Engine for Archetype Models of Schools (DREAMS) is a novel school building stock model framework based on data-driven building theoretical archetypes.

Figure 1 schematically illustrates the study design. DREAMS is based on two extensive and detailed databases of school buildings including both observational (building thermal properties) and measured data (energy consumption). By analysing the data from a national survey of the schools estate, the

Department for Education (DfE)'s Property Data Survey Programme (PDSP) [6], school buildings were classified into groups based on a series of buildings characteristics. The study then proposes an automated process, through which a set of school building archetypes were developed and defined as representative of the whole stock in England. These archetypes were subsequently simulated using EnergyPlus – a dynamic thermal simulation tool which is widely tested and used both in the industry and academia [39]. Predicted building performance analysis was then carried out for energy use performance, and simulation results were then compared with measured energy consumption derived from Display Energy Certificates database (DEC) [41], to ensure the model accurately predicts the stock's performance.

The main novelty of DREAMS lies within its use of databases and its application to the school building sector: It is the first time that a nation-wide school building stock has been modelled and simulated using the archetype approach, based on an extensive, nation-wide, detailed database of school buildings. It is believed that the granularity of the data will lead to a more accurate depiction of the stock.

3.1. DREAMS framework development

Fig. 2 illustrates the method behind the development of the DREAMS framework. Section 3.1.1 provides an overview of the databases that were used in this study, and their statistical analysis. Section 3.1.2 introduces the concept behind the development of the seed models. Section 3.1.3 explains the data processing and the procedures underlying the generation of the school archetype models. Last, section 3.1.4 details the steps of building simulation, post-processing, results analysis and comparison to measured energy consumption figures.

3.1.1. Initial data analysis – The data sources

The Property Data Survey Programme (PDSP), was originally commissioned by the UK Government's Partnerships for Schools (later part of Education Funding Agency) [6]. It collected information on the physical condition of the education estate of England between 2012 and 2014. While not all the English school building stock was surveyed (e.g. recently built or modernised schools were excluded), the resulting database includes information on 18,970 establishments across the country. This represents 85% of the school stock in England, and includes primary and secondary schools, as well as nurseries and special institutions as shown in Table 1. For primary schools - the focus of this study- these surveys cover around 90% of the total stock.

The programme was not originally designed to collect building energy use information. However, part of the collected data is useful for investigating school building thermal performance. This includes the following key variables: the number of buildings in each school's premises, building footprint area, number of storeys, average Window-to-Wall Ratio (WWR), and building construction age. Outside of the focus of the present study, data gathered as part of PDSP include information on external areas, internal finishes and sanitary services [40].

In addition to PDSP, another key source of school building stock information used in this study was the Display Energy Certificates (DEC) database, acquired from the bulk public release. Introduced in 2008, DECs provide standardised and normalised performance benchmarks for large non-domestic public buildings in England and Wales [41]. Variables within the DEC database include the building internal environment: Heating, Ventilation and Air Conditioning (HVAC) systems, main heating fuel, occupancy levels and measured annual energy consumption data, presented as electricity and fossil-thermal energy intensities (kWh/m^2). A number of
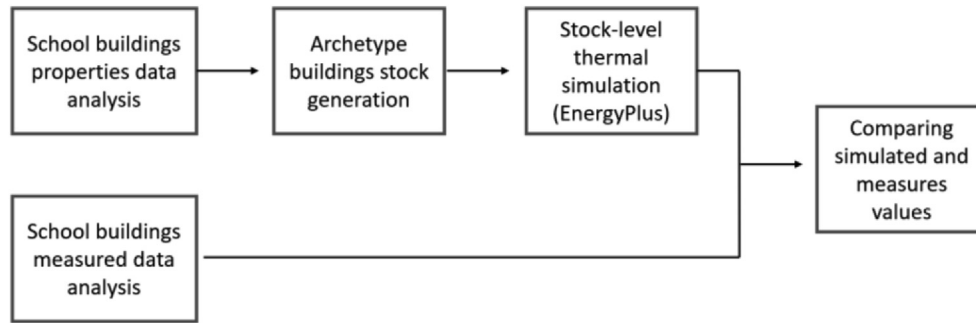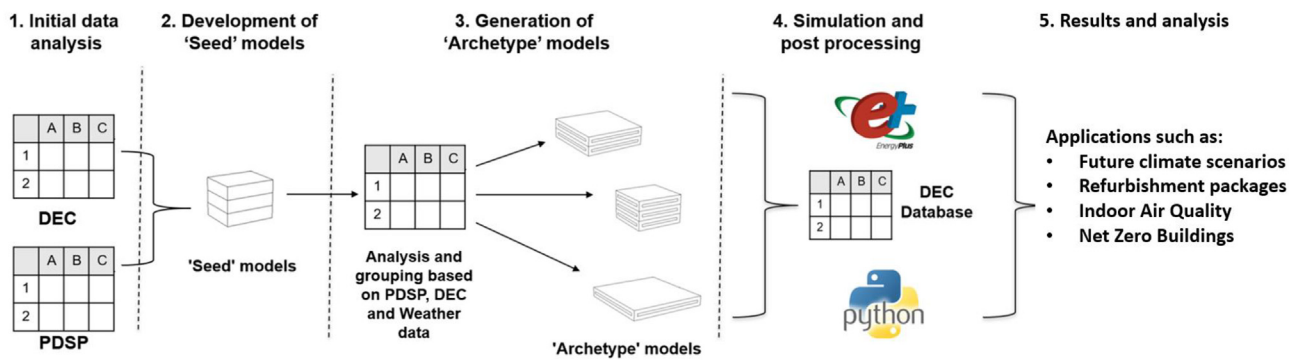
Fig. 1. The study design.

Fig. 2. The DREAMS framework: archetype development.

Table 1
An overview of the surveyed schools' data within the PDSP dataset.

Phase of Education	Survey Year			All
	2012	2013	2014	
	N	N	N	N
16 Plus	20	78	3	101
All Through	14	16	13	43
Middle Deemed Primary	4	5	9	18
Middle Deemed Secondary	37	59	39	135
Not applicable	160	457	365	982
Nursery	106	193	93	392
Primary	2,489	6,324	6,479	15,292
Secondary	507	935	565	2,007
All	**3,337**	**8,067**	**7,566**	**18,970**

studies presenting statistical analyses of school DECs have been published to date [42,43]. The DEC data used for this study comprise 44,127 certificates for primary schools, lodged between 9 March 2010 and 1 October 2016.

Considerable data processing was necessary to use the separate datasets for the purposes of archetype generation. This included pre-processing of the PDSP and DEC files separately, matching the two datasets, and post-processing. The steps are detailed below:

- Pre-Processing: Prior to linking to PDSP, the DEC data was processed using methods developed by [44] to exclude records that could introduce uncertainties to the analysis, or that are for buildings outside of the focus of this study. The process involved checking the records for data formatting and completeness, and excluding any records with the following characteristics:
- Duplicate entries
- Records that were updated in less than 6 months from the previous record.

- Records with unusual normalised energy figures (the 'operational rating'), including DEC with operational ratings above 1,000, below 5 (unusually high and low respectively) or 200 or 9999 (default values)
- Records with floor area lower than 50 m^2
- Mixed-fuel use buildings with no electrical or fossil-thermal energy use
- Composite DECs (where a building has significant areas of different uses that are not sub-metered, these may be produced under a 'composite' methodology)
- Postal address matching: Next, the DEC database was matched to the PDSP database based on postal addresses of the entries in each database. This was carried out at the level of each individual school, using the address data from each file. Address-matching was undertaken in a semi-automated, geographically scaled manner: Code, written in SAS 9.4 [45], calculated Levenshtein distances between pairs of addresses prioritised based on the geographic data, using the postcodes as primary identifiers. Following this, manual inspection of each pair was performed, to validate the address-matching process.
- Post-processing: Following the address-matching process, schools without matched building characteristics and energy data (from the PDSP and DEC databases, respectively) were removed. To improve the alignment of the two datasets, DECs based on surveys that took place between 2012 and 2014 (the period when the PDSP data were collected) were selected, where available. Last, since this study focuses on primary schools, entries for other school types were removed.

Following the processing and matching steps detailed above produced a 'combined school dataset' with 9,551 primary schools, from which the archetype models were produced for DREAMS. References to the schools data in the remainder of this paper refer to

this combined dataset, rather than the raw, separate DEC or PDSP files.

