A classification of building energy performance indices

David B. Goldstein · Charles Eley

Abstract Energy performance indices are used around the world to evaluate and monitor residential and commercial building energy performance during design, construction, renovation, and operation. The two most common indices are Asset Ratings and Operational Ratings. Asset Ratings are based on modeled energy use with uniform conditions of climate, schedules, plug loads, occupancy, and energy management. Operational Ratings are based on measured energy use, often normalized for relevant variables like climate and level of energy service. Surprisingly, there is almost no discussion in the literature about the technical basis of these ratings and what they are attempting to measure. This paper analyzes the merits and weaknesses of the common ratings and introduces additional energy performance indices, in particular the Operation and Maintenance (O&M) Index, which is the ratio of the energy consumption as measured at the meter to the simulated energy performance, calibrated for the actual operating conditions of the building. We provide examples of how such indices are currently used, although we do so as examples to illustrate our hypothesis as to what indices are most helpful to improve energy management, rather than as a comprehensive review. We show how these indices work together to provide better feedback to energy managers. The Operational Rating answers the question: "how does the energy intensity of this building compare to its peers?" The Asset Rating answers the question: "how efficient is this building?" The O&M Index answers the question: "how well is this building being managed?" These questions are useful to answer in the context of a comprehensive energy management program, such as would be required by an Energy Management System standard.

Keywords Benchmarking · Energy performance indices (EnPIs) · Managing energy performance · O&M index · Operational ratings · Asset ratings

D. B. Goldstein (✉)
NRDC, 111 Sutter Street, 20th Floor, San Francisco,
CA 94104, USA
e-mail: dgoldstein@nrdc.org

C. Eley
Eley Consulting, 342 Green Street, San Francisco,
CA 94133, USA
e-mail: charles@eley.com

Introduction

ISO 50001

Energy Management practices can be implemented systematically through Energy Management System standards. The International Organization for Standardization issued the first version of its standard on Energy Management System, ISO 50001, in June 2011 (ISO 2011). It requires the use of Energy Performance Indicators, or EnPIs, to monitor whether a building is on track in meeting goals of continual improvement in energy performance. EnPIs are the key feature of the data-driven approach in the Standard. The overall structure requires the qualifying organization:

- to adopt an energy policy at the executive management level;

- to create a cross-divisional management team led by a representative who reports directly to top management;

- to undertake an energy planning process that includes an energy review;

- to create baseline(s) of the organization's energy use;

- to track energy performance indicators against the baseline to measure progress;

- to develop energy objectives and targets for energy performance improvement at relevant functions, levels, processes or facilities;

- to develop action plans to meet those targets and objectives;

- to establish operating controls and procedures for significant energy uses;

- to measure , manage , and document energy performance for continual improvement for energy efficiency; and

- to report progress periodically to top management.

This Standard does not require any specific EnPIs, but this paper recommends a structure for how to select, measure, and compute EnPIs for buildings. The intent is to characterize building energy performance with indicators that span the full range of issues that energy managers must address and to allow the energy management plan to address both physical changes to the building and operational effectiveness in the context of a plan for continual improvement. A similar structure of EnPIs also applies to industry (Goldstein et al. 2011).

Different forms of energy performance indices (EnPIs) are used around the world to evaluate and monitor the energy performance of both residential and commercial buildings. These indices are intended to inform decisions while the building is being designed, constructed, renovated, or operated.

Building energy management is most effective when it is based on quantitative measurements and predictions. Building energy performance is the result of the interaction of an engineered system with operation and maintenance (O&M) practices and with occupant demands and behavior. Since each of these three dimensions of energy performance—the engineered system, O&M practices, and occupant needs—is largely independent, three EnPIs are necessary to describe and manage the building and isolate these factors. Attempts to characterize building energy performance with fewer than three EnPIs are likely to fail, as a three-dimensional space cannot be described with less than three parameters.

Building energy management may be undertaken in accordance with an energy management system standard. A prominent example of such a standard is ISO 50001, which is described in the sidebar.

