Accommodating new calculation approaches in next-generation energy performance assessments

Citation for published version:

Digital Object Identifier (DOI):
10.1080/19401493.2024.2307634

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Publisher's PDF, also known as Version of record

Published in:
Journal of Building Performance Simulation

Publisher Rights Statement:
© 2024 The Author(s).

General rights
Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Accommodating new calculation approaches in next-generation energy performance assessments

David Jenkins, Peter McCallum, Sandhya Patidar & Sally Semple

To cite this article: David Jenkins, Peter McCallum, Sandhya Patidar & Sally Semple (24 Jan 2024): Accommodating new calculation approaches in next-generation energy performance assessments, Journal of Building Performance Simulation, DOI: 10.1080/19401493.2024.2307634

To link to this article: https://doi.org/10.1080/19401493.2024.2307634

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 24 Jan 2024.

Submit your article to this journal

Article views: 65

View related articles

View Crossmark data


**Accommodating new calculation approaches in next-generation energy performance assessments**

David Jenkins, Peter McCallum, Sandhya Patidar and Sally Semple

School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, UK

**ABSTRACT**

Building energy policy, such as the Energy Performance of Buildings Directive (EPBD), has a direct effect on use of building models and related parameters. Modelling annual energy consumption of a building is a different task to characterizing the demand of that building at a transient level; to do so at scale requires additional complexity. With the ubiquity of Energy Performance Certificates (EPC) across Europe, there is a tendency to use these to communicate building energy demand to policy. However, there is growing evidence of EPCs being applied to areas which they were not designed to serve. By comparing alternative techniques with current methodologies, this study proposes future directions for standardized energy assessment of dwellings, proposing a framework for critiquing such techniques. New methods are formulated that make use of simulation and statistical techniques developed by the authors, and allow for urban-scale modelling that is consistent with traditional energy assessment.

**Key innovations**

- An approach for critiquing new forms of energy assessment in the built environment
- A review of building energy assessments that are cognisant of latest developments of building modelling and new requirements of such models
- New simulation and modelling techniques for next-generation energy assessment of buildings

**Practical implications**

The proposed framework suggests a new way of designing standardized energy assessments, in context of the EPBD. Using some of the proposed approaches, this allows for end-users of EPCs to have actionable information on crucial areas of building and energy system design, where traditional EPCs are not able to do so.

**1. Introduction**

Arguably, the EPBD (European Commission 2002) placed building modelling at the centre of low-carbon building policy in Europe. However, the need to achieve market transformation at scale has resulted in quite simplified forms of building modelling being used to assess our buildings. Whilst this may have been appropriate for a compliance tool focussed on energy efficiency, at a time where the spectrum of modelling tools may have been narrower, new challenges (in energy system design) and opportunities (in software/hardware and data) raise questions about whether this approach is still fit for purpose.

This paper uses modelling of the authors and elsewhere to illustrate new opportunities for simulation and statistical modelling within the broader area of energy compliance of buildings. It places more recent developments of building modelling, both physical and empirically-based, into the context and requirements of EPCs, noting that the questions we are asking of EPCs are not necessarily the same as in the dawn of the EPBD, 20 years ago. The role of more advanced building simulation is therefore discussed, and an approach proposed for re-evaluating what we need from standardized energy assessments.

Rather than make recommendations of what EPCs should be, the paper notes current movements within Technical Committees and research projects relating to EPCs at European level that are already happening, and suggests techniques available to respond to these ‘next-generation EPC’ requirements – where this term is now established within that research and policy community.

With this moving picture of what constitutes an energy assessment (specifically an EPC), continuing
innovation within the building simulation community, ever-improving access to empirical data, and greater uses of (and challenges on) EPCs to reach ambitious carbon targets, there is a clear need to bring this together in an updated approach to energy assessment. This study, building on a previous conference paper (D. Jenkins et al. 2021), attempts to define what new assessment approaches might look like, how that might change our use of both simulation and data analysis (with specific examples), and how we might judge new methods of assessment within this changing picture.

2. Review of assessment
The form of assessment used for EPCs has evolved with specific applications in mind. Any review, and associated critique, of this method should therefore reflect on these applications prior to evaluating the suitability of EPCs as a vehicle for action – and whether changes are required for those EPCs to be applied for new targets and challenges; that is, EPCs should not be judged against end-uses that they were not intended to satisfy. Even within this important caveat, the following section suggests improvements that could be made and, by doing so, the need for other forms of modelling.

2.1. Standardizing energy assessment
The EPBD was introduced by the European Union in 2003 and was revised in 2010 and 2018. The recast directive (European Commission 2010) further established the requirements for the production of EPCs. In particular, the EPBD states that the ‘energy performance of a building shall be determined on the basis of the calculated or actual energy that is consumed to meet the different needs associated with its typical use’.