3.1.2. The development of seed models

A core component of the DREAMS framework is a set of school-building models, or 'seed' models. These are thermal models that were created based on the five construction eras, as recorded in the PDSP survey; pre-1919, inter-war, 1945–1966, 1967–1976, post-1976. These represent 'distinct eras within school building programmes and schools built within these eras often have similar construction characteristics, maintenance needs and lifecycle expectations'. Each construction era was associated with a typical school's built form (e.g., whether a building has a '+' shape footprint or an 'l' shape) and typical building fabric characteristics (construction build-ups, U-Values), as seen in Table 2. It is

acknowledged that built forms do vary within these periods, for example, by reflecting regional variations in construction trends over time. Work is ongoing to account for these issues with a more disaggregated modelling approach.

The resulting five seed models, in the form of EnergyPlus Input Data Files (*.idf) include information on building geometry, along with the thermal properties of the building envelope, internal loads and HVAC-related pre-set characteristics (e.g., infiltration, systems etc., as described in Table 3).

- The *building geometry* of these representative forms were produced based on [46] and an online survey of schools in England, carried out using Google Maps and Bings Maps [47,48].
- The *building fabric characteristics* - Construction build-ups were defined based on [49–53].

Table 2
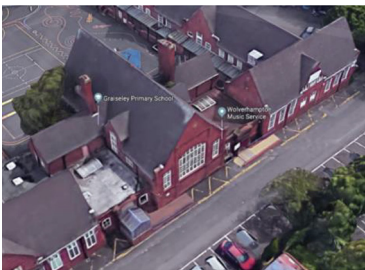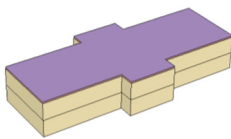'Seed' models (For the U-Values: GF = Ground Floor. EW = External Wall. R = Roof).
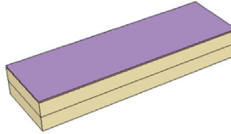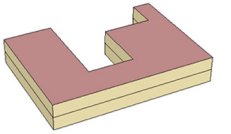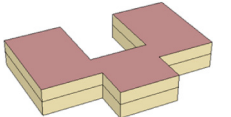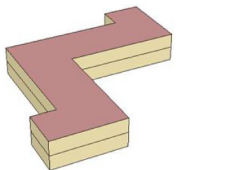
Era	Representative building (Figures by [47])	'Seed' model EnergyPlus file	Build-ups and U-Values [W/m²k]
Pre 1919			GF: 1.5 EW: 1.8 R: 2.9
Inter-war			GF: 1.5 EW: 1.8 R: 2.9
1945–1966			GF: 1.4 EW: 1.7 R: 1.9
1967–1976			GF: 1.4 EW: 1.7 R: 1.2
Post 1976			GF: 0.95 EW: 0.83 R: 0.57

Table 3
Ventilation inputs.

	Value	Time on	Other
Infiltration	6 l/m² exterior surface area	24 h	
Natural Ventilation	8 l/second/person	10 min at the beginning of every hour between 09:00–18:00, if external temperature is not below 19 °C, and difference between indoor and outdoor temperature is not greater than 2 °C	
Mechanical Ventilation	8 l/second/person	09:00 – 18:00	• Pressure drop across the system: 1,000 Pa • Fan efficiency: 70%

- *Internal heat loads* for the seed models were obtained from the National Calculation Method (NCM) [54] and Building Bulletin 101 (BB101, 'Guidelines on ventilation, thermal comfort and indoor air quality in schools') [52], as detailed in Tables 4 and 5.

The school building seed models adopt a 'zone-per-floor' approach with regard to glazed openings and occupancy levels.

Table 4
'Seed models' Internal loads.

	Area-weighted figure in the seed model	Schedule
Lighting	5.1 W/m²	09:00–16:00: 100% 16:00–18:00: 50%
Occupancy	0.3 ppl/m² (with 110 W/person)	09:00–16:00: 100% 16:00–18:00: 50%
Electrical equipment	3.3 W/m²	09:00–16:00: 100% Otherwise: 5%

Table 5
Thermostat inputs.

	Value	Time on
Heating setpoint	20 °C	09:00–18:00
	12 °C	Otherwise

According to this approach, each thermal zone has been defined as an entire floor. While the seed models hold no information regarding window sizes or locations, this approach enables easy manipulation of a range of building characteristics, such as building floor area or window-to-wall-ratio (WWR), later on in the process of theoretical archetype generation. Similarly, school buildings accommodate a mix of functions and spaces that are used in different ways during the day (e.g., classrooms and circulation areas). To allow for this variety of occupancy within the 'single zone-per-floor' model, an area-weighted approach per each seed was taken; whereby an overall average internal environment condition was produced based on the proportion of the total floor area of the particular seed, as associated with each use type.

3.1.3. Archetype model generation
a. Property data Survey Programme database (PDSP) and classification into archetypes

The combined school dataset included data for 9,551 primary schools, across England. The data were then further filtered, classified and divided into groups, or 'archetypes', based on 5 key building characteristics (as illustrated in Fig. 3). These variables are known to impact on energy performance in schools:

- *Location* (degree day regions): Climate impacts on building energy use by directly influencing heat losses and gains. Typi-
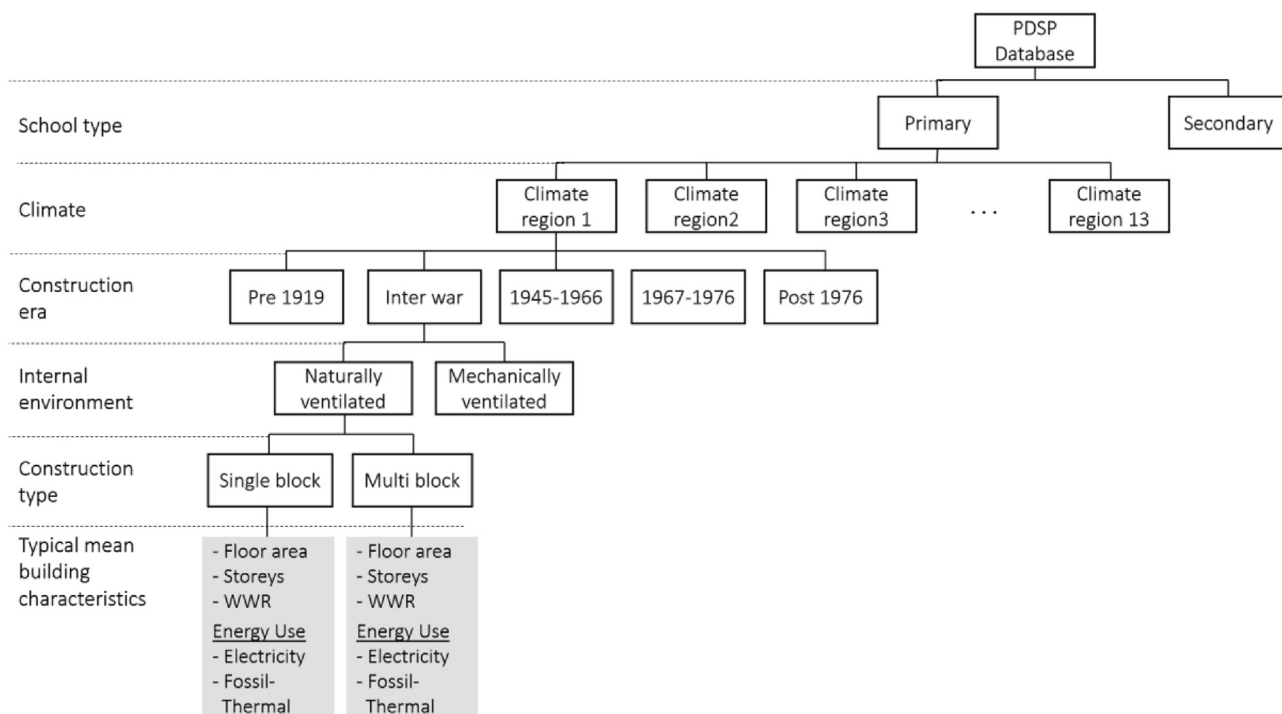

Fig. 3. The filtering and grouping of the PDSP and DEC databases, and the classification into 'archetype' buildings.

cally, when comparing energy performance between buildings in different locations or over different time periods this is accounted for via a 'weather adjustment' process (see e.g. [55,56]) This involves scaling the portion of a building's energy demand associated with space heating, based on the local climate during the energy measurement period. For this study, 13°Day climate regions were used (assigned to each school based on the postal address) in line with the DEC methodology [55]. In line with recent research, an assumption of 80% of fossil-thermal energy associated with space heating was used for weather correction [55] instead of the 55% used in the DEC methodology [55].