This standard helps the user to understand one important context in which the different EnPIs might be used: to evaluate progress towards targets for continual improvement in the various aspects of building construction and operation. An effective EnPI allows management to track progress towards a goal for that metric and to get good feedback as to how the plan is progressing. A less effective EnPI may show improvement when the underlying activity or system is not really improving, or conversely.

Design and construction establishes the inherent energy efficiency of a building. Building managers can improve energy performance or allow it to degrade through operation and maintenance practices.

We use the term "energy performance" as it is used in ISO 50001: to define the problem in terms that are broader than energy efficiency. EnPIs should, we assert, address all three dimensions of energy performance, and to the maximum extent possible distinguish between them. Attempts to reduce building energy use by compromising energy service levels are resisted by the real estate industry and by building occupants because often they do not make economic sense. The value of retail sales in a store, or the cost of salaries of workers in an office building, or the value of

medical services provided in a hospital, exceed the cost of energy by one or two orders of magnitude, so compromising the main function of the building to save energy is clearly a departure from optimality.

We define energy efficiency as the provision of a constant level of energy service while using less energy. We do this because "one very rarely encounters an explicitly stated definition of 'energy efficiency'" (National Academy of Sciences 2010); thus the US National Academy study discusses several alternate definitions. This paper's definition aligns with the primary definition used in that study's discussion of the buildings sector, and is used as that study uses it, to distinguish efficiency from conservation, which includes both improvements in O&M procedures and reductions in comfort or other energy service levels.

This paper does not use the concept of *conservation* because of our desire to distinguish between O&M effectiveness and energy service level. Thus, efficiency, as used in this paper, refers to design and technology. It usually can be controlled, and always can be influenced, by building management.

O&M procedures involve both occupant and management behavior. They are a critical piece of energy management because some companies are able to demonstrate substantial improvement in energy intensity year after year based overwhelmingly on non-capital measures.[1]

Occupants determine the level of energy service that is provided by the building, including hours of operation, density of energy-using equipment, and comfort requirements. These energy service demands are usually outside the scope of what a building energy management plan can address, so they are taken as a given. Energy performance can therefore best be monitored if there are EnPIs that can normalize metered energy use for a constant level of energy service.

There are two common types of building EnPIs (Maldonado 2011). The Asset EnPI or Asset Rating is based on modeled energy use (taking into account physical measurement of relevant characteristics of the building) with uniform conditions of climate, schedules, plug loads, occupancy, and energy management. Asset Ratings are analogous to the Coefficient of Performance

(COP) rating for air conditioners (as reported to the consumer based on laboratory tests under standard conditions) or to fuel economy ratings for automobiles (which similarly are derived from standard protocols based on laboratory tests).

Sidebar on Asset and Operational Ratings

There has been a lot of confusion about the relative meaningfulness of Asset Ratings as compared to Operational Ratings. The following examples for television sets and automobiles illustrate the differences and provide insights on how each may be used.

Televisions

Televisions are a rapidly growing source of energy consumption as screen size becomes bigger and people continue to watch more video content, both at home and in commercial establishments. The energy consumption of TVs depends on many factors. Some, such as the brightness of the screen, hours of operation, and program content, are related to how the TV is used. Other factors such as LED back lights, electronics, and standby controls are features that affect energy efficiency under all conditions.

An Asset Rating for a TV focuses on its inherent energy efficiency features and attempts to rate TVs using a standard pattern of use. This is achieved by measuring the short-term energy consumption of a television under a tightly-defined set of conditions ("test procedure"). The test procedure includes assumptions about the brightness level and the program content. The test involves a specific video loop. Predictive equations are then used to estimate annual energy consumption. As such, an Asset Rating is useful to potential buyers, as they can compare the energy efficiency of products under a standard set of operating conditions.