Within this guidance, variation already exists in terms of responses by individual countries – though still guided by the requirements of standardization within the EPBD (ISO Standards 2017). As noted by the Building Performance Institute Europe (BPIE 2014), 14 EU states use only theoretical/physical methods to assess energy performance. Other states have an option for using measured energy consumption, though different criteria were used to identify which buildings could qualify for this. A report commissioned by the EU (European Commission 2015) noted that the 28 member states of the EU had 35 calculation methodologies in place which could be used to generate EPCs for new and existing buildings. This is, perhaps, surprising for an initiative that places standardization at the forefront, and demonstrates that the EPCs are not necessarily a common currency across different parts of the EU.

One commonality is that every EU state provides at least one calculated method for assessing energy performance, whether for new or existing buildings. For existing buildings, where not all construction details are available, most EU states provide a calculated method which uses assumed values for unknown construction details. These may be separate calculation methods or variations on the calculation method used for new buildings. Generally, all techniques can be categorized as one of the below:

- calculation using detailed construction information
- calculation using assumed construction information
- assessment based on similar reference dwelling
- measured energy consumption

2.2. Identifying flaws with current approaches
When critiquing EPCs, it is important to remember what they were originally targeted towards (and whether they achieve that purpose – that is, a compliance not a design tool), whilst also noting where ‘mission creep’ exists such that EPCs are put forward for less suitable applications.

The Performance Gap, the difference between modelled and measured energy demand, is well documented (Bordass 2013; de Wilde 2014). However, there is an argument that EPCs were not designed to accurately generate real energy bill estimates and, rather, are merely the result of an energy compliance tool that allows for indicative ratings of building assets (for a typical household) to be estimated. An ‘Asset Rating’, in this context, is an estimation of the energy the building might use under some standardized conditions and with standardized occupancy (using assumptions discussed elsewhere (ISO Standards 2017)). Where this argument becomes problematic is the growing use of EPCs for applications beyond basic energy compliance, such as detailed design of buildings, structuring loan repayments for energy efficiency investments, or punitive actions on homeowners based on their energy rating.

An arguably more fundamental problem for a standardized energy assessment is that of consistency. The EPBD requires that a replicable, standardized assessment can be carried out in the same way for any building. With many EPC approaches removing, or at least dampening, the impact of household behaviour (by focussing on the building asset), some of the less standardizable aspects of residential energy use are ignored. Whilst this might imply a clear path to consistency when comparing similar buildings, studies of EPC lodgement databases suggest this is not always the case. Previous work by the authors (UK Government 2014) using multiple assessments of a small sample (29) of dwellings show different assessors making different assessments (and models) of the same
home. A study of a larger number of assessments for a single dwelling (Tronchin and Fabbri 2012) also suggested a lack of consistency emanating from the energy assessors.

Other studies (Hardy and Glew 2019) have taken larger databases of previously assessed buildings and identified more statistically robust variation, introduced concepts of measurement error in EPCs (Crawley et al. 2019), and looked at changes over time in those databases (Pasichnyi et al. 2019) that suggest quality control issues. With such existing variations, one might suggest that other forms of data collection and modelling could actually improve consistency in energy assessment rather than making it more difficult to control. This could involve the use of CityGML files (Rosser et al. 2019) to better describe the building stock, or efficiently determining the thermal characteristics of a large community of buildings that could play a role in the energy compliance process for individual buildings (McCallum et al. 2019).

There is, therefore, a suggestion that, as well as designing future energy assessment for future requirements/challenges, there are already existing problems with standardized energy assessment that could be addressed by different forms of modelling and data collection.

### 2.3. Current innovations in energy performance metrics

The full toolkit available to a building modeller to help understand energy demand is vast. Even noting some of the requirements of the EPBD, there is still a broad spectrum of responses that can satisfy this initiative. However, the EPBD itself is not static; new recommendations and innovations can be seen in various recasts of that directive. An example of this is the desire to create new metrics and indicators that an EPC could generate, where such metrics provide different forms of advice to end-users that are not currently available in existing EPCs.

Two well-documented new metrics for next-generation EPCs are the Smart Readiness Indicator (SRI) (Ramezani, Gameiro da Silva, and Simões 2021) and the Operational Energy Rating (D2EPC 2022). The SRI is designed as an overview of the ability of a building to control its energy and environmental conditions in a flexible way, and has been investigated and tested by a number of projects across Europe (such as X-tendo (Zuhaib 2020), TIMEPAC (TIMEPAC 2023), and ePANACEA (Borragan and Legon 2021)). Rather than being a detailed, technical, thermophysical assessment of the building, relatively simple information (e.g. existence of smart control systems, listing presence of certain technologies etc) is collected and a rating (or score) calculated that indicates whether the building would be a good candidate for grid flexibility services (though that is not the only ambition of the indicator). Therefore, though one might consider demand flexibility planning to require higher resolution data about demand profiles, the SRI is very much a top-level indicator that does not (at this ambition) require access to other forms of physical modelling. It does, however, raise a question about whether, to have more meaningful assessments of demand flexibility, other forms of modelling should be integrated into this or other approaches.