- *Building age* (pre 1919; inter war; 1945–66; 1967–76; post 1976): For this study, building age data was taken from the PDSP dataset. Construction age impacts on energy use, by acting as a proxy for the thermal properties of the building envelope. Additionally, as building design is influenced by regulations and architectural trends, other attributes such as built form will also vary with construction age. In this study, construction build-ups and the school's built form are associated with the

Table 6
Climate regions and their weather files [58].

	climate region	CIBSE weatherfile
1	Thames Valley	London_TRY
2	South-eastern	London_TRY
3	Southern	Southampton_TRY
4	South-western	Plymouth_TRY
5	Severn Valley	Swindon_Brize_Norton_TRY
6	Midland	Birmingham_TRY
7	West Pennines	Manchester_TRY
8	North-western	Newcastle_TRY
9	Borders	Newcastle_TRY
10	North-eastern	Leeds_TRY
11	East Pennines	Nottingham_TRY
12	East Anglia	Norwich_TRY
13	Wales	Cardiff_TRY

school's construction era (as detailed in Table 1). Consequently, all schools built in a certain era, were assumed to have the same build-up and initial built form. This is illustrated in the representative building forms for each seed model.

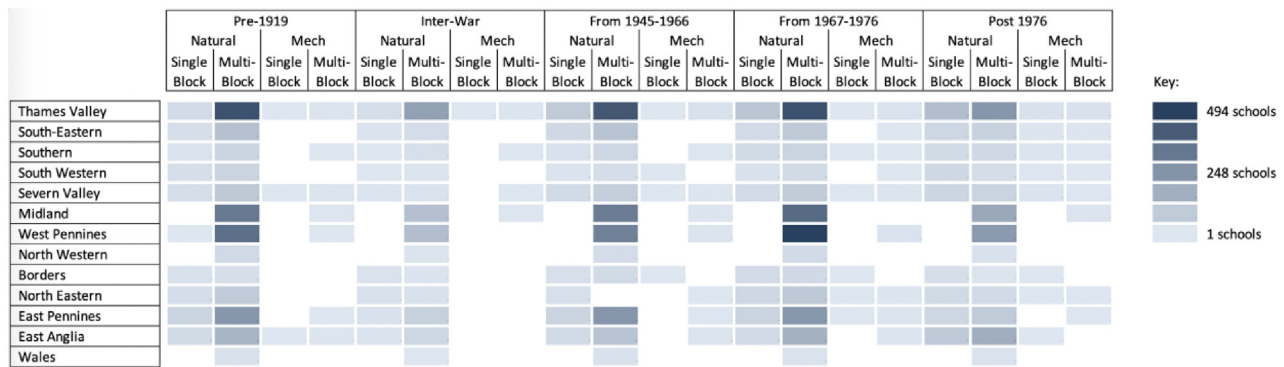**Fig. 4.** Frequency distribution of each archetype across climate regions in primary schools across England.
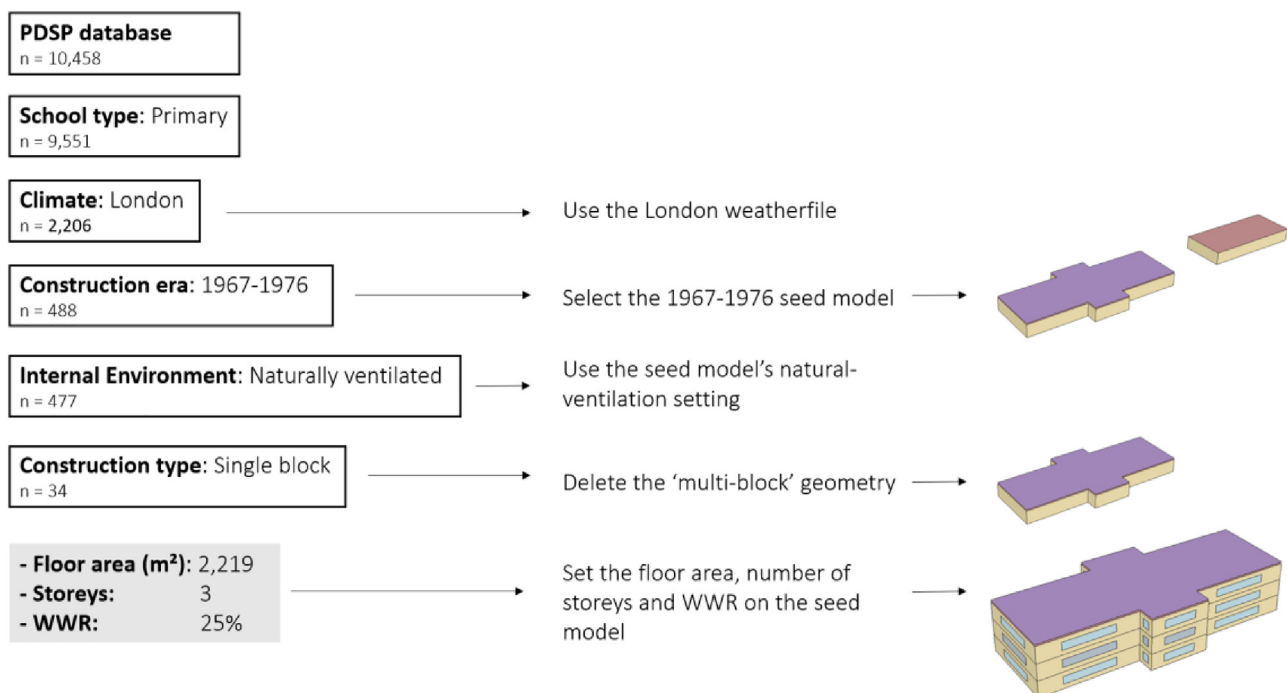

Fig. 5. An example of the manipulation of a seed model and the creation of a theoretical archetype building (n = number of schools).

- *Internal environment* (natural or mechanical ventilation): While mechanical ventilation remains relatively uncommon in UK schools, past studies have shown that the type of ventilation is associated with a significant difference in energy consumption [42,43]. For primary schools, mechanical ventilation was found to be associated with 12% higher electricity intensity, and 15% lower fossil-thermal intensity on average compared with those with natural ventilation. For the purposes of this study, the internal environment field available from the DEC database was used, which categorises buildings as air conditioned, mechanically or naturally ventilated, or mixed mode.
- *Construction type* (single- or multi-block school): The PDSP database includes both single-block schools and multi-block ones. To account for the possibility of different construction dates or internal characteristics between school blocks within a single school, this study generated two building entities within each seed model: an 'original' building, which is assumed to be the largest building in a school, and an additional building, which is an aggregation of the floor area of the rest of the buildings in the school. If a specific school is a 'single-block' school, the additional building is deleted from its seed model.

Following the creation of the above list, each 'seed' model was manipulated to accurately represent the group of school buildings that shared the same characteristics. e.g., all pre 1919 London primary schools that do not have any extension and are naturally ventilated were modelled by manipulating the 'pre 1919' seed model:

- applying the appropriate primary school usage profile, using the London weather file, assigning the pre 1919 construction build-ups,
- enabling natural ventilation in the simulation (setting up windows opening schedule),
- ensuring only the main school building is modelled and simulated, and
- ensuring its dimensions (floor area and number of floors) are the average of all the schools it represents.

Each school archetype, thus, corresponds to a unique combination of the above variables. A full list of the theoretical archetypes,

including frequency counts across the stock, is provided in Fig. 4 below. It should be noted that, unsurprisingly, school building characteristics of schools vary geographically across England. As a result, some theoretical archetypes do not exist in certain regions within the database.

Typical building energy performance modelling inputs were then calculated for each archetype, including: average floor area (m^2), number of storeys (n), WWR (%), electricity (for lighting and fans, in case of mechanical ventilation) and fossil-thermal energy use intensities (energy for space and water heating, expressed in kWh/m^2), for all schools within each archetype, as seen in Fig. 4 below. Further details are provided in the Appendix.

b. The manipulation of seed models and creation of the archetype models

Following the above procedures, the five seed EnergyPlus models were processed using a customised archetype-model-generation script using Python 3.6.2 [57], assigning the relevant categories described above (namely, floor area, number of storeys and WWR). A computer program was developed (using Python 3.2) to automatically read the seed models and modify the relevant parameters, as illustrated on Fig. 3 above. The 9,551 primary schools could now be represented by 168 theoretical archetype models. Fig. 5 offers an example of the manipulation of a seed model and the output of the customised theoretical archetype model generation process. The figure shows an example of the generation the archetype of single-block, London-based, naturally ventilated primary schools, that had been built between 1967 and 1976. It represents 34 schools that have an average overall floor area of 2,219 m2, 3 storeys and an average of 25% WWR.

3.1.4. Simulation controller and building energy performance analysis

Once the full set of archetypes thermal simulation modelling input definition files (*.idf) were generated, a second script was developed to automate a batch-simulation process and results post processing. EnergyPlus (version 8.9, [39]), one of the most widely used thermal simulation tool in built environment research, was used to perform the thermal simulation analysis. Based on two energy and load simulation tools (BLAST and DOE-2), EnergyPlus and its IDF Editor enable relatively easy access to its input files and a quick and simple manipulation of input parameters.