By comparison, an Operational Rating for TVs measures the energy consumption over some period of time as the TV is subjected to a variety of operating conditions. The measurement could then be compared to an average for all TVs of that class or to some other baseline or standard. The measurement is easy to make, but it is difficult or impossible to discern what is causing one set's measurement to be high and another's low. It could be that in one case, the TV was used for longer hours or in a bright room, while in another case the measurements could be low because the TV was used in a dark room (lower brightness) and for fewer hours.

There is a role for field measurements, but to use such measurements for a standard would result in considerable expense and offer little intellectual clarity on the causes of high energy use or what could be done to reduce it. By contrast, the simple test protocol of Asset Rating predicts the energy consumption for a typical user, based on a handful of measurements.

Everyone who buys a television ought to recognize that the energy consumption of the TV in their own home or business may differ, often radically, from the energy consumption on the rating. A consumer that watches the television half as many hours per year as the test procedure obviously will use about half the energy. A consumer who watches the television in a dark room will generally consume much less energy than one who operates it in a bright room, since the screen brightness in many new TV models automatically adjusts to real-time measurements of ambient brightness. The owners of a sports bar will recognize that not only are the hours of operation likely to be much longer, but the level of brightness that the bar manager chooses may exceed the level in the test procedure.

---

[1] Personal communication with Steven Schultz, energy manager for the 3M Company, August 2009.

Thus, the Asset Rating is used to evaluate which set to buy and the wise energy manager will adjust the values in the Asset Rating to assess actual usage patterns. Likewise, Operational Ratings or in-situ measurements would be useful to a sophisticated energy manager in a facility with lots of similar televisions to identify sets that are using more energy than expected and to correct the problem.

Another issue with Operational Ratings is self-selection bias. For instance, larger TV sets tend to be used in brightly lighted spaces for longer hours, *e.g.,* in retail establishments and transportation terminals. The conclusion might be that larger TVs are less efficient, but this is the wrong message, especially since Asset Ratings tend to get better (in energy use per square centimeter of screen) with size.

Automobiles

Asset Ratings for automobiles have been used for decades in many countries. In the U.S. the metric is miles per gallon (MPG), while most other countries use liters of fuel per 100 kilometers of travel. These are widely accepted as measures of a car's fuel economy (we do not use the word "efficiency" here because the metric does not control for size or performance). Both are determined by driving the car through a test track or dynamometer at precisely controlled speeds such that every car is subjected to the same conditions. Asset Ratings for cars are taken into account by most buyers when they make a purchase, although all understand that "your mileage may vary".

One could also measure a car's fuel economy by tabulating the liters of fuel used per year. This would be an Operational Rating. However, the rating would be affected by the distance driven, the maintenance record of the car, the type of driving done (for example, pulling a trailer up mountains or hauling heavy tools versus commuting to work on an expressway), and the care taken by the driver (such as smooth driving versus lots of sudden acceleration and deceleration). The Operational Rating could be useful to compare the driving habits and needs of one family to another, but would be of little value in informing a purchase.

With automobiles, there is likely a large problem of self-selection bias, like for TVs. One would expect a high-performing hybrid to be used preferentially for families that need to drive a lot or for taxicab services, while one similarly would expect a light truck to be used for hauling equipment and tools to job sites. Both would artificially inflate the expected fuel consumption compared to a method that normalized for energy service level and O&M quality. There are other potentially offsetting factors: some have argued that hybrid drivers are more sensitive to fuel economy and drive more gently, or use their bicycles more. The point is that the complications are hard or impossible to remove from an analysis that depends solely on Operational Ratings.

We would no doubt find some spurious but statistically significant correlations. For example, if we measured the number of liters of fuel consumed by an automobile as a function of its age, no doubt we would find that 50-year-old cars use the least energy. But it would be wrong to infer that the best efficiency policy would be to get everyone to renovate cars from the 1950s.

The Operational EnPI or Operational Rating is based on metered or measured energy use. The Operational Rating takes account not only of the physical characteristics of the building (the building asset) but also the level of energy service provided and how it is operated and maintained.

This paper discusses and contrasts the merits of Asset Ratings and Operational Ratings, and also suggests the use of a two additional EnPIs, which we call the O&M Index and the Energy Service Index. The latter two EnPIs are not ratings that are useful to disclose, but rather ratios useful for energy management decisions.