The Operational Energy Rating is a proposed, standardized, way of accounting for real energy consumption data of a building. With the vast majority of EPCs based on theoretically modelled energy consumption (or metrics relating to that), the Operational Energy Rating uses measured energy consumption (within specified boundaries) to construct an empirical energy rating that depends on the use of the building, not just the asset itself. Unlike the SRI, there are multiple examples of this kind of approach elsewhere, whether through Display Energy Certificates in the UK or the NABERS scheme in Australia. However, to have this directly part of an EPC is a new approach (and still being tested) and does raise the question of whether both the current assessment and the assessor are likely to be serve this type of metric appropriately; Performance Gap research tells us that real and modelled energy consumption metrics are likely to be vastly different, and current EPC assessors do not generally have the need, or training, to analyse real energy consumption data and link that with real user behaviour.

There are currently a number of European research projects (crossCert project 2023b) responding to these innovations, testing how these indicators may work with current EPCs and/or future versions of those EPCs. Before incorporating such innovations, it is necessary to consider the impact and reliance on the calculation engine of the EPC, the assessors themselves (and training level), the overarching framework and accompanying legislation, and the end-user of those indicators. The crossCert project (crossCert project 2023a), by comparing the details of different EPC assessments across Europe, is investigating the ability of different countries to accommodate these innovations. This problem is often overlooked; there is no single European EPC (despite the universal starting point of the EPBD and recommendations often provided to all signatories of that directive). A lack of harmonization means that assessment frameworks should be tested individually for any new innovation (and this is underway in some European countries (European Commission 2023)), but this does not just apply to new output metrics. The ability of different, current, EPC frameworks (both the assessment and assessor) to accommodate new calculation techniques is quite
different. Whilst posing a challenge (particularly for goals of harmonization of assessment across Europe), these different countries do provide a wide range of case-studies for what an EPC should be, and the calculations and assessors required to meet a stated end goal of assessment.

The next section takes two particular areas of innovation in terms of identifying alternative calculation engines. Crucially, and of relevance to this study, different forms of calculation can help generate new metrics of assessment (such as the above), or be more robust/reliable than existing methods in creating those indicators (with suggestions for critiquing this reliability discussed in Section 4).

3. Formulating replicable methods

The following techniques have been chosen as they have been tested in prototype form within a number of research projects, but they also respond to some of the concerns of existing modelling techniques within the energy compliance framework.

3.1. Dynamic simulation at scale

Although steady-state modelling is more common for EPC generation, dynamic building simulation has been used in the UK for some non-domestic building EPCs, as well as non-EPC system design.

For the residential sector, it is uncommon for simulation to be used for single dwellings at any stage. Urban-scale simulation of multiple dwellings is, however, of growing interest in academia and with different end-use audiences, such as local authorities. This provides the ability to aggregate intra-day demand patterns for sections of building stock, and opens up new applications that current steady-state models cannot speak to. These applications can involve energy network constraint issues, integration of renewable energy at different scales, use of storage, and more advanced analyses of the growing complexity of the relationship between supply and demand.

Dynamic simulation as an engine for urban energy modelling has been developed and demonstrated by several research teams. Previous work (Sola et al. 2020) has highlighted the use of various dynamic simulation tools for such applications at a timestep of less than 1 h. One of these tools, TEASER (Remmen et al. 2018), falls within the broader Integrated District Energy Assessment by Simulation (OpeniDEAS) framework, incorporating a Time Use Survey-derived stochastic occupancy model – StROBe (Baetens and Saelens 2016). There is also a clear growth in the use of GIS data in such research (De Jaeger et al. 2018).

An approach by the present authors (McCallum, Jenkins, and Vatougiou 2020a) demonstrates that the provision of more detailed simulated urban-scale demand can be achieved using existing data sources. Moreover, one of the key data sources is the UK EPC Register, as generated in its present form. Three distinct data types are used in this approach:

1. Large scale building stock survey records, i.e. EPCs, augmented by supporting databases such as Energy Savings Trust’s Home Analytics
2. GIS data, specifically the Digimap UK Ordnance Survey service, provided by EDINA (EDINA 2023)
3. Smart meter data

The approach uses two models which combine to deliver large quantities of EnergyPlus simulations for batch processing:

1. ParaDwell (McCallum, Jenkins, and Vatougiou 2020a); used to achieve the automated characterization of the physical and thermal properties of the dwelling or stock
2. ECHOsched (McCallum, Jenkins, and Vatougiou 2020b); a model designed to characterize the heating behaviours of the building occupants of each dwelling, or throughout the stock

These models feature entirely distinct processes, in part, due to the taxonomy of the data sources used, where ParaDwell uses EPC and GIS data, and ECHOsched uses smart meter data. In addition, it is useful to treat the outputs of ParaDwell and ECHOsched in distinct ways when carrying out both baseline simulations and scenario-based simulation sets, the latter offering the potential to explore prospective adaptations to each dwelling. More specifically, ParaDwell observes and interprets the data records relating to each dwelling, and carries forward a single parametric description of that dwelling. ECHOsched, on the other hand, should be treated as a tool for generating stochastic occupancy and system schedules for the simulation engine (EnergyPlus), to create statistically relevant patterns of energy behaviour. ParaDwell is a deterministic model – repeated runs will yield the same description of each building (assuming no model configuration parameters are changed). In contrast, ECHOsched is intended as a stochastic model – the system schedules produced by the model will change in a statistically consistent manner upon repeated runs. By treating the dwelling heating and system schedules as stochastic, a much closer reflection can be generated of dwelling usage in the real world, providing that an appropriate
statistical description can be obtained from suitable smart meter data.