Table 7
Archetype buildings stock-level sample size. * Nat = Naturally ventilated building. Mec = Mechanically ventilated building. Single = Single block. Multi = Multi block.

Building era	School description	Stock-level sample size (actual number of schools in the database)	Number of Archetypes across all of England's climate regions
Pre-1919	Nat-Single	261	10
	Nat-Multi	1,984	13
	Mec- Single	5	3
	Mec- Multi	33	7
Inter war	Nat- Single	106	9
	Nat- Multi	672	13
	Mec- Single	2	1
	Mec- Multi	11	4
1945–1967	Nat- Single	329	9
	Nat- Multi	1,730	12
	Mec- Single	7	4
	Mec- Multi	50	8
1967–1976	Nat- Single	392	9
	Nat- Multi	2,138	13
	Mec- Single	11	5
	Mec- Multi	62	9
Post-1976	Nat- Single	501	9
	Nat- Multi	1,166	13
	Mec- Single	50	8
	Mec- Multi	41	8

Test Reference Year (TRY) weather files from the Chartered Institution of Building Services Engineers (CIBSE) were used in this study [58]. TRY files describe typical weather conditions based on 30-year measurements (1984–2013) in 13 cities around the UK and are used for assessing Building Regulations compliance. It is noted that weather data collected in urban locations may not represent the climate for schools at locations of suburban and rural areas. These weather files have been applied to the overall school stock based on the degree-day regions defined in the CIBSE methodology, as noted previously [55]. Table 6 shows the list of climate regions and the CIBSE TRY weather files that were used to represent these regions. It should be noted that the reference to 'Wales' corresponds with schools that have been matched to the Wales climate region but are still physically located within England.

4. Results analysis and interpretation

The section below details the building performance simulation results and the comparison against the measured data. The analysis starts with a wide overview and performs a stock-level analysis, describing each archetype and the number of schools they represent across the country. Next, a high-level cross-country comparison between the simulated and measured energy performance is presented. Last, an analysis at a climate-zone is presented (London), to enable a more detailed investigation of the school stock at a more confined area.

4.1. Stock-level analysis – Archetype description

Table 7 demonstrates the breakdown of the school archetype models and the number of actual buildings they represent at the national level. The data show that most schools in the sample are naturally ventilated (9,279 and 272, respectively, or 2.9% of the entire English primary school stock). Although most schools in the stock include building extensions, an appreciable number of school buildings do not have an extension (1,664 schools with no extension, compared to 7,887 schools with extension, or 21.1% of the entire English primary school stock). Assuming that some of the schools that had never been extended will require additional space at some point, this leaves room for potential interventions that will improve the overall performance of the individual schools, as well as that of the stock as a whole. Last, the summary table reveals that 81% of the sample (7,793 buildings) was built before 1976, and only 1,758 built post that year. Despite the large sample size (9,551 represents half of all English primary schools at the time of PDSP), some biases will exist, reflecting the underlying data. Notably, as outlined previously, schools constructed or significantly refurbished in recent years were excluded from PDSP, so the sample will under-represent more recently built schools.

4.2. Fossil-thermal energy consumption breakdown

Once the Archetype models had been generated, the entire English primary school stock was simulated and its simulated energy performance was compared to measured data, based on the DEC database of those schools. Table 8 presents the simulated and measured fossil-thermal energy consumption for each archetype. While most of the climate regions in Table 8 show a good match between the measured and simulated fossil-thermal energy consumption, some areas show larger differences (e.g., Plymouth, Cardiff and Norwich). It is suggested that these discrepancies are associated with the small number of schools in these areas. It is hypothesised that an archetype-based building stock model, such as the one developed in this study, may perform less well in areas where the sample

Table 8
A breakdown of the simulated and measured fossil-thermal (kWh/m²/y) performance, by climate region, construction period and building characteristics.

Period	Vent	Block	London Thames Valley		London South east		Southampton Southern		Plymouth South-west		Swindon Severn		Birmingham Midland		Manchester West Pennins		Newcastle 1 North west		Newcastle 2 Borders		Leeds North-east		Nottingham East Pennins		Norwich East Anglia		Cardiff Wales	
			Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr	Sim	Msr
Pre-1919	Nat	Single	123	138	131	115	133	114	122	80	146	94	157	–	144	152	–	–	156	190	162	156	153	147	150	122	–	–
		Multi	131	139	136	130	140	120	125	79	149	111	162	136	149	137	164	139	163	133	155	151	161	148	163	125	157	122
	Mec	Single	128	92	–	–	–	–	–	–	136	162	–	97	152	92	–	–	–	–	–	–	–	–	156	47	–	–
		Multi	125	132	–	–	141	103	–	–	142	117	–	–	–	–	–	–	–	–	–	–	160	147	160	119	–	–
Inter war	Nat	Single	134	151	136	115	159	86	141	74	173	122	–5	–	168	–	186	–	166	137	155	138	179	146	173	126	164	126
		Multi	135	147	135	125	154	136	134	80	152	123	173	147	–	141	–	137	189	130	173	142	176	149	169	122	–	–
	Mec	Single	130	79	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
		Multi	127	111	–	–	133	218	–	–	143	99	159	130	–	–	–	–	–	–	–	–	–	–	–	–	–	–
1945-1967	Nat	Single	126	135	125	116	131	106	118	78	135	114	148	–	144	–	158	–	152	158	147	153	149	148	143	124	138	137
		Multi	127	136	127	122	133	113	118	85	141	124	135	133	–	133	–	118	156	152	–	–	153	146	146	133	–	–
	Mec	Single	116	128	–	–	–	–	119	66	130	83	–	–	136	144	–	–	134	150	–	–	–	–	–	–	–	–
		Multi	122	116	–	–	124	148	–	–	129	109	143	119	–	–	–	–	–	–	132	96	142	142	138	122	–	–
1967-1976	Nat	Single	122	137	122	117	130	116	115	70	136	102	143	–	139	–	153	137	147	137	142	141	142	148	141	120	142	78
		Multi	123	137	123	121	129	116	115	81	137	104	–	127	–	123	–	–	151	159	143	136	149	139	140	117	–	–
	Mec	Single	117	115	–	–	124	140	112	37	127	109	–	–	–	–	–	–	0	–	133	110	135	131	–	–	–	–
		Multi	117	135	124	83	126	119	–	–	132	105	–	–	133	109	–	–	–	–	137	124	142	188	130	117	–	–
Post-1976	Nat	Single	115	130	116	111	121	106	108	80	127	107	135	–	132	–	141	–	137	152	132	146	136	140	131	117	129	129
		Multi	118	131	118	117	123	94	110	77	131	112	129	117	–	126	–	138	141	131	134	147	139	142	134	119	–	–
	Mec	Single	106	150	110	100	116	115	102	116	121	114	–	–	–	–	–	–	120	90	117	157	–	–	128	180	–	–
		Multi	113	110	116	163	118	115	102	89	120	91	–	116	–	–	–	–	–	–	125	106	135	178	–	–	–	–

*Nat = Naturally ventilated building. Mec = Mechanically ventilated building. Single = Single block. Multi = Multi block. Sim = Simulated. Msr = Measured through DEC – Display Energy Certificate.

size is particularly small, as variations in the characteristics and energy use of individual buildings may dominate.

Figures 6-7 and Table 9 compare the average mean simulated and measured energy consumption, for fossil-thermal energy use, for each archetype. Since each archetype represents a number of schools, the simulated energy consumption of each archetype was compared to the average measured consumption (based on DEC) of the schools represented by the archetype. Values were calculated by summing up the measured energy consumption of all schools for each archetype (kWh/m^2) and dividing it by the total number of buildings. These figures, therefore, express the average energy consumption, both simulated and measured, of each archetype across the country.

Figure 6 shows that archetypes of the later eras (1945–1967, 1967–1976 and post 1976) have a smaller variation between simulated and measured consumption than the pre 1919 and interwar models. The difference between the average simulated and measured, by archetype, is 18.8% (pre 1919), 22.3% (inter war), 10.8% (1945–1967), 8.9% (1967–1976) and 3.2% (post 1976).

Figure 7 and Table 9 show the relationship between the average simulated and measured fossil-thermal energy performance, broken down by construction age band. This shows that DREAMS performs better for schools constructed at later eras (e.g., 1945–1966, 1967–1977 and post 1977); the difference between simulated and measured fossil-thermal energy consumption is lower for these schools compared to earlier construction eras (pre-1919 and Inter war). This is potentially due to the difference between assumed and actual build-ups: while the assumed modelled build-ups consider that the buildings have not gone through any thermal-related improvements, it is likely that many of the pre 1919 and inter war schools will have undergone refurbishments to some degree and that their building envelope's thermal performance has, therefore, improved.