- The O&M Index is the ratio of the energy consumption as measured at the meter to the simulated energy performance from the models used to determine the Asset Rating. But in contrast to the simulation used for the Asset Rating, the O&M Index accounts for the actual conditions of building operation.
- The Energy Service Index is the ratio of simulated energy performance of the rated building at its observed level of energy service to the energy performance of the rated building at the standard level of energy service assumed for the Asset Rating.

We show that the Asset Rating, the O&M Index, and the Energy Service Index may be used together to provide better feedback to energy managers on inherent building energy efficiency, operation and maintenance, and occupant demands.

In contrast, the Operational Rating is a more holistic yet simple-to-derive value that can encourage better energy management practices, particularly when used at the senior executive level as part of an Energy Management System approach, but that includes too many factors to be very helpful as a tool to accomplish or accurately measure progress toward any specific energy performance goal.

The recommended EnPIs provide quantitative answers to relevant questions about building energy performance management. The Operational Rating provides an overview of the effect of all aspects affecting building energy performance. While it can motivate better energy management, it does not offer clear directions for how to do so: Low energy use might be an indicator of high efficiency, or it might be a consequence of exceptionally effective O&M in an inefficient building. Alternately, it might be an indicator of very low tenant demands for energy services.

Operational Ratings may also be susceptible to self-selection bias. The PlaNYC study (Hsu et al. 2012) found that new buildings had higher EUIs than older buildings. One could attribute this trend to the hypothesis that newer buildings are less efficient. But this hypothesis would amount to saying that energy codes do not save energy,

when the bulk of evidence overwhelming corroborates the hypothesis that they do. Instead, virtually every expert we have talked to on commercial real estate in New York, both business people and building science experts, argues that newer buildings are more likely to be Class A buildings that provide more amenities—more energy services—and also attract tenants with higher needs for IT, comfort, catering services, etc.

Nonetheless, the Operational Rating is an extremely valuable EnPI to report to top management and to the public and thus can motivate more detailed analysis at the operational level of building management that may rely more heavily on the other EnPIs. The selection of effective EnPIs varies depending on the scope of responsibility of the user of that EnPI (Goldstein and Almaguer 2013).

There is surprisingly little discussion in the literature about the reasons why a particular index is useful or optimal, or how it should be derived in principle.[2] A few papers and books discuss how Asset Ratings or Operational Ratings can be developed in a particular context, but there seems to be no source offering a framework discussion on what these ratings are intended to accomplish, or of why they are best derived in one particular way rather than another. These sorts of policy discussions appear to be limited to informal or implementation-focused articles (Crowe and Falletta 2012; Graden et al. 2008), conversations, and write-ups, without the opportunity for serious scientific discussion on the theory and methodology.

As a result of this thin literature, discussions about the relative merits of the different EnPIs have been informal, and have not resulted in scientific clarity or the ability to resolve differences in opinion based on the scientific method of forming clear hypotheses and analyzing data in a way that is intended to corroborate or falsify a rigorous hypothesis. This paper attempts to fill this gap by providing such a framework. It offers the possibility of forming hypotheses that can be tested against the data gathered both in formal studies and in analysis of raw data provided by examining the outcomes of real-world rating and labeling

systems. And it suggests hypotheses about energy simulation that can be tested scientifically, allowing a continual improvement process of "Plan, Do, Check, Act" as required in ISO 50001, to upgrade the quality of both energy models and of the assumptions used in their inputs.

While this paper presents examples of how its concepts are realized in various places globally, it is not a review paper in the sense that we do not try to infer relationships inductively by examining best practices. Nor do we attempt to be comprehensive in reviewing rating systems worldwide. Instead, we propose hypotheses about what types of EnPIs ought to be effective, based on building science and on existing Management System Standards, and then look to regional examples to see the extent to which these trials validate or refute these hypotheses (Table 1).