In the following subsections, outline descriptions of ParaDwell and ECHOsched are provided, as applied to a theoretical case study of Kirkwall, under the ReFLEX Orkney project (ReFLEX n.d.). Proposed application in the dynamic simulation-based EPC context is then described. The demonstrated modelling approach (i.e. the case study and other previous work in this area (Vatougiou, Jenkins, and McCallum 2021) for an Orkney case-study), has only previously been applied in the context of stock-level energy efficiency studies, and has explored aggregate energy demand estimation for distribution grid modelling. In the aforementioned Orkney case-study, for example, an applied study is demonstrated using ParaDwell and ECHOsched, which showed that the extensive insulation retrofit and conversion to air-source heat pumps across a community of 322 dwellings could result in a 58% reduction in energy consumption (80% of the dwellings currently use electric storage or resistive heating). The Discounted Payback Period (DPP) for these retrofits still proved to be a challenge, with large numbers of dwellings recording a DPP of over 30 years. Following the exclusion of these ineligible properties, the distribution of DPP is shown in Figure 1 for different retrofit options.

3.2. Paradwell

The characterization of the geometric and thermal properties of a target dwelling or stock can be interpreted computationally from GIS data and EPC data using the ParaDwell model. Via the specification requirements in Table 1, the model is built around the idea of generating simple, parametric dwelling descriptions that can be both machine generated and machine read, in order to create model input files for a physics-based simulation engine (specifically EnergyPlus in the present implementation). The output of the model is a series of long character strings which contain the parameters that can fully define a simplified dwelling. Rectangular, T-shaped, L-shaped and C-shaped building footprints can, for example, be described using five parameter: width, depth, T-projection, T-offset left and right (see Figure 2). When combined with further parameters (for dwelling height, storeys, adjacencies, age, construction type, and heating system), a description of the dwelling can be recorded using a long character strings which can be stored and used as required to generate EnergyPlus models, using functions contained within the ParaDwell model.

Further capabilities of ParaDwell include the extrapolation of dwelling descriptors based on neighbouring areas. Identifiers such as postcodes and dwelling adjacencies can allow for the interpretation of dwelling age based, for example, where missing or outlying data can be investigated; see Figure 3.

3.3. ECHOsched

The stochastic model for Electricity, Cooling, Heating and Occupancy scheduling (ECHOsched) is defined by the specification outlined in Table 2. The procedure involves the partitioning of heating-dominated smart meter data into 24 h periods, followed by k-means clustering of the data points to discern time records of comfort and setback heating system states, see Figure 4. The inference of heating set point state can then be achieved for consecutive days, over the course of one year, for example; a single week is illustrated in Figure 5. A series of rule-based assessments (see elsewhere (McCallum, Jenkins, and Vatougiou 2020b)) are made on the resulting heating control records to determine whether operation of the heating system is programmer led (by the thermostat), ad hoc (by user intervention), continuously on, or idle. The results of an example study are provided in Figures 6 and 7.
4. Application of dynamic simulations for EPCs

Previously described applications of ParaDwell and ECHOsched relate specifically to the batch creation and simulation of large quantities of dwellings, for the purpose of studying aggregated energy demand. The following constitutes an outline procedure for using the same tools within the new context of generating dynamic simulation-based EPCs:

1. Obtain geometrical and location data of the target dwelling via GIS data and automatically generate a simplified parametric description (e.g. ParaDwell)

2. Obtain thermal characteristics and service (e.g. heating) technology information from survey data (including existing EPC input surveys), to augment the parametric description with more complete dwelling characteristics (e.g. ParaDwell)

3. Generate multiple occupancy profiles from available smart meter data to infer transient activity schedules (e.g. ECHOsched)

4. Create an ensemble of dynamic simulations input files via an embedded suite of automated tools, then running these models in a stochastic manner using the calculation engine (e.g. EnergyPlus) to treat occupant-driven uncertainty
Figure 3. Use of extrapolation routines within ParaDwell, to interpret missing or outlying data. Raw EPC data is provided on the left figure; the figure on the right shows sanitized data for the entire stock.

Figure 4. Clustering of gas consumption values for heating control evaluation, one ‘dwelling-day’ (McCallum, Jenkins, and Vatougiou 2020b).

