

SYNTHESIS

Metrics for energy performance in operation: the fallacy of single indicators

Bill Bordass¹

Abstract

Many countries and organisations have now endorsed the climate emergency. New and existing buildings must play a big part in tackling this, though past history has been disappointing, e.g. with major gaps between predicted and actual energy performance. What metrics should be used to understand a building's energy and carbon performance in operation? Here there is uncertainty. For example, the United States is introducing carbon metrics, the UK has used them for many years, while the European Union recently made primary energy the common standard. Even though the reduction of greenhouse gas emissions may be the prime objective, UK experience suggests that undue concentration on any single headline metric can lead to severe unintended outcomes. The paper outlines the history and some results of various energy and carbon metrics used in UK policies and publications for non-domestic buildings since the 1973 oil crisis, with a few examples from other countries. It suggests how multiple indicators may help resolve future problems, what metrics might be used and how to make the underlying detail more accessible, e.g. with component and system benchmarks.

Policy relevance

Recent UK policy on the climate impact of buildings has been largely framed in terms of CO₂. This seemingly sensible paradigm has had unintended consequences. (1) Contributions from low-energy 'passive' design; efficient equipment; good construction, commissioning and handover; effective energy management; and renewable and low-carbon energy supplies are conflated. There is no target for energy consumption itself. (2) This has divorced building professionals from the realities of in-use performance and deprived many of the necessary agency to improve it. (3) The limited amount of feedback means that policies can favour measures that look good in theory, but which do not work well in practice. This can make buildings too complicated, with high operational and management costs. To stimulate sustainable investment in truly low-carbon buildings, a suite of metrics and benchmarks needs to focus on performance in practice and motivate all the players involved. Elements of a viable approach are presented. Once in-use energy performance becomes reliably visible, action can become more effective.

Keywords: benchmarks; buildings; carbon emissions; carbon metrics; energy; landlords; metrics; net zero; non-domestic

1 Introduction

If you don't measure the right thing, you don't do the right thing.

(Joseph Stiglitz, cited in Goodman 2009)

The world is in a climate emergency (UNEP 2020). Greenhouse gas (GHG) emissions need to be stemmed substantially by 2030 and eliminated by 2050, preferably much earlier: ideally they would be reversed. Buildings will play an important part in this process, not just new ones but the existing stock, particularly in developed countries.

While the need is urgent, unintended consequences must be guarded against, otherwise efforts may disappoint, while ineffective solutions with high levels of embodied carbon may exacerbate problems. In the past, market forces and lobbying power have tended to promote add-on technology. However, some policy-makers now recognise the need

¹ Usable Buildings, 10 Princess Road, London NW1 8JJ, UK. ORCID: 0000-0001-6559-7870
Corresponding author: Bill Bordass (bill@bordass.com)

to reduce embodied energy and carbon, and not just operational. The various actors will need rewarding for using more time, skill, care and thought to get more from less.

An effective low-carbon transition requires clarity about where we are, what we may need to do and how we are doing. The European Union's (EU) Green Deal (European Commission 2019) stresses the need for reliable, comparable and verifiable information. But where is the evidence that the rapid deep renovation it advocates, using largely technological and economic measures, is a good path to follow? It barely mentions people and management. Deep renovation suits certain contexts, but is rapid scaling up justified, given its high embodied carbon and prevalent performance gaps (Gram-Hanssen & Georg 2018)? Could something simpler but more broadly applied be more cost- and carbon efficient? Many non-domestic buildings can save in the region of 20% here and now at little cost, largely by solving minor problems and improving control, operation and management. This applies to many new buildings too, where controls and user interfaces are often deficient; and systems seldom properly tuned up after handover (Waide Strategic Efficiency 2014).

The countless origins of energy performance gaps include building regulations requirements that are met in theory but often bear little relationship to in-use outcomes; and limited understanding of how buildings really perform in operation. What information is available is often poorly used, with case studies wrongly dismissed as anecdotal (Flyvbjerg 2006) and warnings overlooked. Policy-makers often say: 'we need more evidence' or 'we want a statistically significant sample'. This may not help much: bulk data on buildings are often of poor quality, and statistics seldom capture enough context to appreciate the reasons for the outcomes. 'If you are not careful, more data just gives you better medians.'¹

This paper identifies a need for more clarity in describing and benchmarking predicted and in-use energy performance, both to motivate action and to help understand what works and what needs improving. Transparent reporting must run through all stages of briefing/programming, design and construction, in general and in detail. The insights could then influence practices of clients, designers, builders, managers, occupiers, investors and government, and of their service providers. More and better information needs to be freely available, not locked up in commercial sources, including outsourced government databases (*e.g.* Cohen & Bordass 2015). While simple headline metrics have their attractions, if the underlying detail is not rich enough to address the individual requirements of the wide range of players that need to come together, the large reductions in energy consumption and CO₂ emissions now sought are unlikely to materialise.

This paper uses the following terms:

- **Metric:** a quantifiable system of measurement. This includes:
 - **Performance indicator:** typically calculated by dividing a **numerator** specifying an amount (*e.g.* kWh/year) by a **denominator**, a measure of intensity (*e.g.* occupancy) or extent (*e.g.* m² floor area), to produce kWh/m².
 - **Benchmark:** against which a performance indicator may be compared. These may be derived from statistics, case studies and empirical or theoretical calculation.
 - **Rating:** comparison of a performance indicator with the appropriate benchmark, often by numerical ratio or in rank order percentile.
 - **Grade:** placing the rating on a simple scale, for example A–G, pass/fail for building regulations; poor–fair–typical–good–excellent; or by quartile or decile.

Local variations make it impossible to apply these stringently. For example, the EU calls its A–G grades Classes, while Australia has Star Ratings.

The paper is structured as follows. The next section discusses elements of energy and carbon metrics. This is followed by energy and carbon performance indicators (CPIs) and weightings. The paper then considers approaches to benchmarking and some developments in the UK. The final section provides conclusions and implications for policy and practice.

Not covered here are the important topics of embodied energy and carbon in construction, maintenance, repair and upgrading; load profiles and peaks; or how time of use affects the primary energy and carbon content of electricity.

2 Elements of metrics for operational energy and carbon performance

This section outlines the ingredients of performance indicators: the numerator, conversion factors, the denominator and various types of boundary. It also introduces the concept of 'Base Buildings' and landlord's energy statements. The indicators themselves are reviewed in section 3.

2.1 The numerator

Building energy use is often expressed as kWh of annual delivered energy (US = site energy, EU = final energy); primary energy (US = source energy, with a slightly different definition); energy cost; or in terms of its climate impact, often in units of CO₂ equivalent (CO₂e). These may also be split by energy source, including on- and offsite renewables. Over the years, UK policy has used many of these, with differing consequences (**Table 1**).

Even energy units may be uncertain. For combustion fuels, the default UK convention is the gross (higher) calorific value, but net (lower) is widely used in other countries. The difference is large for hydrogen-rich fuels, owing to latent heat in the combustion products.

Table 1: UK energy and carbon metrics and some consequences.

Date	Example	Metric	Some consequences
1974–90s	'Yellow Booklet' Normalised Performance Indicators (NPIs)	Stress on reducing total delivered energy (EU = final energy, US = site energy)	Normalised for weather and occupancy-hours. Some later revisions included fuel/electricity splits. Caused some buildings to change from fuel to much more expensive and high-carbon electricity. Normalised performance was sometimes confused with raw
1980s	<i>e.g.</i> Monergy campaign, 1986	Stress on cost and efficiency : not conservation, which politicians still do not favour	Considerable effort on energy management. Motivation slackened when energy prices fell. Once price competition was introduced, contract negotiations could save a lot of money easily, so why bother with investment and management?
1991–2001	Energy consumption guides	Separate benchmarks for delivered fuel and electricity	Some guides included cost and CO ₂ indicators, and breakdowns by end-use. For a while this improved the focus of designers on all elements of energy performance in use
2002–08	In-use benchmarking neglected by government and its agencies	Dominated by building regulations requirements in CO₂ metrics	Policy emphasis on CO ₂ shifted activity towards low-carbon and renewable energy, too often at the expense of basic energy savings. A 'design for compliance' culture burgeoned, with design thinking narrowed to modelled calculations of 'regulated loads': the heating, hot water, cooling, ventilation and lighting prescribed in the EU Energy Performance of Buildings Directive
2008 on	In-use benchmarking continues to be neglected	Regulations and building energy certificates largely based on CO₂	Compliance culture continues. Poor support to Display Energy Certificates based on metered energy use, which also failed to be extended to the private sector (Cohen & Bordass 2015)
2018 on	Revisions to building regulations pertaining to the conservation of fuel and power (Part L) MHCLG (2019)	Primary energy (Source Energy in the US is similar) and CO₂e	Could create new unintended consequences, in particular too fast a shift from fuel and heat to electricity, even though electricity is much more expensive and its capacity to do useful work should not be squandered

2.2 The denominator

To create an energy-use indicator (EUI) or carbon performance indicator (CPI), the numerator (kWh, kgCO₂, cost *etc.*) per time interval (usually a year) must be divided by something. **Table 2** shows widely used denominators, and a few of their strengths and weaknesses. Internal floor area is a widely used starting point (UBT 2011): it is usually recorded (*e.g.* in leases) and is easier to audit than occupancy, though even it is subject to uncertainties. Other denominators are often better used in secondary performance indicators.