Analytic framework: Asset Ratings

Asset Ratings are based on simulated energy performance. The simulation is based on physical measurements of characteristics such as wall areas, window areas, thermal conductance, air leakage, etc., combined with reported measurements from manufacturers, such as the efficiency of a boiler or the wattage of a motor (Maldonado 2011). An Asset Rating is reported as a value of energy consumption, usually a ratio but sometimes an absolute number. We make the case later that Asset Ratings generally are best expressed as ratios. Examples of Asset Ratings include energy code compliance that uses the performance approach, the American "Home Energy Rating System" or "HERS index" that establishes a score of 100 for a house that meets the model US energy code as of a fixed date and 0 for zero net-energy home (RESNET 2012), and various Asset Ratings in use in both the USA and many member states of the European Union.

Asset Ratings are the exclusive energy rating method for most common energy-using systems such as automobiles, refrigerators, and clothes washing machines, for reasons noted in the sidebar. They are also most common for single-family residential buildings. Point scores or check-lists of efficiency features are not Asset Ratings, although the ratios we recommend often are reported as star ratings or letter grades.

Asset Ratings, if correctly implemented, isolate the effect of the building asset by assuming standard operating conditions for energy service and O&M. Asset Ratings are necessarily derived from simulation, since

---

[2] After the authors performed an extensive literature review themselves and found virtually nothing relevant other than the sources used to inform the sidebar on "Energy Use per Square Meter as an EnPI", we consulted other experts in the field to make sure that we were not missing something. In personal communications with David Eijadi, Prasad Vaidya, Philip Fairey, and Liu Xiang in November 2012, these practitioners in the energy simulation field were unable to identify any relevant literature in this field, either.

Table 1 Relationship of EnPIs to the dimensions that determine energy performance

EnPI	Dimensions of Energy Use in Buildings		
	Level of Energy Service	Operation and Maintenance	Efficiency of Building Asset
Asset Rating			███ (black)
Operational Rating	███ (red)	▓▓▓ (red hatched)	▓▓▓ (red hatched)
O&M Index		███ (black)	
Energy Service Index	███ (black)		

███ (black) EnPI focuses on single dimension

███ (red) Some EnPIs adjust for energy service, but the adjustments are incomplete (see Table 2)

▓▓▓ (red hatched) EnPI includes the effects of both the building asset and operational and maintenance practices

this is the only reasonable way to apply standard and identical operating conditions such as weather, O&M, and energy service. Physically measuring energy performance under such controlled conditions would be prohibitively expensive in general, and perhaps impossible for climate, whereas simulation is a simple and inexpensive way to assure that efficiency differences are not being confounded with differences in building operation or weather.[3]

Asset Ratings usually are expressed as the ratio of the energy performance of the rated building to the energy performance of a baseline or reference building. The baseline building normally is assumed to have the same conditioned floor area and general configuration as the rated building, although a few energy codes limit the maximum size of residential baseline buildings.

The energy performance of both the rated building and the baseline building are determined through energy models using standard schedules of operation, plug loads, temperature settings, and other operational characteristics. These and other required assumptions are a key part of a successful Asset Rating system. We refer to these standard characteristics as neutral independent (NI) operating assumptions (*neutral* because they are the same for both the baseline building and the rated building and *independent* because they are prescribed independently of any choice made for the rated building). This nomenclature is more explicit and less subject to misinterpretation than more commonly used and parallel terms such as "normalized" or "standardized" because all possible combinations of

neutrality and independence may be used in deriving the EnPIs discussed here. The technical basis of the asset rating may be expressed as shown in Eq. 1.

$$Asset\ Rating = Asset\ EnPI = \frac{EP_{RB,NI}}{EP_{BB,NI}} \qquad (1)$$

where

$EP_{RB,NI}$ The energy performance of the rated building determined from an energy model. The "NI" subscript means that neutral independent modeling assumptions are used.

$EP_{BB,NI}$ The energy performance of the baseline building determined through the same modeling procedure. The same neutral independent modeling assumptions are used as for the rated building.