Figure 5. Week-long gas meter record from a single site, with periods of machine-inferred high set point (McCallum, Jenkins, and Vatougiou 2020b).
Various enhancements from the above points can be introduced in order to assist with the process efficiency, scalability, and consistency of modelling from the perspective of practitioner involvement. Whilst some of these enhancements relate directly to innovations introduced by the discussed modelling framework, other benefits will originate from the implicit modern implementation of tools such as ParaDwell/ECHOsched, as cloud-deployable services. This can allow cross-linked Application Program Interfaces (APIs) between models such as ParaDwell/ECHOsched and their GIS and EPC data sources, with the potential to integrate other databases such as Land Registry. Enhancements on the above outline procedure include:

- Regarding point 1, new-build and existing dwellings will feature different considerations – existing dwellings will (in most cases) allow the use of existing GIS information to help generate the simulation model, whereas the practitioner will be required to generate geometric descriptions of new-build dwellings from scratch. In both cases, ParaDwell offers a simple and intuitive web-deployable interface to create and edit dwelling plans, and to generate the resulting EnergyPlus model files.

- Regarding point 2, prior existence of an EPC will dictate whether the practitioner can access historic information on the target dwelling (although this can confound surveying errors, much as it does at present when an EPC is being updated). A more perceivable benefit will be the opportunity for the practitioner to easily access accurate information (original construction date, for example, from Land Registry or similar records) via the same web-based interface as the model itself. There may also be benefits from accessing various building characteristics from the local area; historic EPCs from within a single housing development or block of flats should highlight outlying and inaccurate survey assumptions.

- Regarding point 3, significant opportunities and difficulties have been raised elsewhere in this paper relating to the measured energy consumption of a dwelling. The application of smart meter data for a target dwelling, for the purpose of dynamic simulation-based EPC creation, can be readily achieved using tools such as ECHOsched. Discussions on whether the fair assessment and potential for punitive action is appropriate have been provided under Criteria C4 in Section 4. Other concerns over the use of smart meter data include the visibility of sensitive and personal household patterns. It is worth noting that ECHOsched could, however, apply normalized behavioural schedules, which would anonymize and provide a level basis for comparing dwellings via EPCs. This could introduce a fair EPC system with enhanced representation of intra-day energy consumption, as compared to the present process for EPCs. For example, a flat with one bedroom could be assigned statistically relevant behaviour for a small household of one or two persons. A five bedroom house, on the other hand, could be...
Regarding point 4, successful lodgements of EPCs could be entirely cloud-based, allowing the potential for home owners/occupiers access to lightweight working models via smartphone or web-apps, potentially making use of their own smart meter data, in private, to assess how they can reduce their energy demand or carbon footprint. With respect to the centralized curation of the EPC register as a whole, greater consistency and integration of EPC recording and inspection via a cloud-based system can allow for better auditing processes. Clear and obvious errors on the practitioner part could, for example, result in the reassessment of failed lodgements, and result in practitioner licencing reviews.

4.1. Statistical treatment of empirical data

There is a growing evidence base (Zhao and Magoulès 2012; Ferrari et al. 2019) for using empirical data, rather than purely theoretical physical modelling, to assess energy use in buildings, though limits exist for the degree to which this can be standardized. This can be seen in models such as MARKAL (MARKet ALlocation) (Taylor et al. 2014), the UK TIMES Model (Fuso, Francesco, and Strachan 2017), and the CREST demand model (McKenna and Thomson 2016) where empirical energy data can be used (or energy inferred from other empirical data), though not necessarily focussed on the individual dwelling. Egwim (Egwim et al. 2022) investigated a range of high performing machine learning algorithms for developing predictive models for accessing building energy efficiency ratings. Most recent studies focused on estimating the building energy performance or rating using machine learning algorithms but there is very limited work focusing on the prediction of EPC using data-driven approaches.

Wood and Standring (Wood and Standring 2022) have demonstrated the potential of machine learning for predicting energy efficiency (and EPCs) of properties that do not have any official rating. The data used for the model development were collected from several sources, such as Council tax valuation data, land registry price paid, tenancy deposit protection scheme, energy company obligation scheme, central fit register (CFR) feed-in-tariff scheme, national energy efficiency data framework: consumption data, socio-demographics, and meteorological data. Quite often information at this level is not easily accessible and obviously challenging to process. Some recent developments in this area includes application of artificial neural network (ANN) models as an alternative approach for classifying a building’s EPC label (Tsoka et al. 2022).

The richness of empirical data now available through, for example, smart meter data, allows for an attempt to correlate energy demand with socio-economic class/status, dwelling properties, and appliances (Beckel et al. 2014; Gajowniczek, Ząbkowski, and Sodenkamp 2018). Crucially, high-resolution (e.g. 5 min and below) electricity demand data can be linked to behavioural characteristics that would not be discernible through the use of purely physical modelling. To ensure a robust model development, in addition to acquiring high-quality and diverse datasets, it is essential to identify suitable approaches and develop a systematic procedure that involves: (i) data pre-processing – handling missing data, outliers, and categorical variables; (ii) feature engineering – applying domain knowledge and suitable data-driven approaches for feature extraction, selection and often dimension reduction (principle component analysis, etc.); (iii) model development – involving application of potential data-driven models often combining multiple models in a hybrid structure for improved accuracy; (iv) model validation – two-fold validation involving training and testing of models as part of routine development and transportability of models across different datasets. Previous work has proposed application of a Hidden Markov model (HMM), with a Seasonal-Trend Decomposition procedure based on Loess process’ (STL), and a Generalised Pareto (GP) distribution for simulating such dynamics for high-resolution electricity demand profiles (Patidar et al. 2019). The application of STL facilitates temporal decomposition of stochastic components of demand profiles from deterministic features, noting how this differs from seasonal features. This level of information can characterize specific activities occurring at specific times in a household, which itself provides clues to causation (Torriti 2020). Building on this, templates of activity can be developed from these causal factors which could lend themselves to a more standardized approach for energy assessment, one which attempts to account for household (i.e. occupants) as well as the building itself.