Area definitions vary between sectors and jurisdictions, complicating comparisons. Although standard units are desirable, common usage also needs taking into account. For example, UK designers usually refer to gross internal area in metric units (m²), while the commercial property market uses net lettable area in imperial units (ft²).

2.3 Premises and energy-related boundaries

For construction work, it is usually clear what a 'building' is: the project. In use, it becomes more complicated: premises may consist of a group of buildings or parts, or a floor in a rented building. There may also be outdoor services, *e.g.* floodlights. Ideally the boundary of the premises, its management and its metering would be identical, but blurring commonly occurs, for example, where:

- individual buildings on a site and on district heating/cooling systems are not separately metered
- premises in multi-tenanted buildings have unmetered shared services, commonly heating, ventilation and air-conditioning (HVAC) and hot water
- stored fuels such as oil, coal and biomass are not measured in day-to-day use.

In the first two cases, several premises can sometimes be aggregated to a more distinct boundary, *e.g.* a campus.

Where there are onsite active renewable systems (*e.g.* turbines, photovoltaics (PV), site-grown biomass), energy used in the premises no longer equals what it imports, while combined heat and power (CHP) (co-generation) systems shuffle the pack. Carbon policy-makers may think all they need is the demand the premises puts on the national infrastructure, but building-related insights will be hampered if the additional detail is not available.

Table 2: Denominators often used in energy-use indicators (EUIs) and carbon performance indicators (CPIs).

Denominator	Strengths	Weaknesses	Comments
Floor area (m ² or ft ²)	Measure of useful space. Often recorded, but not always accurately	May reward lightly used buildings, unless intensity of use is also taken into account in some way	Floor area conventions (<i>e.g.</i> gross, net, usable, internal, external, treated (heated) <i>etc.</i>) and definitions can vary widely between sector and country
Volume (m ³ or ft ³)	Used in some sectors, <i>e.g.</i> historically in UK health buildings, but its justification is not at all clear	Not routinely recorded. Tall ceilings help natural ventilation and light. With air-conditioning, lower ceilings aggravate differences in EUI	Not normally very helpful. May suit sectors (<i>e.g.</i> warehouses) where height can contain useful volume. Often better to have separate sector benchmarks by area
Number of workstations, occupants or occupant-hours	Indication of ‘productivity’ of the building in some sectors	Occupant numbers (and occupant-hours) are difficult to count reliably. Overestimates have often been used to ‘improve’ EUIs	Best as a secondary indicator, or where occupancy metrics are robust (<i>e.g.</i> school rolls). Should become more useful as systems for monitoring occupancy improve
Volume of production or sales	Useful where production or sales are well defined	Relationship is often quite weak	Used for some industrial processes, restaurant meals and supermarket sales
Other commercially relevant factors	Relates to business drivers	Relationships need to be demonstrated to be relevant and useful	For example, type and quality of hotel and number of bedrooms

Standard EN 15203 (CEN 2005) on energy ratings therefore recommends reporting on-site active renewable energy separately. Premises energy use (PEU) can then be calculated by adding this to energy purchases. Ideally, CHP/co-generation would be treated similarly, with all its energy inputs and outputs metered. **Figure 1** shows the different boundaries for the PEU and the operational rating: the energy imported (and where relevant exported) across the premises boundary.

Legislation may, however, presume buildings, not premises. For instance, the Energy Performance of Buildings Directive (EPBD) (European Parliament & Council 2002) requires certificate display in many public buildings >500 m². The UK’s Display Energy Certificates (DECs) are based on metered energy use, renewed annually, but an over-literal interpretation of the EPBD requires a school, for example, to certify each relevant building and exclude smaller ones (Cohen & Bordass 2015: postscript). As most schools only have one set of utility meters, the results include much

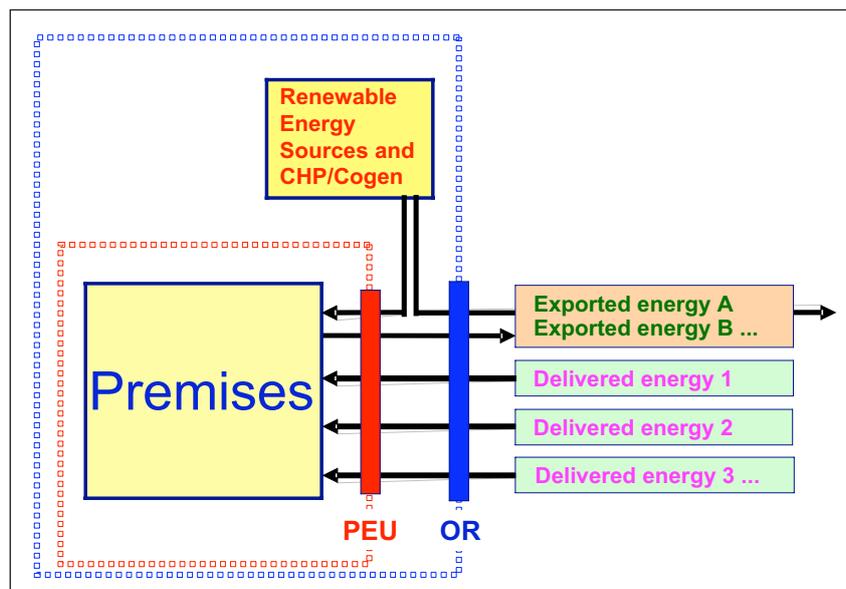


Figure 1: Boundary of premises energy use (PEU) and operational rating (OR).

Note: CHP/Cogen = combined heat and power/co-generation.

Source: Cohen, Bordass, & Field (2006).

estimation, so the individual DEC's are often unsound. A site DEC would be cheaper, better and improve comparisons. In contrast, users of the United States' EnergyStar (2020) Portfolio Manager system can set their own boundaries.

2.4 Landlord and tenant boundaries

Organisational boundaries are also important. Where responsibilities are blurred, energy is likely to be wasted. For example, the author has surveyed premises where facilities managers have never seen the fuel bills, because their employer purchases utilities centrally. In multi-tenanted buildings, many parties contribute to the final outcome: landlords have managing agents, consultants, maintenance contractors and facilities staff. So may each tenant. The principal-agent problems also afflict tenanted buildings (IEA 2006), as intermediaries can have different motives from the developer, landlord and the tenants. Landlords also have no incentive to go beyond legal minima if they are unable to recover their extra investment and management costs, ideally as higher rents and capital values. Tenant departments such as information and communication technology (ICT) and catering are often driven by service not economy, and may get their energy free, so it has little influence on purchasing and operational decisions.

In UK prime offices, tenants often install their own fitouts, including HVAC systems that use the landlord's core services but are locally controlled. The landlord may then lack a clear overview and turn into a 'dumb provider of 24-hour heating, ventilation and cooling', as a NABERS expert stated when visiting a well-regarded new London office building. In some other countries (*e.g.* the US), landlords are more likely to provide serviced space, or undertake fitouts on behalf of tenants.

2.5 Base buildings

To improve energy management in multi-tenanted buildings, landlord and tenants need agency over what they can control, and good information on how they each are doing. The Australian NABERS (2020) Base Building rating, launched by the Sustainable Energy Development Agency of New South Wales (NSW) in 1999, shows the way for landlord's services in rented offices. Since landlords' services in NSW usually had separate utility meters, data and benchmarks were available at the outset.

NABERS started voluntary, supported by some major property companies. By 2004, it had established a foothold. The federal government then drove the market by requiring any new office it rented to be 4 stars or better (at the time, the median was 2.5 stars and leading edge was 5 stars). Ratings have continued to improve ever since (Cohen, Bannister, & Bordass 2015, Cohen *et al.* 2017), with a short setback when they became mandatory. A new 6 star grade has had to be added, half way from 5 stars to zero carbon.

While NABERS ratings are also available for whole-office buildings and individual tenants, they are not as widely used, probably because the markets are more diffuse. Co-assessment may increase uptake, rating tenants at the same time as the landlord. NABERS is also extending into highly managed buildings with relatively few key players, including data centres and public hospitals. Base Building ratings are also available for shopping centres and apartment blocks.

2.6 Commitment agreements

The success of NABERS ratings created a problem: what about new buildings that have no operational performance record? The solution was the commitment agreement (CA), where a developer and its design, building and management team sign up to produce a Base Building with a declared operational rating. Early CA projects were gruelling, but most met, and some surpassed, their commitments, though always after tune-ups and with a few requiring expensive alterations. Today, tune-ups are still necessary (they always will be, but without CAs they seldom happen), but the process is smoother, because the property, building services and contracting industries have learnt what to do.

CAs require careful modelling of HVAC systems and controls, with a review of the design and the final outcome by independent assessors. Over the years, HVAC systems have become more efficient, better specified, commissioned, handed over and fine-tuned. Engagement with the outcomes has helped designers focus on what works: this can also be smaller and less complicated, which helps to cover the cost of the process and of more efficient plant. IPD (2013) found that high-rated offices rented faster, had fewer vacancies, and commanded higher rental and market values. A good in-use energy rating had become a proxy for overall quality.