The energy models and assumptions used in certain Asset Rating systems are quite good at predicting metered energy use (Hassel et al. 2009; Johnson 2003), on average. For residential buildings in the USA, the Asset Rating was within 3 % of the metered average for cooling energy and 4 % for heating energy in Houston. This agreement is a consequence of two factors: the accuracy of the simulation model and the validity of the operating assumptions that the Asset Rating system being evaluated requires modelers to use. Both factors are essential for Asset Ratings to have the value as effective EnPIs that we discuss next. More discussion of the ability of energy models to predict measured energy use accurately, both on average and for particular buildings, is provided in the section "Accuracy of asset rating systems and of energy models".

Simulation models used in the context of Asset Rating systems are even better at predicting relative energy use; that is, the difference between one design option and another, while keeping operational factors

---

[3] There have been only a few research projects that compare modeled results to metered results for unoccupied buildings, carefully controlled to maintain identical conditions. One could also compare results in which the analyst allows occupants if their behavior is monitored on an hourly or more-frequent basis. But these are research projects that are orders of magnitude too complex and expensive to be used for ratings of real buildings.

neutral between the options. Expressing the Asset Rating as a ratio of energy performance using the same modeling tool, climate data, and operating assumptions for both the rated building and the baseline building takes advantage of this benefit. If the model predictions are high or low, or if the weather data is a little off, energy performance predictions for both the rated building and the baseline building are off in the same direction and the ratio between the two is relatively unchanged. Using a ratio allows errors to cancel out to first order, whether the errors are due to weaknesses in the simulation algorithms, errors in the input of the building characteristics, or errors in specifying typical operating conditions.

Note that some modeling differences are due to how controls are treated. Asset Rating standards typically contain control credits, in which a given control (such as occupant-accessible manual dimming of specific luminaires) is assumed to be used in a fixed way to reduce energy use. Modeling rules for control credits are used consistently with both the rated building and the reference building so this difference is also more neutral when the Asset Rating is expressed as a ratio.

Some elements of energy service can be re-introduced in that the standard conditions used in an Asset Rating depend on generic categories of energy service demand. Thus, the Asset Rating of an antiques store will be based on higher lighting energy use than for a clothing store, and the asset rating of a warehouse used for storing small objects will be based on higher lighting energy than for a bulk materials warehouse where the operators do not need to read fine print, and the Asset Rating for a nursing home may be based on higher winter temperatures than that of a university dormitory.

Asset Ratings are appropriate for a number of purposes:

• When one is evaluating the efficiency of a building that is being constructed or when one is considering purchasing or leasing, it makes most sense to evaluate the building and its comparables in the real estate market while making identical assumptions concerning energy service levels, operational conditions, climate, and maintenance. If different buildings are evaluated with different operating regimes or practices, then reliable comparisons of energy ratings are not possible, since there are too many uncontrolled variables.

• Asset Ratings give normatively "better" ratings (lower energy use) for more advanced technologies and designs, independent of variables that the owner or developer cannot control, such as the need for energy-intensive services such as Information Technology (IT) or hotel laundry services.

• Asset Ratings allow an apples-to-apples comparison of the efficiency of one building to another. It is very difficult to compare metered data to a baseline and also control for differences in energy service and/or operations and maintenance.

• Asset Ratings are essential in developing energy management programs and objectives for new buildings and major renovations, since they allow predictions of energy savings that will occur due to features that have not yet been installed. They also allow for quantitative comparisons of how far a building has gone compared to leading-edge practice (NBI 2012; DOE 2011) in adopting efficiency measures and design techniques, as seen next.

Most Asset Rating systems for both residential and commercial buildings are expressed as the ratio of the energy performance of the rated building to that of the baseline building. This ratio allows meaningful comparisons of buildings across sizes and occupancy types: a large office building that scores 20 % lower energy use than the baseline (an index of 80) can be considered more efficient than a small retail building that scores 10 % higher energy use than the baseline (an index of 110).