This is similar to ideas of a Domestic Operational Rating (DOR) scheme (Lomas et al. 2019), which uses daily smart meter data alongside contextual information collected from an energy/household-based survey. As we attempt to have better integration and communication between energy demand and supply in future low-carbon energy systems, there is considerable value to being able to categorize energy use in buildings in this way.

Weather correlation can provide additional problems for standardizing energy use, particularly across larger geographies. For example, the authors recently demonstrated a successful integration of a ‘climate module’ within the framework of the STL-HMM-GP model (Patidar, Jenkins, and Peacock 2021). The underpinning partial
Summary statistics of original fintry dataset.

<table>
<thead>
<tr>
<th>Daily load CoV</th>
<th>Average annual load factor</th>
<th>Total Consumption (KWh/yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.48</td>
<td>0.29</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.17</td>
<td>0.09</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>10th Percentile</td>
<td>0.26</td>
<td>0.20</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>0.37</td>
<td>0.24</td>
</tr>
<tr>
<td>Median</td>
<td>0.48</td>
<td>0.27</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>0.60</td>
<td>0.34</td>
</tr>
<tr>
<td>90th percentile</td>
<td>0.67</td>
<td>0.39</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.90</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Variance from three components of PCA processes.

<table>
<thead>
<tr>
<th>Proportion of Variance</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost function related</td>
<td>0.9799</td>
<td>0.01996</td>
<td>0.00017</td>
</tr>
<tr>
<td>Load function related</td>
<td>0.5364</td>
<td>0.2752</td>
<td>0.1883</td>
</tr>
</tbody>
</table>

The month of February 2017 is taken from the above dataset to form clusters around the three PCA components extracted from the seven annual features. Larger periods of time are currently being explored by the authors, but the month above provided a complete and reliable dataset. To test the suitability of the three PCA parameters (PC1, PC2 and PC3), the total variance was noted for cost and load functions, as shown in Table 4.

Nearly 98% of the cost function related information is captured in PC1, and more than 80% of load function-related variance is captured in PC2. This suggests that these three principal components are adequate for characterizing the energy demand variability.
related information is captured in PC1 and PC2 combined. Therefore, the k-means clustering is deemed to be suitable for basing around four variables: PC1 for cost function, PC1 and PC2 for load function, and a degree day correlation variable. An elbow method is applied to obtain an optimal cluster number for the 56 dwellings, with the k-means procedure and its implementation in R (using ‘fviz_cluster’ in the R package ‘factoextra’ (R-package 2020)) described elsewhere (Kassambara 2017). Using this method, the intention is to produce household demand categories clustered into clear, distinct groups, that could inform energy classification schemes. The results of the clustering analysis are summarized in Figure 8, using Within Cluster Sum of Square (WCSS), the squared average distance of all points within a defined cluster, and Between Cluster Sum of Square (BCSS), the variation between the clusters in terms of squared average distance between all centroids. These are described in Equations 2 and 3 respectively.

\[
WCSS = \sum_{i=1}^{n_k} z_{ik} (x_i - \bar{x}_k)^2 \quad (2)
\]

\[
BCSS = \sum_{k=1}^{K} \frac{n_k}{n} (\bar{x}_k - \bar{x})^2 \quad (3)
\]

A smaller WCSS ensures less dispersion within a cluster (indicating that the cluster may be suitable for classifying the data) whereas a larger BCSS indicates a good separation achieved across the different clusters (indicating strong definitions between the clusters). A Silhouette (\(-1 < Si < 1\)) analysis (Table 5) is also carried out to measure how well each observation is clustered, where a Si value of 1 indicates observations are very well clustered, 0 indicates the observation lies in two clusters, and a negative value indicates the data is likely assigned to an unsuitable cluster.

Figure 8 visualizes these clusters for the chosen proxy PCA variables (Dim1 and Dim2). Different sensitivities, and numbers, of clusters can be proposed, but this study demonstrates the result of using four clusters for the 56 dwellings. Figure 9 indicates the Silhouette width of individual properties within each cluster.