2.7 Landlord's Energy Statements (LES) and ratings

In 2006, the UK government said DEC's would be mandated for public buildings from October 2008, and might be extended to commercial buildings. This would cause difficulties in multi-tenanted buildings, as landlord-only utility metering is not widespread and sub-metering is ragged, so robust Base Building ratings were not practicable.

The British Property Federation (BPF) (2007) instead developed the Landlord's Energy Statement and Tenant's Energy Review (LES-TER). **Figure 2** outlines the LES process. Its output (see Appendix A) tells each tenant how much of each type of energy the landlord has used on its behalf, what for, the associated CO₂e and how it all has been apportioned. By adding LES data to its own, each tenant can obtain its own DEC: the exact position of the landlord–tenant boundary no longer matters.

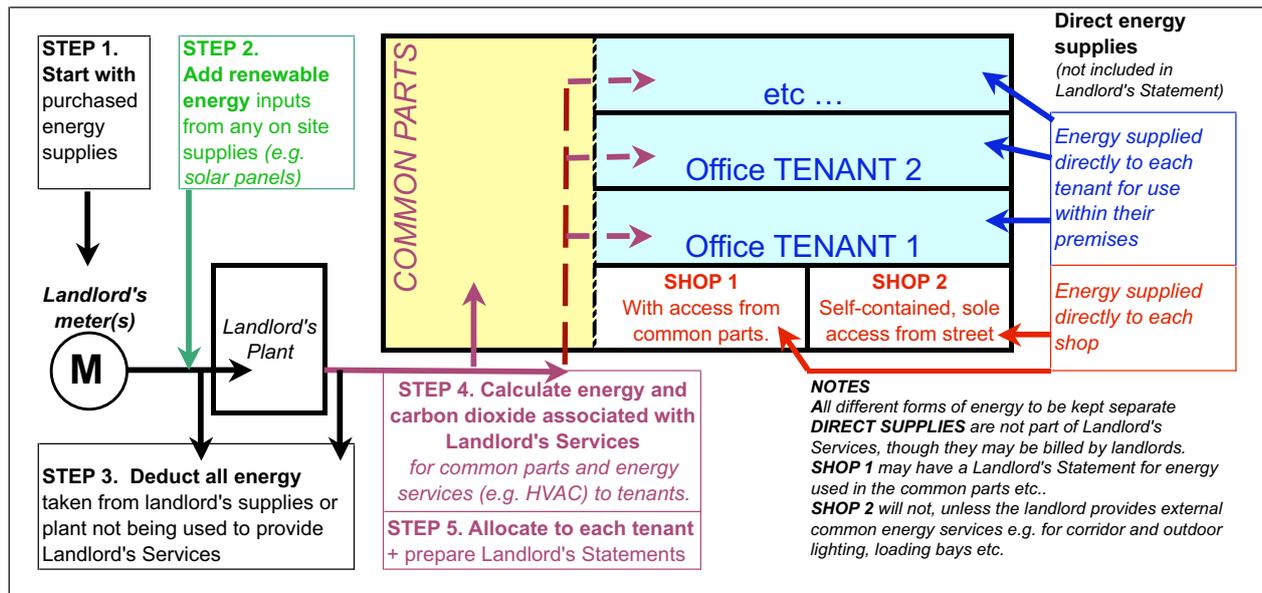


Figure 2: The Landlord's Energy Statement (LES) process for a rented office.
Note: HVAC, heating, ventilation and air-conditioning.

In spite of strong industry support, DEC's were not extended to commercial buildings (Cohen & Bordass 2015). This eliminated the regulatory driver for the LES. Disappointed, leading property industry members of the Better Buildings Partnership (BBP) sought a NABERS-style voluntary rating. Feasibility studies showed that ratings (like the LES) that required some estimation would not convince investors. Base Building metering would be necessary, but it proved too expensive to retrofit to many existing offices, where the configuration of HVAC and electrical systems was unsuitable.

For new buildings and major refurbishments, good metering could be designed in at little or no cost. For these, BBP (2020) developed Design for Performance (DfP), the UK equivalent of NABERS CAS. LER, the associated Landlord Energy Rating, uses standard weighted energy (SWE) (see section 3.4).

The LER follows NABERS, EnergyStar and other systems by grading in stars, from 1 star (poor) to 6 stars (market leading). While the A–G scale used for DEC's suits new manufactured products, the property market likes stars as they provide a more positive message: Would you prefer to have a 3-star building or a D-rated one? Ten leading property companies are now using DfP, with leading engineering practices declaring their support. It is hoped that the focus on in-use energy performance, uncomplicated by carbon factors, will help to overcome the UK's stultifying 'design for compliance' culture.

3 Operational energy performance indicators and weightings

3.1 Introduction

The elements in section 2, split between landlord and tenants as necessary, can be assembled into a variety of energy demand, energy consumption, cost, CO₂ emissions and other indicators. These may include normalisations and take account of on- and offsite renewables and CHP. A headline indicator may be required for market engagement, but any single perspective will obscure others that may also be informative. A more rounded view will help professionals and policy-makers to make better choices.

Performance indicators may be applied at many scales, from a whole site to a particular responsibility (e.g. landlord's services), system (e.g. heating), area (e.g. kitchen) or element (e.g. light fittings). They may also be aggregated to stock levels, e.g. a street, city, region, country, building type or management portfolio.

3.2 Weightings in the EU

For EU energy certificates, Standard EN 15203 (CEN 2005) expresses Operational Ratings (ORs) as the sum of the weighted annual consumption of each form of energy supplied (imports less exports) per m² of usable floor area. For the EPBD, EU member states could choose their own weightings, based on primary energy factors, energy cost, CO₂ emission factors or other policy drivers. However, a recent amendment (European Parliament & Council 2018) now requires common reporting in primary energy units. England (MHCLG 2019) has therefore added a secondary CO₂ indicator, while the UK Green Building Council (2019) advocates kWh/m² total delivered energy, although its appendix B Reporting Template includes its components by source including renewables. These different perspectives indicate a need for multiple indicators.

The UK's concentration on CO₂ led to difficulties. Will the EU's switch to primary energy be any better a motivator to improve building performance? Both primary energy and CO₂ indicators have important purposes, but they conflate

the performance of a building and its energy supplies, which clouds international comparisons and allows energy with a low-carbon or primary content to hide an inefficient building. More transparency is required.

3.3 Conversion factors for electricity

The primary energy consumption and CO₂ emissions in making electricity vary greatly: by source (fossil or renewable), region (*e.g.* Poland's electricity is largely generated from coal and Norway's from hydro), over the years, and from minute to minute. The marginal carbon and primary energy burden of adding load (or benefit of removing it) at some times can be huge, because the marginal power station may well be less efficient and use a higher carbon fuel (*e.g.* Wattime & Rocky Mountain Institute 2017).

In the UK, the carbon emission factor for mains electricity has been falling rapidly due to changes in energy sources: coal giving way to gas, wind growing quickly offshore, a significant nuclear legacy, and some hydro and PV. From 0.519 kg CO₂e/kWh in the current (2012) edition of the Standard Assessment Procedure (SAP), the factor in the 2018 draft was 0.233, similar to mains gas at 0.210. The latest draft (BRE 2019) says 0.136 kg CO₂e/kWh, using projections to 2025. Its primary energy factors are 1.501 for electricity purchased and 0.501 for renewable electricity exports. Publishing all these factors (which are also used in English building regulations) to three decimal places indicates a blindness of policy to the fundamental uncertainties.

The new UK factors might well drive unintended consequences, *e.g.* a 'dash to electricity' in the name of sustainability. But electricity currently accounts for just 17% of UK delivered energy use (BEIS 2019). If unsupported by other policy measures, a rapid increase may create bottlenecks in national and local distribution (Vivid Economics & Imperial College London 2019). If growth exceeds that of renewable supplies and the associated balancing capacity, CO₂ factors may even rise. Electricity is also valuable thermodynamically, being almost pure capability for doing work. It must not be squandered just because it has a nominally small carbon content. UK electricity also remains expensive, typically four times the price of gas per kWh, though gas should really carry a much higher carbon penalty.

3.4 Standard energy weightings?

As part of its work on EU energy certification, in 2004 the EPLabel (2006) project suggested that a set of simple, constant standard weighted energy (SWE) factors would permit the energy use of premises anywhere in the world to be compared, whatever the local primary energy and CO₂ factors. Property companies with international portfolios liked the idea, so it was included in the LES (BPF 2007). *Design for Performance* (BBP 2020) uses a similar approach, expressed as 'electricity equivalent'.

SWE accounts in a rudimentary way for the thermodynamic value of different energy sources, in particular that delivered heat comes with upstream losses, while electricity is almost pure work. The proposed simplified weights were:

- combustion fuel 1.0
- hot water 1.25, to account for some combustion and distribution losses
- chilled water 1.25 too, for simplicity
- electricity 2.5; some commentators in Australia and the US suggested 3.0 or even 3.5, but 2.5 was the consensus figure.