The calculation process also provides an absolute measure of energy use (typically measured in GJ), or of emissions associated with that energy use, or of standardized operating costs (energy use by type weighted by a standard schedule of cost by type and often by time of use). These absolute measures are more prone to error than ratios, but when ratings are used to compare different buildings, the comparison provides a ratio implicitly.

In the past, Asset Ratings have been relatively expensive to generate because the building simulation software requires the user to specify large amounts of data to describe the proposed or actual building and then to do this again to describe the reference building. This need not be the case in the future. Quality assurance programs such as Residential Energy Service Network (RESNET) or Commercial Energy Services Network (COMNET) both reduce input costs dramatically and

also add more confidence by requiring that the baseline building be automatically generated and that neutral modeling assumptions be uniformly applied. For example, RESNET ratings typically cost less than US$ 500 for a 200-m^2 single-family house, and most of the cost consists of on-site air leakage diagnostic tests. COMNET is a specification for nonresidential energy analysis software and users that use COMNET accredited software should be able to perform performance analyses in less than half the time it currently takes.

RESNET and COMNET are interesting models because of their emphasis on specifying the details of simulation software that is sufficiently accurate and input assumptions that are permitted or required to be used in generating the Asset Rating. These assumptions are intended to reproduce typical conditions of building energy service, controls functionality, and operational conditions and thus generate predictions that will be equal to average metered energy use to the extent possible. The systems' specifications are open to public review so that newly discovered discrepancies between energy consumption predicted using modeling results and measurements of energy consumption can be corrected, whether they are the result of the assumptions for operating conditions in the Asset Rating or inaccuracies in the simulation model algorithms or methods. Such an effort is valuable for providing the most meaningful information to the market on likely energy use and cost.

The assumption of on-site, post-construction inspection as a part of an Asset Rating is worth noting. Some of the problems with Asset Ratings have been a consequence of using as-designed parameters rather than as-built to calculate the ratings. RESNET distinguishes between the two by the terms "projected rating" and "confirmed rating" and by requiring that all projected ratings be accompanied by the disclosure on the first page of the report: "Projected Rating Based on Plans—Field Confirmation Required" (RESNET 2013). RESNET ratings also typically include estimates of the costs and energy savings (in both energy and monetary units) of a set of recommended efficiency upgrades.

Using these protocols, Asset Ratings are easier to generate because they are based on the same physical characteristics and diagnostics as would be required to demonstrate compliance with energy codes, and because the software standards (RESNET and COMNET) require that most of the inputs be applied automatically in the software as neutral independent or neutral dependent. Entries from users are mainly limited to the parameters that would appear on energy code compliance forms, such as U values and areas of envelope assemblies, rated efficiencies of heating and cooling equipment, power ratings of fans and lights, etc. Thus, the amount of time spent inputting data on the building's energy characteristics is minimized.

Simulation is one of the primary tools being used to implement the Energy Conservation Building Code (ECBC) in India (Bureau of Energy Efficiency 2013). The Bureau of Energy Efficiency (BEE) is preparing an on-line Asset Rating simulation model that automatically generates the reference building, similar to COMNET's requirements. The program has been beta-tested and BEE's consultants report that an engineer can input an entire building in about 1 h of billable time, if they are willing to accept some conservative assumptions used to simplify input.

Asset Ratings are used more often than Operational Ratings in Europe, with both types of ratings being ways to comply with the Energy Performance in Buildings Directive of the EU. Asset Ratings are used in twice as many member states as Operational Ratings, while three or four member states used both by 2010 (Maldonado 2011).

Asset Ratings can apply to separate building systems—envelope, heating, ventilation, and air conditioning (HVAC), and lighting. This procedure is used to document energy code compliance for California's nonresidential Title 24 energy standards for nonresidential buildings (California Energy Commission 2012). It is also used by Australia's Commercial Building Disclosure (CBD) program, where an Asset Rating of the lighting system is used for tenant assessments while an Operational Rating is used for the whole building (Australian Government 2013).