Figure 8 and Tables 10–11 compare the simulated and measured consumption figures based on ventilation strategy – natural and mechanical ventilation. The results show that while the naturally-ventilated simulations have managed a good level of accuracy (between 1 and 10% difference between predicated and
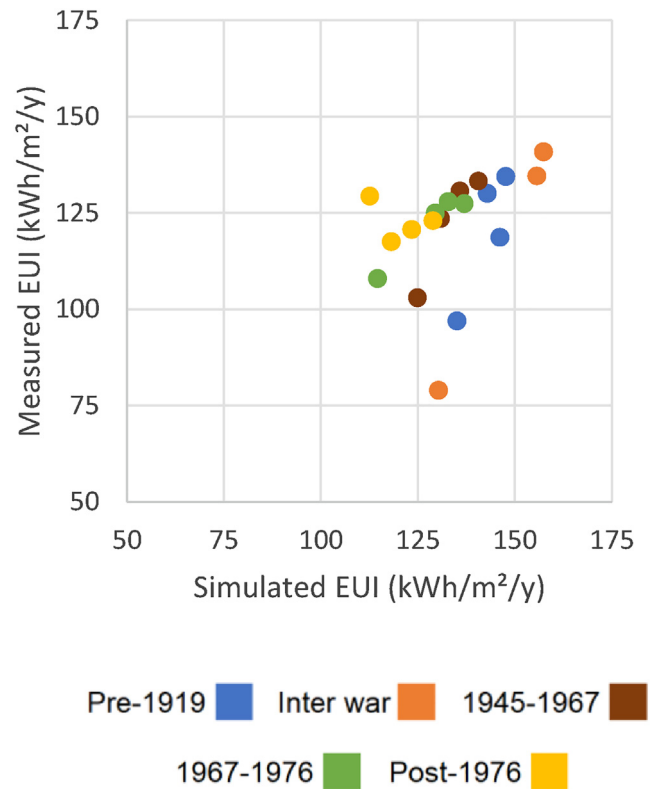

Fig. 7. The relationship between simulate and measured fossil-thermal performance in archetypes primary schools in England, by construction era.

measured figures), the predictions for the mechanically-ventilated schools show a wider range of differences – between 0% and 26%.

It is important to point out, nevertheless, that mechanically ventilated schools account for only 2.9% (272 schools in total) of the entire sample. This suggests that for 97% of the school stock
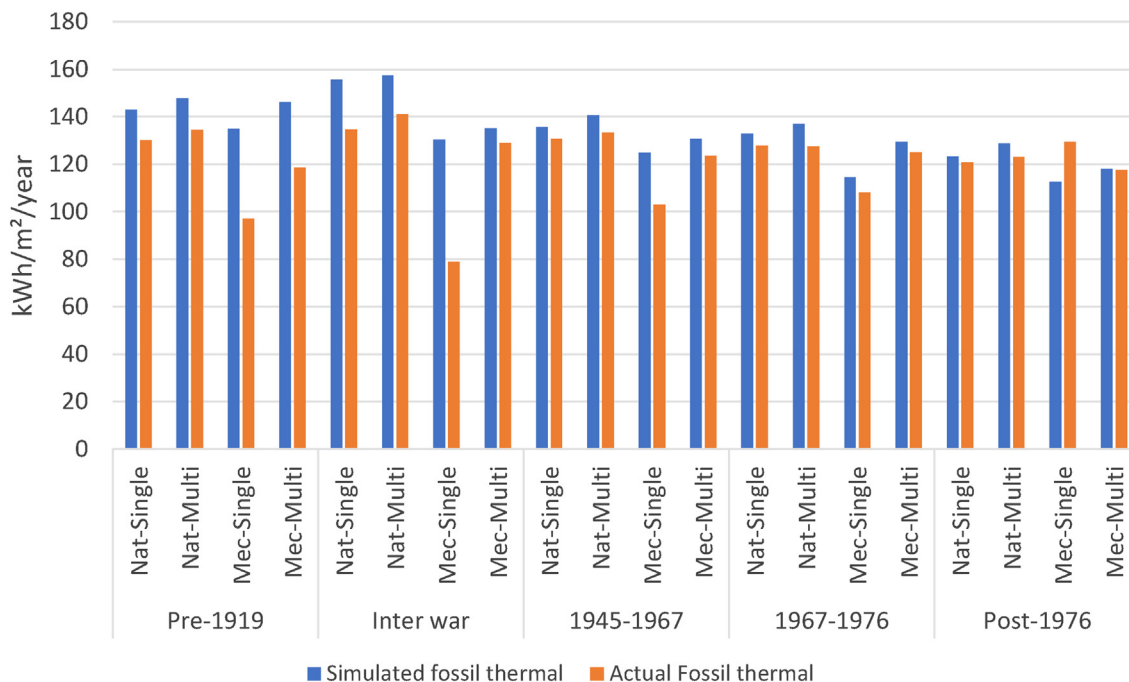

Fig. 6. A comparison between simulated and measured fossil-thermal (gas) energy consumption for the primary school stock in England. * Nat = Naturally ventilated building. Mec = Mechanically ventilated building. Single = Single block. Multi = Multi block.

Table 9
Average simulated and measured fossil-thermal energy consumption of primary schools in England, by construction era.

Construction Era	Simulated consumption (kWh/m²/y)	Measured consumption (kWh/m²/y)	Difference	Sample size (n)
Pre-1919	147	134	10%	2,283
Inter war	157	139	13%	791
1945–1967	140	133	5%	2,116
1967–1976	136	127	7%	2,603
Post-1976	127	123	3%	1,758
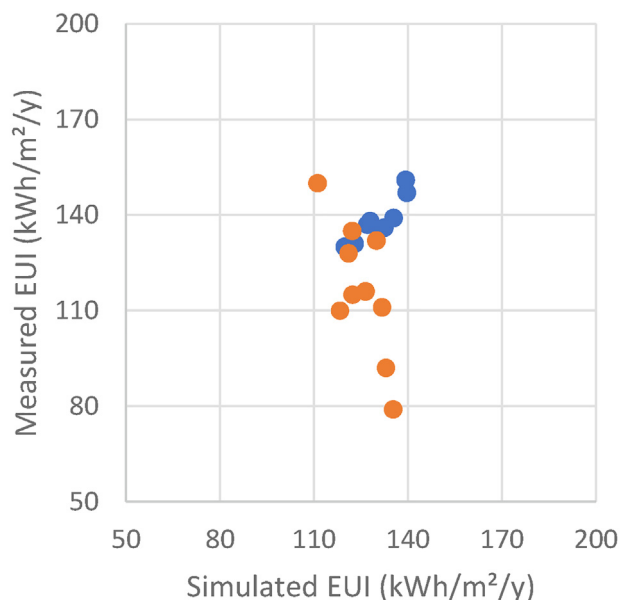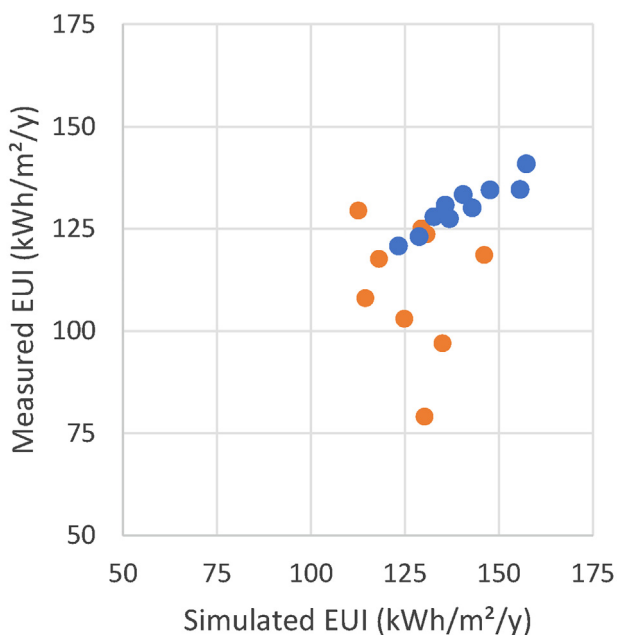
Fig. 8. Simulated and measured fossil-thermal performance in primary schools archetypes in England, by ventilation strategy.

Fig. 9. Simulate and measured fossil-thermal performance in primary schools in London, by ventilation strategy.

(the blue entries in Fig. 8), modelling achieves a high level of prediction accuracy. It is suggested that the lower agreement between simulated and measured consumption in mechanically ventilated schools is associated with the small sample size (272 entries, split into 10 sub-categories, based on the classification procedure as described in Section 3). Notwithstanding these preliminary results,

further investigation needs to be undertaken to understand and improve the predictions for those schools.