Exergy analysis (assessing the ability of an energy source to do useful work in a particular context) might produce more rigorously based multipliers, and help stop precious sources (such as renewable electricity) being turned into heat prematurely.

3.5 Discussion

Whatever the merits of any specific weighting, too much stress on any one set may well prove troublesome. At best, people will 'game the system' to obtain the best result with the least effort, *e.g.* choosing low-carbon fuels rather than making a building efficient. Multiple metrics can help to avoid this. If a rating system takes carbon offsets and dedicated offsite renewable supplies into account, their influence should always be reported separately (as in the LES and with Green Power in NABERS (2020)), and not rolled into the headline indicator.

A single headline value also fails to expose the potential for multiplier effects (ACE 2001) where, for example, if one were to:

- halve demand, *e.g.* by questioning standards (*e.g.* do we really need this space, hot water here or to light the entire room to 500 lux?) and using passive measures
- double efficiency, which is possible today for some energy end uses and
- halve the GHG emissions from the energy supplies

the footprint for that end use would fall to one-eighth: a dramatic change. Reporting and benchmarking by component can help here (see sections 4 and 5).

Supplementary information should accompany any headline indicator and not be hidden away. This could include, as in UK Green Building Council's (2019) reporting template:

- Total annual imports (and exports) of energy to the premises by type
- Total onsite active renewable generation and CHP, by energy type
- Units used in performance indicator denominators, particularly floor area

Unweighted data need to be available too, as, for example, is shown in the LES (see Appendix A), so underlying detail can be scrutinised and transactions can take place between parties, *e.g.* landlord and tenant. Different indicators can also be calculated by applying new weights to the raw data. Accuracy indicators should also be considered for kWh values, *e.g.* for estimated readings, stored fuels with long intervals between deliveries or biomass not accurately measured for amount, moisture, calorific value or carbon content.

3.6 Normalisation

Performance indicators may be normalised, *e.g.* for weather, climate, and sometimes exposure and occupancy. Normalisation allows indicators for buildings in different contexts to be put into better rank order. However, CIBSE (2012: Section 19.5) warns that normalisation should be used with care and only where relationships are proven, to avoid introducing unhelpful distortions. Normalisation can easily be abused, and normalised indicators confused with raw ones.

Weather adjustments help energy managers to review monthly consumption against targets, but climate correction for building location may be less useful, as outlined below. Seeking comparability between buildings in different climate zones, EUIs are often corrected to a single national heating degree-day standard. However, UK data (*e.g.* BEIS 2018) suggest a flatter relationship. It appears that even where the regulations are the same, thermal envelopes and heating systems receive more attention in colder places. For example, when commercial condensing boilers were new to the UK, sales were much stronger in the colder north than the richer south-east.²

The Europrosper (2002) study considered a standard EU climate correction, but its national reviews discovered that buildings in cold regions could use less heat than in milder ones—where good thermal envelopes and efficient systems were less critical to survival. Similarly, air-conditioning in hot climates (where systems are more single-purpose) can use less electricity (sometimes not just relatively) than in milder ones which afford more opportunities for waste, *e.g.* running unnecessarily, or with heating fighting cooling (*e.g.* Bannister and Zhang 2014).

In order to permit corrections whilst retaining the raw data, in 2002 the author developed a technique to adjust the benchmark itself, an approach adopted for UK DECs (see section 5). Alternatively, adjustments to the raw data may be presented graphically, as in Stage 2 of TM22 (CIBSE 2006).

4 Approaches to benchmarking

4.1 Introduction

A benchmark is a point of reference for measurement. Building operational energy and CPIs can be benchmarked against many references, including:

- a ratio to a notional building, as in UK regulations
- its previous performance, as in routine energy management exercises
- its location in a peer group, with results expressed as a percentile (*e.g.* Energy Star, US); a grade (A–G in the EU and stars in NABERS); or as typical (median) or good practice (varies), in UK consumption guides
- by calculation, using methods ranging from sophisticated simulation (and even 'digital twins'), via semi-empirical models to rules of thumb, all at various levels of detail.

A widely used indicator is annual weighted energy use and its components per unit floor area (see section 3.2). Other aspects can also be benchmarked, in particular energy demand profiles over a day, week, month and year for the entire premises, and for its systems. These are beyond the scope of this paper.

Chapter 20 of CIBSE (2012) classifies benchmarks as:

- **Overall:** usually at good practice and typical levels for fuel and electricity.
- **Component:** often expressed in W/m² of installed capacity (*e.g.* for heating, cooling, ventilation, fans, pumps, lights and office equipment), and/or as efficiency indicators, for example W/(litres per second) for mechanical ventilation and W/m² per 100 lux for lighting. Some are also expressed as product grades, *e.g.* A–G for electrical equipment in the EU.
- **End use:** some UK consumption guides showed their benchmarks split by:
 - fuel: typically into heating, hot water and catering
 - electricity: heating, hot water, cooling, fans pumps and controls, lighting, office equipment, communication and server rooms, and catering

- sometimes 'special' areas and end uses were also included, particularly computer rooms, swimming pools, process loads and external lighting.

ISO (2013) Standard 12655 includes a list of 12 main end uses.

4.2 System and component reporting and benchmarking

Benchmark values can be populated top down, starting with annual consumption by fuel; or bottom up, by component. The TM22 Energy Assessment and Reporting Method (CIBSE 2006) uses an iterative approach to reconcile top-down data (mostly from utility meters and sub-meters, if any) with bottom-up estimates (and/or measurements) of system and end-use values, augmented as necessary by spot measurements and short-term logging. A development of TM22 software (IUK 2012) can also import half-hourly electricity demand data and reconcile this with estimated demand profile characteristics for each end use.

TM54 (CIBSE 2013) uses a similar component-based approach to estimate energy use at the design stage, and can incorporate results from modelling. Both TM22 and TM54 allow users to start with small amounts of data and add more detail as it becomes available, or as time and budget allows. TM22's Excel software gives a provisional result at every step: a development version also includes an audit trail.³

The associated 'tree diagrams' (Figure 3) illustrate the multipliers (Field *et al.* 1997) on the basis of load × equivalent full-load hours. These can be used to present and compare results of in-use energy surveys and/or estimates for new buildings in a simple manner. The data can come from any source, from rules of thumb to sophisticated modelling and monitoring. Each box can also be used to show system, end-use or component benchmark values as well.

4.3 An example

Figure 3 summarises, in rounded numbers, the predicted and actual annual energy use by lighting an air-conditioned office which had low-energy aspirations:

- at the top: actual annual electricity was over three times the design estimate
- centre left: the connected load was one-third more than predicted
- centre right: the annual equivalent full-load running hours were 2.4 times the predicted level; the bottom level shows why
- bottom left: lighting was brighter than predicted (perhaps allowing for a maintenance factor) and somewhat less efficient
- bottom right: occupied hours were 33% longer (for cleaning and some weekend working), while the lights were on at their equivalent full-load output for 90% of the occupied period, not the 50% anticipated.

Findings from this particular exercise included:

- the original estimates for installed power were not updated once the design was finalised and specific fittings selected; this commonly happens: design expectations are seldom managed throughout the procurement process
- the estimated hours of use were optimistic as designers often make no allowance for cleaning or any eventing or weekend working.

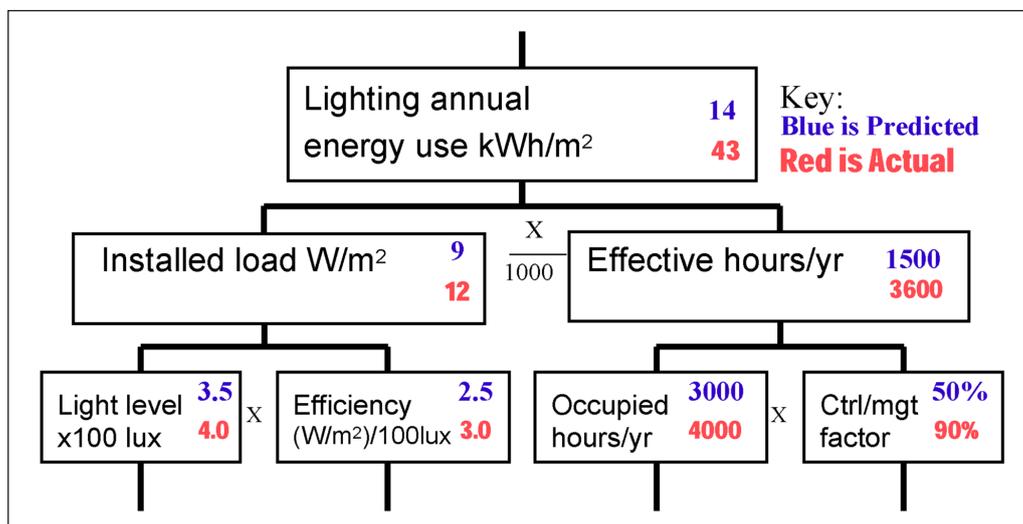


Figure 3: Chartered Institution of Building Services Engineers (CIBSE) TM22 'Tree diagram' for lighting in an office.