Thus, a simple k-means clustering, if applied systematically (in this case involving a two-step PCA analysis with feature selection) has the potential to provide valuable insights in energy efficiency of buildings and could organize buildings based on their energy performances for assisting EPC estimation process. Although k-means clustering is just one approach investigated here, there are several potential data-driven approaches that could be tested for EPC estimation. This clustering cannot solely be used as a predictive model for EPC estimation but could help development of more complex models as part of a pre-processing strategy for organizing buildings with similar features. There is a need for continued research in this area to identify the potential for developing new approaches for addressing the existing challenges related to availability of good quality data, regional variations (such as regulations) effecting model transferability, engineering more comprehensive feature sets, and novel predictive model developments, e.g. combining physics-based and data-led approaches.

### 5. Critiquing new methods – a framework

Having introduced techniques for characterizing building energy demand at scale, and noting many more, there is a need to critique those methods against some criteria of suitability. An associated conference paper by the authors (D. Jenkins et al. 2021), defined ‘suitability’ based on likely future (and relevant current) requirements of energy performance assessments; this was carried out within the wider purpose of EPC functions, though not constrained to existing EPC techniques. Table 6 summarizes some of these criteria, with brief discussions provided below. They are designed such that proposals emanating from some of the ‘next-generation’ EPC projects, already discussed, could be critiqued in a standardized way. The criteria also help distinguish different applications that are suitable for different modelling approaches, moving away from a binary definitions of ‘wrong/right’ when critiquing models and, instead, showing understanding of the importance of specific application when judging suitability of approach. A lengthier discussion of these criteria is provided elsewhere (D. Jenkins et al. 2021).

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size of cluster</td>
<td>21</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>WCSS</td>
<td>37.22</td>
<td>29.15</td>
<td>39.23</td>
</tr>
<tr>
<td>BCSS</td>
<td>188.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Si width</td>
<td>0.20</td>
<td>0.34</td>
<td>0.19</td>
</tr>
</tbody>
</table>

### Table 5. Summary statistics of k-means clustering.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Alignment with reality</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>C2 Flexible demand rating</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>C3 Accommodates new technology</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>C4 Suitability for punitive action</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>C5 Extrapolating and standardizing</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>C6 Quality of input information</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
</tbody>
</table>

### Table 6. Evaluated ability of different assessment approaches meeting selected criteria.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Alignment with reality</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>C2 Flexible demand rating</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>C3 Accommodates new technology</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>C4 Suitability for punitive action</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>C5 Extrapolating and standardizing</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>C6 Quality of input information</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
</tbody>
</table>
Figure 8. 56 dwellings organized in four clusters through the application of k-means clustering approach coupled with PCA.

Figure 9. Silhouette coefficient plot with all positive values indicating coherent clustering.
6. Alignment with reality (C1)

All theoretical models can report performance gaps when compared to real energy data, but the advantage of dynamic modelling is the ability to directly characterize transient physical processes within a building. The simplifications adopted by steady-state models do not allow for this. This justifies a judgement that dynamic simulation satisfies this criteria to a medium level, compared to the low level of steady-state models. Empirical characterization, by definition, has a stronger relationship with real energy data and, with aforementioned Operation Energy Ratings becoming more part of the discussion of where EPCs go next, this category should be viewed somewhat differently than with traditional forms of standardized energy assessment.

7. Flexible demand rating (C2)

Whilst not seen as part of basic energy compliance, there is growth of interest in how existing assessment approaches consider demand flexibility – as seen with the aforementioned SRI metric in current proposals for EPCs. Considering existing calculation engines, it is not the purpose of steady-state modelling to account for dynamic aspects of energy use and, therefore, it cannot be used as a basis for understanding demand flexibility in a direct way – hence the low rating proposed here for meeting this criteria. Dynamic modelling, with higher temporal resolution, can model key drivers of flexibility (e.g. thermal mass, thermal storage, and occupancy control/behaviour), though this will depend on the temporal resolution of both inputs and outputs being used for such modelling. Likewise, the ability of empirical characterization to adequately reflect demand flexibility issues will depend on the resolution of collected data (e.g. from smart meters), but there is great potential to do so from high quality datasets. As already noted, most aspects of the SRI (in particular) do not require that level of analysis; rather, it infers demand flexibility potential from the presence of certain technologies in the building.

8. Accommodates new technology (C3)

In the UK, procedures exist for quantifying the impact of new technologies (Products Characteristics Database 2020), such that they can be accommodated within the existing SAP model. It is more difficult to judge whether a given modelling approach is actually suitable for accommodating that technology – or an aspect of that technology that may require a certain type of modelling (e.g. the ability to reduce peak demand would not be characterized in a standard steady-state model). Measures such as district heating, heat pumps, onsite storage and home-charged electric vehicles (which, in the UK are least, are seen as technologies likely need to considerable growth to meet carbon targets) may not be well-served by steady-state models calculating/averaging over long time durations. Dynamic modelling can operate with calculation time-frames more suitable to any technology with a strong diurnal cycle of variation. This justifies a high C3 rating for dynamic modelling, as opposed to the medium rating of steady-state modelling. The suitability of using empirical data in this way could be dependent on the level of sub-metering for large numbers of buildings; however, even without this, pattern recognition techniques (D. P. Jenkins, Patidar, and Simpson 2014) could still be effective for isolating signatures from individual technologies.