Overall, results show that the average primary school in England is predicted to consume 139 kWh/m²/y of fossil-thermal energy, compared to 130 kWh/m²/y of measured consumption.

Table10
Average simulated and measured fossil-thermal energy consumption for **naturally ventilated** primary schools in England, by construction era.

Construction Era	Simulated consumption (kWh/m²/y)	Measured consumption (kWh/m²/y)	Difference	Sample size (n)
Pre-1919	147	134	10%	2,245
Inter war	157	140	12%	778
1945–1967	140	133	5%	2,059
1967–1976	136	128	7%	2,530
Post-1976	127	122	4%	1,667

Table 11
Average simulated and measured fossil-thermal energy consumption for **mechanically ventilated** primary schools in England, by construction era.

Construction Era	Simulated consumption (kWh/m²/y)	Measured consumption (kWh/m²/y)	Difference	Sample size (n)
Pre-1919	145	116	25%	38
Inter war	134	121	11%	13
1945–1967	130	121	7%	57
1967–1976	127	122	4%	73
Post-1976	115	124	7%	91

This indicates that the simulated results predicted 7% higher energy consumption than measured figures.

4.3. Analysis of the London climate region

DREAMS enables a detailed investigation of the stock-level performance at a climate-region level. This section presents an example of such an analysis, for London. Fig. 9 and Tables 12 and 13 compare the simulated and measured fossil-thermal energy consumptions for London's primary schools. While most archetypes have achieved between 3 and 7% difference between simulated and measured figures, it is noted that, similarly to the analysis of the UK as a whole (Table 10), three archetypes have performed worse (12, 17 and 41%). However, these are each mechanically ventilated buildings, and only represent a total of 50 schools out of the 2,206 London-area schools in the database.

Table 14 shows a comparison between the DREAMS' simulated and measured fossil-thermal energy consumption for the primary school stock for London. This demonstrates that there is a greater difference between simulated and measured performance in archetypes, where the sample size is small.

5. Discussion

The development of the DREAMS framework for the energy performance prediction of the English school building stock was presented in the previous sections. Comparison of the simulated heating consumption with measured consumption figures show that the model performs satisfactorily: Simulated heating energy consumption for a typical primary school in England is only 7% higher than the measured consumption (139 kWh/m²/y simulated, compared to 130 kWh/m²/y measured). While no previous studies exist that use an archetype-based approach to modelling the school stock, this result is comparable to the discrepancies between measured and simulated data in equivalent residential stock modelling (Famuyibo et al. [21]– 8%, Caputo et al. [32] – 4%, Ren et al. [16]- 9.2%, Stefanović et al. [59]- 2,2%).

The study shows that, while the overall difference between simulated and measured energy use for the stock is small, there is a relationship between the school characteristics and model accuracy. For example, older schools have larger differences between predicted and measured consumption, than more recently built schools.

Table12
Average simulated and measured fossil-thermal energy consumption for **naturally ventilated** primary schools in London, by construction era.

Construction Era	Simulated consumption (kWh/m²/y)	Measured consumption (kWh/m²/y)	Difference	Sample size (n)
Pre-1919	132	139	5%	477
Inter war	140	149	6%	233
1945–1967	132	136	3%	511
1967–1976	128	137	7%	543
Post-1976	121	131	7%	358

Table 13
Average simulated and measured fossil-thermal energy consumption for **mechanically ventilated** primary schools in London, by construction era.

Construction Era	Simulated consumption (kWh/m²/y)	Measured consumption (kWh/m²/y)	Difference	Sample size (n)
Pre-1919	131	112	17%	11
Inter war	134	95	41%	7
1945–1967	124	122	1%	18
1967–1976	122	125	2%	16
Post-1976	115	130	12%	32

Table 14
A comparison between simulated and measured fossil-thermal energy consumption for the primary school stock in London. * Nat = Naturally ventilated building. Mec = Mechanically ventilated building. Single = Single block. Multi = Multi block.

Building type	School description	Simulated fossil thermal (kWh/m²/y)	Measured Fossil thermal(kWh/m²/y)	Difference (%)	Sample size (n)
Pre-1919	Nat-Single	128	138	7%	34
	Nat-Multi	136	139	3%	443
	Mec-Single	133	92	45%	3
	Mec-Multi	130	132	2%	8
Inter war	Nat-Single	139	151	8%	26
	Nat-Multi	140	147	5%	207
	Mec-Single	135	79	71%	2
	Mec-Multi	132	111	19%	5
1945–1967	Nat-Single	131	135	3%	84
	Nat-Multi	132	136	3%	427
	Mec-Single	121	128	5%	2
	Mec-Multi	127	116	9%	16
1967–1976	Nat-Single	127	137	7%	98
	Nat-Multi	128	137	6%	445
	Mec-Single	122	115	6%	2
	Mec-Multi	122	135	9%	14
Post-1976	Nat-Single	120	130	8%	120
	Nat-Multi	123	131	6%	238
	Mec-Single	111	150	26%	17
	Mec-Multi	118	110	8%	15

The study has further shown that a larger difference between simulated and measured performance was noted in naturally ventilated buildings (compared to mechanically-ventilated ones), but also noted that such schools account for less than 3% of the stock. Lastly, the paper showcased DREAMS' performance at a climate-region scale (London). It was shown that the performance of the model for the London region is broadly similar to its performance at national level.

Through the analysis of the modelling and simulation outputs, a few issues were identified to potentially improve the model's outputs as part of future and ongoing work:

5.1. Generalisation

Generalisation of building properties, which is a fundamental principle in stock modelling, creates some limitations. These may reflect a range of factors, including the difficulty of accessing measured building energy use, a lack of knowledge of building thermal characteristics and an inability of models to accurately reflect user behaviour [31].

5.2. Sample size

The archetype models were defined based on an analysis of existing data on real school buildings. Through data filtering and grouping, schools were aggregated into groups with shared characteristics. In some cases, these groups represent a very small number of schools. The analysis shows that the smaller the sample size, the higher the likelihood of a larger difference between simulated and measured fossil-thermal energy consumption.

5.3. Mechanically-ventilated schools

These schools have shown the greatest differences between simulated and measured space heating energy performance. While mechanically ventilated schools currently represent only around 3% of the entire stock, it is possible that mechanical systems will become more common in schools in the future. Consequently, the relative impact of these systems on the stock may be more significant.

5.4. Data availability

Though the archetype models have been based on databases of actual buildings and their measured energy consumption, additional data could further improve the model's accuracy in predicting actual performance. In particular, detailed construction build-up data, internal loads and usage patterns profile may be especially important [9].

Despite the current limitations in the DREAMS framework, this study shows promising levels of resemblance between predicted and measured consumption on a stock level. It is suggested that a framework such as DREAMS could be useful in examining a range of performance-related stock-level proxies in schools, both on a national and regional levels. This could potentially help inform decision and policy makers in recommendations for stock-level school performance-related policies, such as:

- Evaluating schools' performance under difference climate change scenarios
- Examining different schools' refurbishment packages and their suitability on a regional-scale
- Evaluating stock-level Indoor Air Quality and pupils' cognitive performance
- Investigating power-generation in school premises and the potential for Net Zero schools

It is noted that as the proposed methodology will used in a number of modelling applications, further developments would be needed to account for the specific requirements and unique features of each application. This may include domains such as occupancy-based demand response and comfort optimisation, the use of microgrids with renewable energy sources and energy storage, energy and comfort management and others.

6. Conclusions

This paper presented the principles underlying the development of the Data dRiven Engine for Archetype Models of Schools (DREAMS), a novel, data-driven, archetype-based school building stock modelling framework, which was developed within the EPSRC-funded ASPIRE project. To the knowledge of the authors, this is the first time that detailed statistical analysis of two national, large scale building stock databases (PDSP and DEC) is performed in order to create a combination of theoretical and sample archetypes that are statistically representative of the English primary school building stock. Data analysis, category definition and archetype development procedures were outlined in detail. Using the DREAMS framework, 168 school building archetype models were developed that broadly represent 9,551 primary schools in England.

Following the development of the theoretical archetypes, building energy and thermal performance simulation was carried out in EnergyPlus for the baseline English primary school building stock using 13 regional CIBSE TRY weather files.