If design assumptions are made explicit and kept up to date, the reasons for any differences can be understood. These can also be used to develop better benchmarks and rules of thumb.

The biggest discrepancy was in the control and management factor:

- In open-plan offices, when unoccupied zones were switched off, nearby occupants could be distracted or found the interior gloomy. As a result, the lights were programmed full on during the core time. Some later buildings were able to overcome this problem by illuminating perimeter walls efficiently.
- In cellular offices, the presence of detectors switched lights on in low daylight, whether or not occupants and visitors wanted them: absence detectors and manual on/off switches would have been better.
- For a 'business-like' external appearance, the management kept the venetian blinds down (not closed). The occupants would have preferred more control.

Despite its apparent simplicity, tree diagram reporting can give surprising insights. For example, a new building claimed to be an advance on an exemplar that used very little gas. A journalist preparing an article about it requested a few summary values. Its calorifiers had a capacity of 200 W/m², while the boiler power in the other building was 23 W/m². The designer had never done this rule-of-thumb calculation, few do. Not surprisingly, the building's claimed efficiency did not materialise.

UBT's (2006) analysis of predictions and outcomes for buildings reviewed as candidates for a book and an award suggested the simpler the model, the smaller the performance gap. Perhaps sophisticated modelling (often performed by specialists) was distancing designers from the practicalities. To compare predicted results with component benchmarks and rules of thumb can be a useful reality check: might the values be too high, or unrealistically low?

5 Some benchmarking developments in the UK

5.1 Introduction

Benchmarking should not be an end in itself, but an effective way to help good things happen. A drill-down process (as developed in EPLabel 2006) can start with a simple entry level, but motivate users to want more. For example, if premises are used more intensively than an entry level assumes, the prospect of a better rating could be a business driver to dig deeper. It will often reveal new opportunities to save energy too.

5.2 Building Energy Certificates (EPCs)

The EU's Energy Performance of Buildings Directive (EPBD) (European Parliament & Council 2002) requires energy certificates for two different purposes:

- When a building is completed, sold or let. Here the UK's EPC is an asset rating based on modelled annual energy use by heating, hot water, cooling, ventilation and fixed lighting (known as 'regulated loads' in the UK), under standard conditions. (The UK uses similar calculations for building energy regulations, but with the benchmark based on a notional building similar to the proposed design, but with default efficiency values.)
- For display in certain public buildings > 1000 m² useful floor area (later cut to 500 m²). England and Wales chose DEC's using an Operational Rating based on metered energy for all end uses, and renewed annually to fit normal reporting cycles and motivate continuous improvement.

Performance ratings are calculated as follows (see CIBSE 2009 for DEC's):

- Annual energy supply requirements are calculated by energy source, each expressed as kWh/m² of usable floor area.
- These values are multiplied by national policy-determined factors (the UK chose CO₂e) and summed into a single energy performance indicator.
- This is compared with a benchmark. DEC benchmarks are derived from stock medians, as outlined in section 5.3.
- An efficiency rating is produced by multiplying the performance indicator (net of any onsite renewables) by 100 and dividing it by the benchmark.
- The rating is graded A–G, in increments of 25, so Grade A represents a rating 0–25, and G anything > 150.

As implemented, the asset-rating process had some unfortunate consequences. Designers tended to concentrate on modelled regulated loads only, contributing to the performance gaps now endemic. An emphasis on CO₂ allowed an inefficient building to be concealed by nominally low-carbon energy. The model used for building regulations also gave more credit to making active systems more efficient than to the careful execution of passive measures. This tempted some designers to add technical systems that may not have been necessary, because the resulting benchmark increase could make regulatory approval easier to obtain.

5.3 Benchmarks for DECs

In the 1990s, the UK government researched and published a wide range of energy consumption guides based on in-use performance, *e.g.* Office Guide 19 (DETR 1998). Many included headline indicators in multiple units—fuel, electricity, cost and CO₂—often broken down into end uses. In 2002, the Carbon Trust took this work over. However, when the European Parliament & Council (2002) mandated energy certificates, the associated benchmarking became a government responsibility. Since the Carbon Trust's remit was to go beyond policy obligations, development ceased and it merely republished the old guides (now archived on the CIBSE website).

In terms of peer groups, the UK guides tended to classify by characteristics, *e.g.* schools with and without swimming pools; and offices with and without air-conditioning (DETR 1998). In the US, EnergyStar (2020) examines statistical distributions and extracts influencing factors by regression. However, statistical analysis can easily be blind to influences that are evident when visiting a building, *e.g.* it has a large restaurant while its 'peers' may not.

A review of UK benchmarks for public buildings (EPLabel 2006) suggested the following:

- Many were out of date, not recognising that fuel use had fallen (with better insulation, more efficient heating and warmer weather), while electricity had risen (more equipment, longer operating hours, and more air-conditioning).
- Corrections for occupancy *etc.* varied between the energy consumption guides and sectors.
- Weather correction was both suspect and had not changed with the warming climate.
- End-use benchmarks were not always consistent with the reported totals.
- Data collection for different sectors was not consistent, so some apparent differences in energy use between similar building types might not reflect the underlying realities.
- 'Special' areas (*e.g.* data centres and swimming pools) varied greatly in size, specification and use. Instead of having different overall benchmarks for buildings that had them, they were better considered on their own merits.

In spite of these shortcomings, the government department responsible for energy certificates did not invest in operational benchmarking, but stuck to its traditional area of building regulations, extended to models and benchmarks for EPCs.

Instead, CIBSE (2008), with help from volunteers and its research fund, developed new, simpler and more consistent provisional values. After 80 stakeholders from public and commercial sectors agreed them to be an acceptable starting point, the government adopted most of the CIBSE proposals. Once DECs had been launched, CIBSE expected the government to review the DEC benchmarks every three to five years, using feedback from the database of certificates lodged. Sadly, this has never happened. However, CIBSE (2019) has launched a new benchmarking website, which includes distribution curves of DEC data for some public buildings, and the older data collated in CIBSE (2012) where nothing newer is available.

The approach adopted for DECs followed a scoping study for CIBSE (Bordass & Field 2007). This reviewed what existed and recommended starting again, with:

- a limited set of fuel and electricity use-related benchmarks, reflecting simple buildings of their type, in normal use, with typical occupancy levels and no special areas
- a simple entry level, based on energy use per m² of useful area, separately for fuel, heat and electricity, with a headline rating based on UK policy weightings, dimensionlessly graded A–G, with supporting detail included
- mandatory adjustments, in particular for regional weather year on year; adjustment for climate was not recommended, but some was finally included
- legislated but optional adjustments, available for use where the entry level did not take proper account of the situation, *e.g.* dense occupancy, or 'special' items such as regional server rooms and restaurants
- optional adjustments would only be accepted if examined rigorously using accredited procedures, including sub-metering special areas and end uses and reporting on their own energy efficiency and potential for improvement.

CEN (2005) defines the energy performance rating R as the ratio of the chosen headline indicator to the benchmark value in the same units. To assign an A–G grade, CEN put the typical (median) benchmark at the D–E boundary. It also suggested the B–C boundary should show current good practice. As this was not mandatory, the scoping study endorsed the EPLabel recommendation of a linear scale from zero to the median and beyond, graded in increments of 25% of the median. The dimensionless scale was straightforward to establish, would be identical for any energy source, end use or weighting system chosen, and addressed the policy goal of achieving net zero (in policy-preferred units). It also allowed mixed-use premises to be rated simply, using area-weighted sums.

To cover the UK stock of both public and commercial buildings, the scoping study suggested 17 benchmark categories to which different types of building could be assigned, replacing the published benchmarks for over 100 types. Stakeholder consultation introduced new ones, from filling stations to several different defence establishments, so the final publication (CIBSE 2008) includes 29 benchmark categories, each including median annual thermal and electrical use.

5.4 Complementary approaches to support development of DEC benchmarking

The authors of the scoping study saw statutory benchmarking for DECs as one of three complementary approaches:

- **Statistical:** widely used in voluntary benchmarking and by governments, ranking a building against its peers.
- **Technical:** going to the roots of energy consumption, in any depth the user would like, exposing possible causes and allowing 'what-if' calculations.
- **Statutory:** allowing the DEC system to draw on the first two methods. However, its prime concern is rigour, helping to direct users to achieve policy objectives whilst attempting to close loopholes.

Figure 4 shows the relationship between the approaches. The UK DEC benchmarks are based on what a building **does**, not what it is. So air-conditioned premises do not get bigger benchmarks at the entry level, while in voluntary benchmarking systems they can. However, if an air-conditioned building can demonstrate it is more intensively used, its benchmark may be increased.

The statistical approach shows where one is, but seldom why. It may inspire action, but gives no practical guidance. Peer group selection can also be problematic, so a high or low rating may not always reflect efficiency but missing attributes in categorisation or the statistical analysis. Mills (2016) shows how different metrics and peer group references can produce very different outcomes, going on to describe a 'features benchmarking' drill-down approach, where users can segment peer groups progressively. This was implemented as EnergyIQ, using data from California's Commercial End Use Survey (CEUS) (California Energy Commission 2020). For example, an office can first be benchmarked against the whole data set, then only offices occupied by information technology companies, and then only those in a particular region which also have variable air volume (VAV) air-conditioning. By that time, the peer group may however have become very small, so such a system will work best where it can draw upon large numbers of detailed records.