9. Suitability for punitive action (C4)

With the growing need to act to meet ever-closer carbon targets, EPCs (in particular) are not just being used in an advisory capacity; ratings are being used to stimulate mandatory action, judging whether a building can be sold or leased. This new use should require a re-evaluation of the model itself, making a judgement about the fairness of doing this and what the consequences may be. For many (usually steady-state) models used across residential buildings in Europe (related to the EPBD), it is important to note that this is designed for generating approximate energy ratings, does not account for the household specifically, and does not accurately predict energy bills. Enforcing action on the back of this could therefore be questioned (hence a low rating for C4). Dynamic modelling provides the option to explore why and when certain energy uses are higher than legislation may allow, providing some level of accountability and explanation for recommending a punitive action, even if we are still placing great reliance on a purely theoretical model. Justifying action on actual energy use (i.e. incontrovertible evidence that energy is high in a property) suggests a higher level of proof and accountability. Even a modified form of this, using generalized energy patterns from an empirical database to match a given property, would provide a stronger platform than theoretical models.

10. Extrapolating and standardizing (C5)

EPC-based models are designed for largescale, replicable use, and steady-state methods therefore can be ranked highly for C5. Dynamic models (rated medium) have more complex inputs and a higher level of training required for energy assessors – links between training requirements and upscaling the use of a method is being explored
in future work by the crossCert project. It is uncommon to see dynamic simulation used for residential buildings across Europe and its use for non-residential buildings across different countries is also inconsistent, with the associated challenge placed on assessor workforce. Some of the research presented here aims to question some of these assumptions, but it is unarguable that an assessor workforce trained on steady-state models will not be easily transferred to a framework relying on a completely different type of physical model. Empirical data could place an even greater challenge on standardization; even with the suggested clustering approach of this paper, empirical data is tied to the decisions of the householder(s) within the signals generated, unlike asset-based theoretical models. There may therefore be a compromise required on either the level of standardization, or a new approach to rating buildings that is linked to different household (as well as building) categories. Again, the Operational Energy Rating approach may suggest a path to follow in this regard.

11. Quality of input information (C6)

Issues of consistency across different energy assessors have already been documented in this paper. Even with improved quality control, standardized energy assessment procedures, by design, have to make compromises on input accuracy (e.g. the use of generic ‘look-up’ tables rather than building-specific, measured inputs). Dynamic simulation gives more scope for building-specific information to be collated (with more granular data required), and the data collection method in this paper suggests a way of achieving this at scale. An empirical approach to building and energy classification brings with it a higher degree of representation of individual buildings and their energy characteristics. However, this would have to be integrated into some form of categorization technique, with causal factors, that would require the clustering method of this paper to be developed.

It could be argued, therefore, that C6 counterbalances C5; that is, achieving reliable input information for individual dwellings can create standardization challenges.

12. Conclusions

Through reviewing recent research, and also proposing new methods developed by the authors, this study has compared established methods of standardized energy assessment with potential new forms. In doing so, a series of criteria has been proposed that judge whether key requirements of energy assessment can be met by different methods. Crucially, this framework attempts to capture likely future requirements of assessment; noting that what we once asked of energy assessment may be significantly different to the needs of today.

The study concludes that the range of techniques currently described through academic and industry research for assessing energy use in buildings are not necessarily well-replicated in standardized, regulated forms of assessment, particularly in the UK. This is particularly true for techniques involving urban-scale building modelling (adopting time-efficient uses of dynamic simulation) and statistical models. More research is required to judge, and demonstrate, how these different approaches can be used to address current flaws in steady-state energy modelling, whilst reflecting on new challenges emanating from our evolving building stock and surrounding energy systems.

The authors also suggest, on this basis, that the modelling community should review the form of modelling adopted in standardized energy assessment (such as EPCs) and re-visit whether new opportunities exist for quality and efficiency improvements, without compromising the central purpose of such energy assessments to provide useable and useful outputs to aid decision-making across a range of different end-users. The examples provided in the paper are proposed as potentially useful techniques to support this process, describing forms of modelling that although once thought of as entirely separate to compliance tools, could strengthen the toolkit of assessors and modellers to describe buildings more efficiently for a greater range of end-uses.

Acknowledgements

The work described in the paper is the result of a number of projects. The support of the funders, and related organizations, are greatly appreciated.

Data availability statement

The participants of this study did not give written consent for their data to be shared publicly, so due to the sensitive nature of the research supporting data is not available.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

CEDRI Grant reference [EP/R008655/1] was funded by the Engineering and Physical Sciences Research Council (ESPRC) and Indian Department of Science and Technology as part of the Newton-Bhabha programme. The CESI project [EP/P001173/1] was also funded by EPSRC, and ReFLEX [104780] by InnovateUK. Finally, crossCert [101033778] is funded by EU Horizon2020.
References


BPIE. 2014. Energy Performance Certificates Across the EU.”.