- DREAMS' simulated performance achieved good agreement with measured energy consumption (only 7% difference).
- The study showed that there was a higher agreement between predicted and measured heating consumption in the archetypes of newer schools than in those of older ones: Simulated heating energy consumption of post war schools archetypes had achieved between 3 and 7 % higher figures than measured consumption, while pre-war schools archetypes have reached between 10 and 13% difference.
- The study showed that simulated consumption figures of naturally ventilated school archetypes – which account for 97% of the stock – were only 4–12% higher than measured figures, while the 3% mechanically ventilated school archetypes achieved between 4 and 25% difference. While the later can be considered as a significant difference, the small number of mechanically ventilated buildings (less than 3% of the stock) means this should have minimal impact on a stock-level. Still, the study concludes that further investigation should be carried to explore the reasons for the discrepancy between simulated and measured consumption figures in mechanically ventilated buildings.
- Overall, this study finds that more detailed data, with regards to the description of the school building stock, as well as the way it is applied, could further improve the accuracy of the simulated archetypes in predicting actual performance.

Future work should be focused on improving the accuracy of the model's performance predictions. DREAMS was built as a flexible framework that can be adapted to process more granulated data inputs. As part of the ASPIRE project, further work is also planned on retrieving updated and more extensive school stock data. The platform will also be expanded to model secondary schools and predict indoor overheating and air quality levels, under current and future climate change scenarios.

Furthermore, stakeholder engagement can help identifying the potential for model improvements specifically, for testing retrofit

strategies, estimating indoor overheating and indoor air quality, and for estimating energy consumption under current and future climate change scenarios.

CRediT authorship contribution statement

Y. Schwartz: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing - original draft. **D. Godoy-Shimizu:** Data curation, Writing - original draft. **I. Korolija:** Conceptualization. **J. Dong:** Writing - original draft. **S.M. Hong:** Data curation, Writing - original draft. **A. Mavrogianni:** Funding acqui-sition, Writing - review & editing. **D. Mumovic:** Funding acquisi-tion, Supervision, Writing - review & editing.

Declaration of Competing Interest

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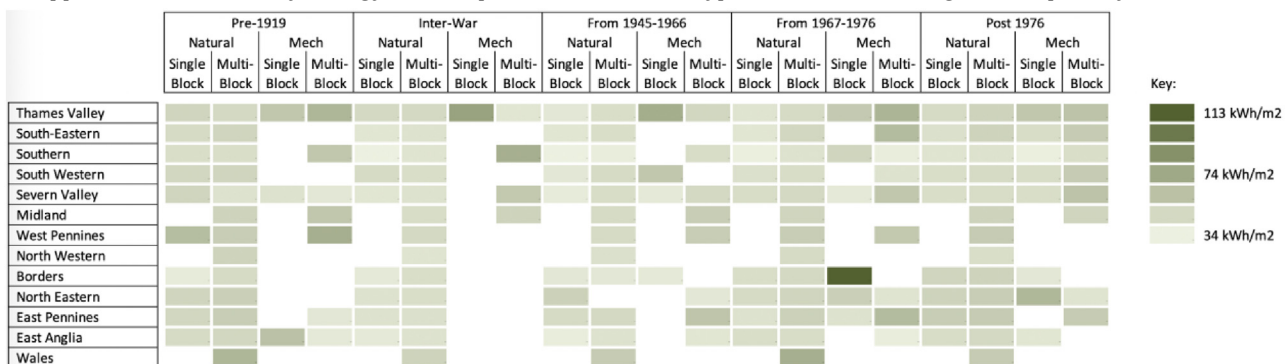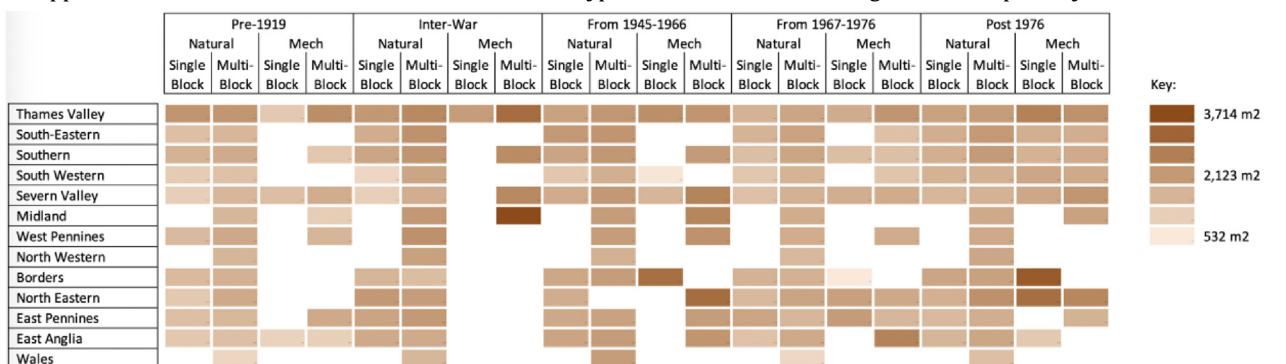
Acknowledgment

Appendix

Appendix A – Fossil-thermal energy consumption of each archetype across climate regions in primary schools across England
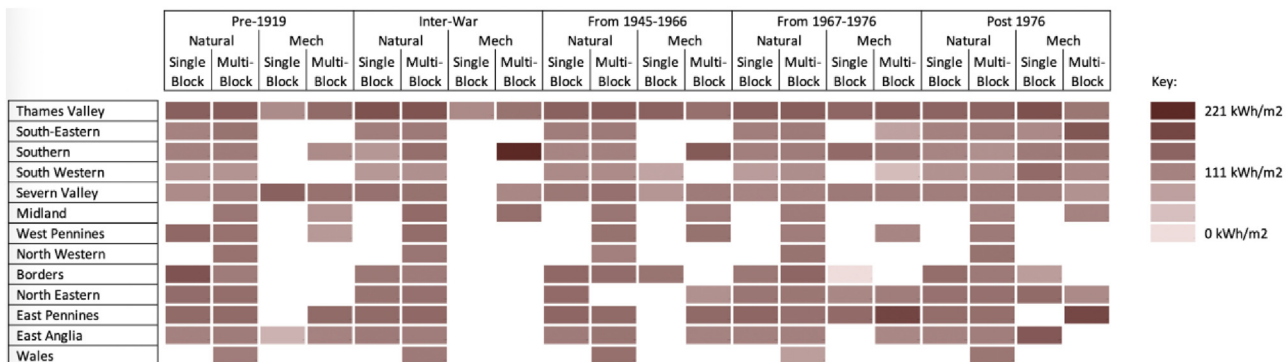
Appendix B – Electricity energy consumption of each archetype across climate regions in primary schools across England

Appendix C – Floor area of each archetype across climate regions in primary schools across England

References

[1] C. Schweizer et al., Indoor time–microenvironment–activity patterns in seven regions of Europe, J. Eposure Sci. Environ. Epidemiol. 17 (2) (2007) 170–181.

[2] E. Csobod et al., SINPHONIE Schools Indoor Pollution and Health Observatory Network in Europe Final Report. Co-published by the European Commission's Directorates General for Health and Consumers and Joint Research Centre, Luxembourg, (2014).

[3] R. Becker, I. Goldberger, M. Paciuk, Improving energy performance of school buildings while ensuring indoor air quality ventilation, Build. Environ. 42 (9) (2007) 3261–3276.

[4] UKCP18, UKCP18 Science Overview Report. (2019). https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf.

[5] D.P. Jenkins, A.D. Peacock, P.F.G. Banfill, Will future low-carbon schools in the UK have an overheating problem?, Build Environ. 44 (3) (2009) 490–501, https://doi.org/10.1016/j.buildenv.2008.04.012.

[6] EFA, Property data survey programme: Summary report. (2015), Education Funding Agency: London, UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/402138/PDSP_Summary_Report.pdf.

[7] A. Montazami, M. Gaterell, F. Nicol, A comprehensive review of environmental design in UK schools: history, conflicts and solutions, Renew. Sustain. Energy Rev. 46 (2015) 249–264, https://doi.org/10.1016/j.rser.2015.02.012.

[8] L. Chatzidiakou, D. Mumovic, J. Dockrell, The effects of thermal conditions and indoor air quality on health, comfort and cognitive performance of students. (2014).

[9] T.C. Trust. Energy efficiency and carbon saving advice for schools, in The Carbon Trust, London. (2018).

[10] M. Kavgic et al., A review of bottom-up building stock models for energy consumption in the residential sector, Build. Environ. 45 (7) (2010) 1683–1697, https://doi.org/10.1016/j.buildenv.2010.01.021.

[11] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (8) (2009) 1819–1835.

[12] W. Li et al., Modeling urban building energy use: a review of modeling approaches and procedures, Energy 141 (2017) 2445–2457, https://doi.org/10.1016/j.energy.2017.11.071.

[13] M. Brøgger, K.B. Wittchen, Estimating the energy-saving potential in national building stocks–A methodology review, Renew. Sustain. Energy Rev. 82 (2018) 1489–1496.