Statistical and technical approaches can be combined by pegging the attributes of a 'typical' building to median values from a statistical distribution. Good (or advanced)-practice benchmarks can then be calculated for identical use, but better fabric, engineering systems, controls, management *etc.* Such transparency between benchmarks and engineering values permits, for example:

- priorities to be assigned, *e.g.* to identify areas and end uses that use the most energy; and those with the greatest potential for improvement
- what-if calculations, *e.g.* 'what if we reduced the installed power density of the lighting from 5 to 3 W/m²?'; or 'what if better lighting controls were to reduce annual full-load operating hours from 2500 to 1500?'
- allowances for different patterns of use, *e.g.* 'what if we had an average occupation density'; or 'what if we didn't work at weekends?'
- identifying and adjusting 'special' energy end uses.

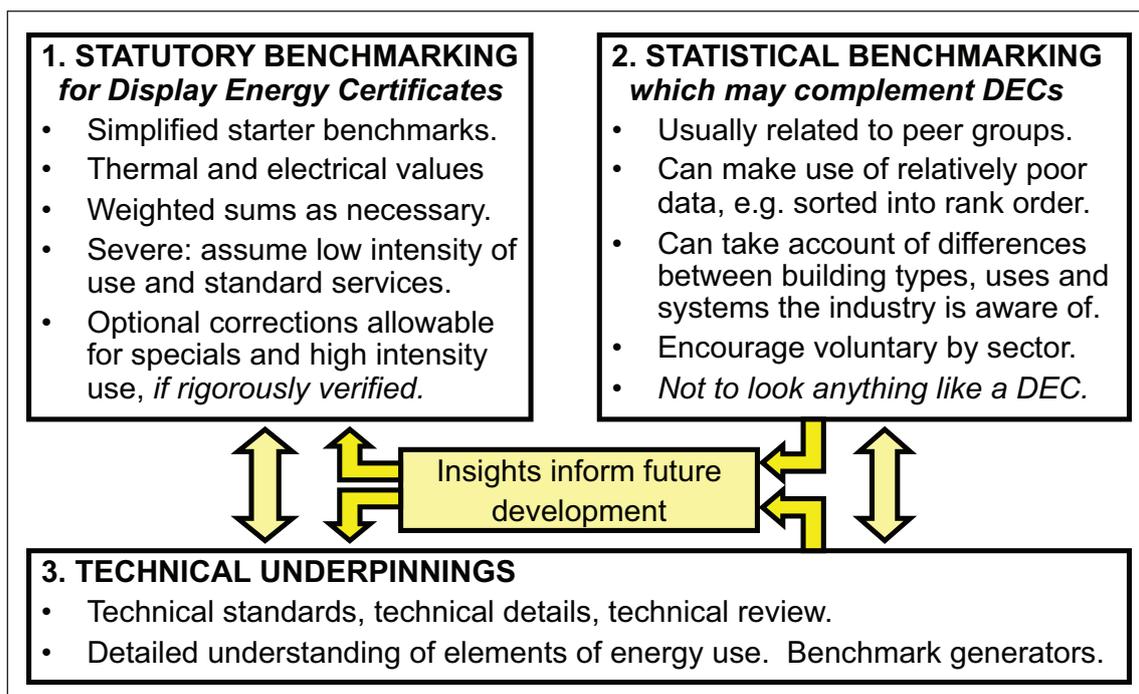


Figure 4: Three complementary approaches to benchmarking.

Note: DEC, Display Energy Certificate.

5.5 Benchmark generators and tailored benchmarks

The technical underpinnings can include benchmark generators that create realistic energy budgets bottom up from end-use and component values. This approach was initially developed in 1989–90 for Energy Consumption Guide 19 for Offices (ECON 19), the first produced under the government's then-new Energy Efficiency Best Practice programme. The background research included collecting technical details from 100 nominally energy-efficient offices, 25 of which became case studies (EEBPP 1994), most of which were published. These details were reconciled with published and statistical data from a range of sources, allowing the guide to include typical and good-practice benchmarks for fuel and electricity, as a whole, and split into nine categories of end use.

The second edition of ECON 19 (DETR 1998) was underpinned by an explicit Excel benchmark generator that used tree-diagram values, cross-referenced where possible to case studies and published rules of thumb (*e.g.* Boushear 2001).⁴ Consumption Guide 18 for Industrial Buildings used a similar approach. Guide 78 for Sports Centres (EEBPP 2001) added some simplified models, *e.g.* for swimming pool energy. 'Design sizing' prototype software was also produced, with Guide 78's component values replaced by ones appropriate for new buildings, and reviewed by practising design engineers. The software allowed stretching but realistic energy budgets to be established before design started. Predictions could then be reality-checked against these and in-use benchmarks, as the design developed.

Studying options for a new generation of consumption guides, in 2001–02, ECON 19 was developed into an Excel and web-based 'tailored benchmarking' prototype (Bordass *et al.* 2014). Typical and good-practice benchmarks were built up from component values and a list of attributes, including simplified schedules of accommodation and occupancy. Annual fuel and electricity use was shown as totals, by end use, and split between landlord and tenants. ECON 19's four types (naturally ventilated: cellular and open plan; and air-conditioned: standard and prestige) were no longer necessary—the software could re-create them.

Ironically, UK funding for in-use technical benchmark development fell between the two stools of government and the Carbon Trust just after Guide 78 was published and as the design sizing and office tailoring prototypes were being completed. The potential of the office system was, however, demonstrated in proof-of-concept EU DEC Excel software (Cohen, Bordass, & Field 2004). From a limited amount of information, this produced not only ratings and grades but also estimates of all end-use tree diagram values. It could then compare these with ECON 19-tailored benchmarks; work out the potential for energy-saving improvements; estimate budget costs and annual savings; and rank possible measures in order of likely cost-effectiveness. When policy-makers expressed concern about the (albeit modest) demands on users, a simple 'quick start' worksheet was added. This allowed the workbook to be initialised with very little input data: building type, size, fuel and electricity purchased, and servicing system. Users could stop there, or subsequently amend the more detailed input sheet where they wished.

The prototype DEC software also allowed experts to overwrite the automated values with their own insights to improve the breakdown of energy into end uses, the potential for making savings and the capital cost estimates. As any data were overwritten, all other estimates were automatically updated, so they remained compatible with the measured annual totals of fuel, heat and electricity consumption. For direct comparison with EPCs and design data, a subset of 'regulated loads' (fixed heating, hot water, ventilation, cooling and lighting—the EPBD's minimum set) could also be calculated and normalised to standard hours of use. Drop-down menus allowed users from different EU countries to select their languages and choose different units and weightings to suit local preferences.

This approach could also allow entry-level DEC certificates to be produced automatically, as the UK government can access gas and electricity meter readings through the Department of Business and property records from the Valuation Office Agency. However, in 2004 the UK regulator Ofgem decided this would burden the utilities, and that premises managers would need to ask and pay their suppliers for anything like this. Ironically, some utilities in the US sought and achieved free routine monthly uploads to Portfolio Manager (2020), regarding this as easier for them than providing data on request.

5.6 Future prospects

In spite of the urgent need to make energy performance in use visible, for nearly 20 years the UK government has not invested in operational energy benchmarking publications and software for general use. Nor have many other countries and regions. In California, funding has run out to maintain the promising EnergyIQ system (Mills 2016). If we do not really understand where we are, how can we know what to do in today's climate emergency?

In Australia, NABERS has demonstrated that operational ratings can motivate management to cut energy use substantially and progressively (Cohen *et al.* 2015). However, its greatest success has been for Base Building performance in offices. This benefited from good data from the outset, a trusted government-operated platform, purposeful engagement of the property industry all along and market pull—aided by the federal government's procurement policy.

In the UK, however, after more than a decade of use (and with a few notable exceptions), DEC in public buildings seem to have become more of a compliance ritual than a spur to improvement. This may not reflect a flawed process, but a total lack of government support in publicising or enforcing DEC, or in keeping the system and its benchmarks up to date. The underlying CO₂ units may not have helped either, distancing people from the realities of their actual energy use. DEC do include supplemental information and indicators, but in less detail than recommended by the technical advisory group, because the government preferred simplicity to transparency.

Tailoring may merit a second look in the climate emergency. It offers a practical and granulated approach to benchmarking and target-setting for existing, new and refurbished buildings, with transparency between policy and design expectations and in-use outcomes, and the ability to calculate a multitude of performance indicators from the one set of data. Bordass & Field (2007) saw prospects for a universal benchmark generator, drawing on an ever-growing library of end-use and component values to create benchmarks for a widening range of premises and Base Buildings.

Tailored benchmarking could be used for:

- establishing energy budgets at the briefing/programming stage
- managing expectations throughout the design and construction process
- establishing operational targets on completion of new works
- reviewing in-use performance and reporting outcomes
- developing and testing rules of thumb and
- putting policy in closer touch with the art of the possible.