[14] W. Tian, R. Choudhary, A probabilistic energy model for non-domestic building sectors applied to analysis of school buildings in greater London, Energy Build. 54 (2012) 1–11, https://doi.org/10.1016/j.enbuild.2012.06.031.

[15] A.A. Famuyibo, A. Duffy, P. Strachan, Developing archetypes for domestic dwellings—An Irish case study, Energy Build. 50 (2012) 150–157, https://doi.org/10.1016/j.enbuild.2012.03.033.

[16] Z. Ren, P. Paevere, C. McNamara, A local-community-level, physically-based model of end-use energy consumption by Australian housing stock, Energy Policy 49 (2012) 586–596, https://doi.org/10.1016/j.enpol.2012.06.065.

[17] X. Li et al., Low carbon heating and cooling of residential buildings in cities in the hot summer and cold winter zone - A bottom-up engineering stock modeling approach, J. Cleaner Prod. 220 (2019) 271–288, https://doi.org/10.1016/j.jclepro.2019.02.023.

[18] M. Olonscheck, A. Holsten, J.P. Kropp, Heating and cooling energy demand and related emissions of the German residential building stock under climate change, Energy Policy 39 (9) (2011) 4795–4806, https://doi.org/10.1016/j.enpol.2011.06.041.

[19] S. Petrovic, K. Karlsson. Model for determining geographical distribution of heat saving potentials in Danish building stock. 3(1) 2014 143-165.

[20] D. Wang et al., CESAR: A bottom-up building stock modelling tool for Switzerland to address sustainable energy transformation strategies, Energy Build. 169 (2018) 9–26, https://doi.org/10.1016/j.enbuild.2018.03.020.

[21] L. Filogamo et al., On the classification of large residential buildings stocks by sample typologies for energy planning purposes, Appl. Energy 135 (2014) 825–835, https://doi.org/10.1016/j.apenergy.2014.09.002.

[22] A.T. Booth, R. Choudhary, D.J. Spiegelhalter, Handling uncertainty in housing stock models, Build. Environ. 48 (2012) 35–47, https://doi.org/10.1016/j.buildenv.2011.08.016.

[23] É. Mata, A. Sasic Kalagasidis, F. Johnsson, Energy usage and technical potential for energy saving measures in the Swedish residential building stock, Energy Policy 55 (2013) 404–414, https://doi.org/10.1016/j.enpol.2012.12.023.

[24] G. Kazas, E. Fabrizio, M. Perino, Energy demand profile generation with detailed time resolution at an urban district scale: a reference building approach and case study, Appl. Energy 193 (2017) 243–262, https://doi.org/10.1016/j.apenergy.2017.01.095.

[25] T. Loga, B. Stein, N. Diefenbach, TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable, Energy Build. 132 (2016) 4–12.

[26] R. Choudhary, Energy analysis of the non-domestic building stock of Greater London, Build. Environ. 51 (2012) 243–254, https://doi.org/10.1016/j.buildenv.2011.10.006.

[27] I. Korolija et al., UK office buildings archetypal model as methodological approach in development of regression models for predicting building energy consumption from heating and cooling demands, Energy Build. 60 (2013) 152–162, https://doi.org/10.1016/j.enbuild.2012.12.032.

[28] G. Happle, J.A. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, Energy Build. 174 (2018) 276–292, https://doi.org/10.1016/j.enbuild.2018.06.030.

[29] J. Sokol, C. Cerezo Davila, C.F. Reinhart, Validation of a Bayesian-based method for defining residential archetypes in urban building energy models, Energy Build. 134 (2017) 11–24, https://doi.org/10.1016/j.enbuild.2016.10.050.

[30] H. Lim, Z.Q. Zhai, Review on stochastic modeling methods for building stock energy prediction, Build. Simul. 10 (5) (2017) 607–624, https://doi.org/10.1007/s12273-017-0383-y.

[31] C.F. Reinhart, C. Cerezo Davila, Urban building energy modeling – A review of a nascent field, Build. Environ. 97 (2016) 196–202, https://doi.org/10.1016/j.buildenv.2015.12.001.

[32] P. Caputo, G. Costa, S. Ferrari, A supporting method for defining energy strategies in the building sector at urban scale, Energy Policy 55 (2013) 261–270, https://doi.org/10.1016/j.enpol.2012.12.006.

[33] A. Macmillan et al., Integrated decision-making about housing, energy and wellbeing: a qualitative system dynamics model, Environ. Health 15 (1) (2016) S37, https://doi.org/10.1186/s12940-016-0098-z.

[34] A. Mavrogianni et al., The impact of occupancy patterns, occupant-controlled ventilation and shading on indoor overheating risk in domestic environments, Build. Environ. 78 (2014) 183–198, https://doi.org/10.1016/j.buildenv.2014.04.008.

[35] E. Oikonomou et al., Modelling the relative importance of the urban heat island and the thermal quality of dwellings for overheating in London, Build. Environ. 57 (2012) 223–238, https://doi.org/10.1016/j.buildenv.2012.04.002.

[36] J. Taylor et al., Understanding and mitigating overheating and indoor PM2.5 risks using coupled temperature and indoor air quality models, Build. Serv. Eng. Res. Technol. 36 (2) (2015) 275–289, https://doi.org/10.1177/0143624414566474.

[37] S. Grafakos et al., Integration of mitigation and adaptation in urban climate change action plans in Europe: a systematic assessment, Renew. Sustain. Energy Rev. 121 (2020), https://doi.org/10.1016/j.rser.2019.109623 109623.

[38] U.S. DoE, Buildings energy data book, Energy Efficiency Renew. Energy Dept. 286 (2011).

[39] DoE, U., EnergyPlus energy simulation software v. 3-1-0. US DoE EERE. https://github.com/NREL/EnergyPlus/releases/tag/v8.9.0.

[40] EFA, Property data survey programme: Survey manual part 1 Overview and methodology. (2014), Education Funding Agency: London, UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/276415/pdsp_survey_manual_part_1_a4_revision_20_v9_final_access.pdf.

[41] DCLG, Improving the Energy Efficiency of Our Buildings. A Guide to Display Energy Certificates and Advisory Reports for Public Buildings. (2015), Department for Communities and Local Government: London, UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/452481/DEC_Guidance__rev_July_2015_.pdf.

[42] D. Godoy-Shimizu, D., et al., Using Display Energy Certificates to quantify schools' energy consumption. Building Research & Information, (2011). 39(6): p. 535-552

[43] S.-M. Hong et al., Improved benchmarking comparability for energy consumption in schools, Build. Res. Inf. 42 (1) (2014) 47–61.

[44] S. Hong et al., Assessing the trends of energy use of public non-domestic buildings in England and Wales, Build. Serv. Eng. Res. Technol. 40 (2) (2019) 176–197.

[45] SAS 9.4. (2020) Available from: https://support.sas.com/software/94/index.html.

[46] P. Steadman. Building types and built forms. 2014, Kibworth Beauchamp, Leicestershire: Kibworth Beauchamp, Leicestershire: Matador.

[47] Google maps. (2020). https://www.google.co.uk/maps/.

[48] Bing maps. (2000). https://www.bing.com/maps.

[49] Buildings, B.a., The Building Regulations 1965. (1965).

[50] Buildings, B.a., The Building Regulations 1976. (1976).

[51] H. Government. Approved Document L2A: conservation of fuel and power in new buildings other than dwellings, 2013 edition with 2016 amendments. (2016): Newcastle Upon Tyne. https://www.gov.uk/government/publications/conservation-of-fuel-and-power-approved-document-l.

[52] Agency, E.S.F., BB 101: Ventilation, thermal comfort and indoor air quality. (2018). https://www.gov.uk/government/publications/building-bulletin-101-ventilation-for-school-buildings.

[53] The Charted Institution of Building Services Engineers London, CIBSE Guide-A: Environmental Design. (2006), CIBSE Publications: Norwich. ISBN-10: 1-903287- 66-9, ISBN-13: 978-1-903287-66-8.

[54] NCM, National Calculation Methodology (NCM) modelling guide (for buildings other than dwellings in England). (2016).

[55] CIBSE, TM46: Energy benchmarks. (2008), CIBSE London.

[56] BEIS, Domestic National Energy Efficiency Data – Framework (NEED) Methodology. (2020). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/895139/Domestic_NEED_Methodology.pdf.

[57] Python. (2020). https://www.python.org/.

[58] CIBSE, CIBSE UK Weather Data Sets. (2020). https://www.cibse.org/weatherdata.

[59] A. Stefanović, D. Gordić, Modeling methodology of the heating energy consumption and the potential reductions due to thermal improvements of staggered block buildings, Energy Build. 125 (2016) 244–253, https://doi.org/10.1016/j.enbuild.2016.04.058.