6 Conclusions

Metrics are a means to an end, but always at risk of turning into the ends themselves. While metrics based on outcomes promise a clear goal without saying how to reach it, this paper has exposed the fallacy of single indicators as far as buildings are concerned: there needs to be more to grasp. Too few metrics may even lead people in the wrong direction. Many governments have multiple policy measures to save building-related energy and reduce GHG emissions, *e.g.* energy supply, building regulations, appliance standards, energy management and personal behaviour. These alone necessitate more than a single metric.

However, too many metrics may lead to mayhem. For a particular set of players, the sweet spot may be a selection that helps clarify their mission but lets them 'own' their specific problems. For rented offices, Australia found a leverage point (Meadows 1999) in its NABERS Base Building rating. This motivated a small but influential group of property owners and developers to reduce landlord energy year on year; which also brought along their service providers and helped to train the industry. The NABERS headline indicator was carbon based, but focused on outcomes, so the players soon learnt that energy saving was the cheapest way to start saving carbon. Contrast this with the UK, where the emphasis was on saving carbon in theory, not in practice; and where operational rating systems were neglected.

This paper has argued that more diverse reporting and benchmarking could make a big difference, helping people to play their part in reaching overall goals:

- Policies and reporting systems become less focused on single headline indicators and take more account of supplemental indicators that make underlying contributions more explicit, and avoid masking the raw data.
- A move away from purely statistical benchmarking, *e.g.* with buildings in the top quartile seen as efficient, when they may just be the best of a bad bunch.
- Greater transparency between policy, client, and design intent and operational reality, *e.g.* reporting and benchmarking by component at both the design stage and in use, and comparing the results.
- A move from 'design for compliance' to in-use performance, with devices such as DEC's and Landlord's statements and ratings making this visible and actionable by the industry, its clients and the property market.
- Use the Base Building concept for all managed buildings, *e.g.* landlord's services in apartment blocks. Estimates of both total and Base Building energy should be a mandatory, as should metering of the Base Building.
- Government support for this at policy level, including open databases that make high-quality information available to all.

This transparency will help to motivate people and release multiplier effects. It will expose emergent problems that need addressing. It will also reveal unexpected successes. Time and again (*e.g.* Palmer & Armitage 2014), post-occupancy evaluations show that unmanageable complication is the enemy of good performance, and that too much technology brings with it problems of effective control, usability, support costs and premature obsolescence. With care and thought, today's new buildings can also perform much better, as NABERS (2020) Commitment Agreements have shown in Australia. A truly sustainable, low-carbon built environment will need to achieve much more with much less, and will require radical reductions in both embodied and operational energy and carbon.

Notes

- ¹ H. Bruhns, University College London, private communication, 2005.
- ² Atlantic Boilers, private communication, 1992.
- ³ J. Field, private communication, 2020.
- ⁴ The fifth edition is now available.

Competing interests

The author has no competing interests to declare.

References

- ACE.** (2001). *Flying Blind—All you wanted to know about energy in commercial buildings but were afraid to ask*. London: Association for the Conservation of Energy (ACE). Retrieved January 2, 2020, from www.usablebuildings.co.uk/UsableBuildings/Unprotected/FlyingBlind.pdf
- Bannister, P., & Zhang, H.** (2014). What simulation can tell us about building tuning. *Ecolibrium*, December, 33–37.
- BBP.** (2020). *Design for performance*. London: Better Buildings Partnership (BBP). Retrieved January 2, 2020, from www.betterbuildingspartnership.co.uk/node/360
- BEIS.** (2018). *Sub-national electricity and gas consumption statistics*. London: Department of Business, Energy and Industrial Strategy (BEIS).
- BEIS.** (2019). *Digest of UK energy statistics*. London: Department of Business, Energy and Industrial Strategy (BEIS). Retrieved January 2, 2020, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/794590/updated-energy-and-emissions-projections-2018.pdf
- Bordass, B., Cohen, R., Spevak, R., Burman, E., Hong, S., Ruyssevelt, P., & Field, J.** (2014). Tailored energy benchmarks for offices and schools, and their wider potential. *Paper presented at the CIBSE–ASHRAE Technical Symposium*, Dublin, Ireland, 3–4 April.
- Bordass, B., & Field, J.** (2007). Energy and CO₂ emissions benchmarks for non-domestic buildings. Unpublished report for Chartered Institution of Building Services Engineers (CIBSE), London.
- Boushear, M.** (2001). *Rules of thumb*, 3rd edn (BSRIA Technical Note No. 15/2001). Bracknell: Building Services Research and Information Association.
- BPF.** (2007). *Landlord's energy statement and tenant's energy review*. London: British Property Federation (BPF). Retrieved January 2, 2020, from www.les-ter.org
- BRE.** (2019). *The government's Standard Assessment Procedure for energy rating of dwellings (SAP 10.1)* (September). Garston: Building Research Establishment (BRE).
- California Energy Commission.** (2020). California commercial end use survey [CEUS]. Retrieved May 21, 2020, from <https://www.energy.ca.gov/data-reports/surveys/california-commercial-end-use-survey>
- CEN.** (2005). *Standard EN 15203: Energy performance of buildings—Assessment of energy use and definition of ratings*. Brussels: Comité Européen de Normalisation (CEN).
- CIBSE.** (2006). *Technical Memorandum TM22: Energy assessment and reporting method*. London: Chartered Institution of Building Services Engineers (CIBSE).
- CIBSE.** (2008). *Technical Memorandum TM46: Energy benchmarking*. London: Chartered Institution of Building Services Engineers (CIBSE).
- CIBSE.** (2009). *Technical Memorandum TM47: Operational ratings and Display Energy Certificates*. London: Chartered Institution of Building Services Engineers (CIBSE).
- CIBSE.** (2012). *Guide F: Energy efficiency in buildings* (ch. 19). London: Chartered Institution of Building Services Engineers (CIBSE).
- CIBSE.** (2013). *Technical Memorandum TM54: Evaluating operational energy performance of buildings at the design stage*. London: Chartered Institution of Building Services Engineers (CIBSE).
- CIBSE.** (2019). *Energy benchmarking tool*. London: Chartered Institution of Building Services Engineers (CIBSE). Retrieved January 2, 2020, from www.cibse.org/knowledge/energy-benchmarking-tool-beta-version
- Cohen, R., Austin, B., Bannister, P., Bordass, B., & Bunn, R.** (2017). How the commitment to disclose in-use performance can transform energy outcomes for new buildings. *Journal of Building Services Engineering Research and Technology*, 38(6), 711–727. DOI: <https://doi.org/10.1177/0143624417711343>
- Cohen, R., Bannister, P., & Bordass, B.** (2015). Nearly zero energy buildings in reality, not just theory. *REHVA Journal*, May, 57–59.
- Cohen, R., & Bordass, B.** (2015). Mandating transparency about building energy performance in use. *Building Research & Information*, 43(4), 534–552. DOI: <https://doi.org/10.1080/09613218.2015.1017416>
- Cohen, R., Bordass, B., & Field, J.** (2004). Grading non-domestic buildings for their energy performance. *Paper presented at the IEECB Building Performance Congress*, Frankfurt, Germany, April.
- Cohen, R., Bordass, B., & Field, J.** (2006). EPLabel: A graduated response to EPBD energy certification based on an operational rating. *Paper presented at the IEECB Building Performance Congress*, Frankfurt, Germany, April.
- DETR.** (1998). *Energy consumption guide 19: Energy use in offices*, 2nd edn. London: Department of the Environment, Transport and the Regions (DETR). Retrieved January 2, 2020, from [www.cibse.org/getmedia/7fb5616f-1ed7-4854-bf72-2dae1d8bde62/ECG19-Energy-Use-in-Offices-\(formerlyECON19\).pdf.aspx](http://www.cibse.org/getmedia/7fb5616f-1ed7-4854-bf72-2dae1d8bde62/ECG19-Energy-Use-in-Offices-(formerlyECON19).pdf.aspx)
- EEBPP.** (1994). *General Information Report GIR 15: Technical review of office case studies and related information*. Energy Efficiency Best Practice programme. Retrieved April 26, 2020, from www.usablebuildings.co.uk/UsableBuildings/Unprotected/OfficeCSTechReviewMar94.pdf
- EEBPP.** (2001). *Energy consumption guide 78: Energy use in sports and recreation buildings*. Energy Efficiency Best Practice programme. Retrieved May 21, 2020, from <https://www.cibse.org/getmedia/34def23a-c65b-405e-9dff-ce181c0b1e0d/ECG78-Energy-Use-in-Sports-and-Recreation-Buildings.pdf.aspx>

- EnergyStar.** (2020). Portfolio Manager. Retrieved January 2, 2020, from www.energystar.gov/buildings/facility-owners-and-managers/existing-buildings/use-portfolio-manager
- EPLabel.** (2006). The EPLabel project ceased in 2006 but some of its benchmarking work is here. Retrieved January 3, 2020, from https://ec.europa.eu/energy/intelligent/projects/sites/iee-projects/files/projects/documents/the_eplabel_benchmarking_system.pdf
- European Commission.** (2019). *The European Green Deal* (COM(2019), 640 final, 11 December). Brussels: European Commission.
- European Parliament & Council.** (2002). *Directive 2002/91/EC on the energy performance of buildings*. Brussels: European Commission.
- European Parliament & Council.** (2018). *Directive (EU) 2018/844 amending Directive 2010/31/EU on the energy performance of buildings and Directive 2012/27/EU on energy efficiency* (30 May). Brussels: European Commission.
- Europrosper.** (2002). Research into the feasibility of climate corrections in EU Display Energy Certificates. Unpublished.
- Field, J., Soper, J., Jones, P., Bordass, W., & Grigg, P.** (1997). Energy performance of occupied non-domestic buildings: Assessment by analysing end-use energy consumptions. *Building Services Engineering Research and Technology*, 18(1), 39–46. DOI: <https://doi.org/10.1177/014362449701800106>
- Flyvbjerg, B.** (2006). Five misunderstandings about case study research. *Qualitative Inquiry*, 12(2), 219–245. DOI: <https://doi.org/10.1177/1077800405284363>
- Goodman, P. S.** (2009, September 22). Emphasis on growth is called misguided. *The New York Times*. Retrieved from <https://www.nytimes.com/2009/09/23/business/economy/23gdp.html>
- Gram-Hanssen, K., & Georg, S.** (2018). Energy performance gaps: promises, people, practices. *Building Research & Information*, 46(1), 1–9. DOI: <https://doi.org/10.1080/09613218.2017.1356127>
- IEA.** (2006). *Mind the gap—Quantifying principal-agent problems in energy efficiency*. Paris: International Energy Agency (IEA).
- IPD.** (2013). *NABERS Energy Analysis* (June). London: Investment Property Databank (IPD) and the Australian Department of Industry.
- ISO.** (2013). *Standard 12655: Energy performance of buildings—Presentation of measured energy use of buildings*. Geneva: International Organization for Standardization (ISO).
- IUK.** (2012). Unpublished development of the CIBSE (2006) TM22 software, used to collate information from Innovate UK's Building Performance Evaluation programme.
- Meadows, D.** (1999). *Leverage points: Places to intervene in a system*. Hartland: The Sustainability Institute.
- MHCLG.** (2019). *Consultation document, Approved Document L: Conservation of fuel and power, Volume 1: dwellings* (October). London: Ministry of Housing, Communities and Local Government (MHCLG).
- Mills, E.** (2016). Action-oriented energy benchmarking for nonresidential buildings. *IEEE Proceedings*, 104(4), 697–712.
- NABERS.** (2020). Our story. NABERS. Retrieved January 2, 2020, from www.nabers.gov.au/about/our-story
- Palmer, J., & Armitage, J.** (2014). *Building Performance Evaluation Programme: Early findings from non-domestic projects*. Swindon: Innovate UK.
- UBT.** (2006). The edge voluntary energy and CO₂ declaration (VECD) for buildings in use (unpublished final overview report). London: Usable Buildings Trust (UBT).
- UBT.** (2011). Notes on benchmarking building energy performance for occupation density (Unpublished report, August). London: Usable Buildings Trust (UBT) for the Chartered Institution of Building Services Engineers (CIBSE).
- UK Green Building Council.** (2019). *Net zero carbon buildings: a framework definition*. London: UK Green Building Council.
- UNEP.** (2020). *Facts on the climate emergency*. United Nations Environment Programme (UNEP). Retrieved April 26, 2020, from <https://www.unenvironment.org/explore-topics/climate-change/facts-about-climate-emergency>
- Vivid Economics & Imperial College London.** (2019). *Accelerated electrification and the GB electricity system*. Report to the Committee on Climate Change (CCC), April.
- Waide Strategic Efficiency.** (2014). *The scope for energy and CO₂ savings in the EU through the use of building automation technology* (Report for the European Copper Institute). Retrieved April 30, 2020, from http://neu.eubac.org/fileadmin/eu.bac/BACS_studies_and_reports/2014.06.13_Waide_ECI_-_Energy_and_CO2_savings_BAT.pdf
- Watttime & Rocky Mountain Institute.** (2017). On the importance of marginal emissions factors for policy analysis. Retrieved January 3, 2020, from www.bloomenergy.com/sites/default/files/watttime_the_rocky_mountain_institute.pdf

Appendix A: The Landlord's Energy Statement (LES)

Figure A1 shows part of page 1 of the LES (BPF 2007) for one tenant in a multi-tenanted office. Another version (intended for the landlord and property industry statistics) shows total landlord energy use and how it is assigned to each tenant and any other end uses, e.g. the landlord's management office or a mobile phone mast. The LES aims to be explicit about all major components of the energy use and CO₂e emissions reported.

Not shown is page 2, which includes background details, including total landlord consumption (and production, if any) by source, apportionment to tenants, contracted hours of operation for each tenant, any extra hours, and contextual data used in benchmarking. Landlords preferred to keep energy costs separate.

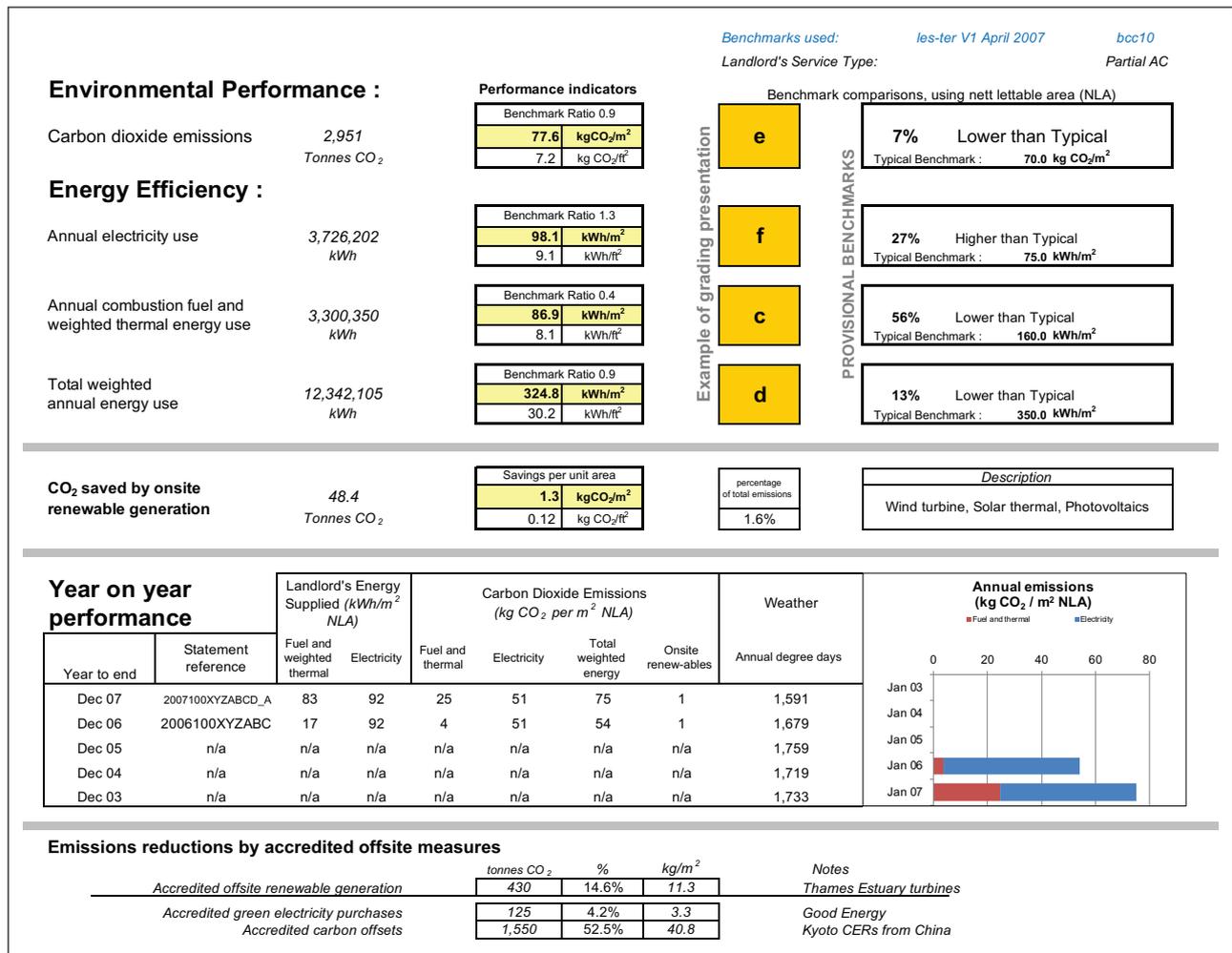


Figure A1: Landlord's Energy Statement (LES), extract from page 1.

How to cite this article: Bordass, B. (2020). Metrics for energy performance in operation: the fallacy of single indicators. *Buildings and Cities*, 1(1), pp. 260–276. DOI: <https://doi.org/10.5334/bc.35>

Submitted: 14 January 2020 **Accepted:** 01 June 2020 **Published:** 30 June 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.