These system-specific ratings are not independent. One cannot calculate the rating for one system without making assumptions about the two others (California Energy Commission 2012). This is because lighting energy use affects HVAC use, and envelope energy use is reflected in changes in HVAC energy on the meter and in lighting energy to the extent that the envelope design inhibits or facilitates daylighting. Thus, while one could attempt to assign lighting energy to tenants while ascribing HVAC and envelope energy to landlords (unless the tenant provides the HVAC system, in which case only the envelope energy applies to the landlord), such an assignment is not determinate.

For example, the Australian choice to use a system more closely resembling Asset Ratings[4] for tenant systems seems necessary because of the difficulty of disaggregating tenant energy use (outside of lighting) from whole-building use, as well as its serving as a near-complement[5] for "base building" use (defined as use for common services). In addition, an Operational Rating for lighting would be hard to normalize for occupancy in a way that distinguishes better energy management from simpler energy service demands. Even with this advantage, the market uptake of tenant ratings has been lower than that of whole-building or base-building ratings (Bannister 2012a).

In summary, Asset Ratings are useful because they isolate one of the three dimensions of energy performance: the efficiency of the building's design and technology. Through modeling, they do an effective job of neutralizing for O&M and the level of energy service.

Analytic framework: operational ratings

Operational Ratings describe buildings in terms of adjusted metered energy performance per unit of conditioned floor area, with the baseline typically representing average energy use for a cohort of buildings of the same function (e.g., office). The most prominent example of an Operational Rating is the ENERGY STAR commercial buildings program (ENERGY STAR 2013). This program is based on comparing metered data for at least a full year of operation to baseline consumption data that is adjusted for the neutral dependent variables of occupancy, weather, conditioned floor area, and hours of operation. Operational Ratings work well at comparing the performance of a building from year to year, or for comparing the performance of one

portfolio of buildings against another. The Australian Building Disclosure system also makes use of Operational Ratings for its whole-building and base building systems; however it also uses an index that is effectively an Asset Rating for lighting systems in tenant spaces.

The technical basis of the Operational Rating is shown in Eq. 2. The energy performance of both the rated building and the baseline building is determined by looking at energy bills. The numerator represents the bills of the rated building over a minimum 12-month period of time, while the denominator represents the average energy bills of the baseline building normalized for climate and certain operating parameters. For ENERGY STAR, average energy bills are based on US "Commercial Buildings Energy Consumption Survey" (CBECS) (Energy Information Administration 2012) data, supplemented by other data when required. The metered energy performance of the baseline building is adjusted to match the climate and operating conditions of the rated building (neutral dependent). A similar methodology is used by the Australian CBD system.

$$Operational\,Rating = Operational\,EnPI = \frac{EP_{RB,EB}}{EP_{BB,AEB}}$$

(2)

where

$EP_{RB,EB}$ The energy performance of the rated building determined from the utility bills. Electricity, gas and other fuels measured at the meter would be converted to common units, such as source energy or cost.

$EP_{BB,AEB}$ The energy performance of the baseline building with the same conditioned floor space as the rated building, but adjusted for the operating conditions of the rated building. ENERGY STAR does this through a statistical analysis of CBECS data.

Comparing asset ratings and operational ratings

Asset Ratings and Operational Ratings are not intended to agree with each other for any particular building, and this is what is observed. Figure 1 compares the

---

[4] The resemblance is that the tenant rating is based on equipment power demand without consideration of operation. As noted in the text, most asset rating systems provide such a non-ratio variant of the energy prediction as a secondary output. However, the Australian rating uses a discreet system of star levels rather than a continuous ratio as defined in Eq. 1.

[5] Tenant energy use is not completely complementary to base building use, because the base building HVAC will be affected by the loads imposed by tenant lighting (as well as plug loads). The interdependency occurs in design as well as operation, because the capacity of the HVAC system—chillers and boilers and also fans and pumps—and in many cases its fundamental design choices, depend on the level and timing of tenant loads that are anticipated.

ENERGY STAR score (an Operational Rating) on the horizontal axis with percent savings calculations for LEED on the vertical axis (an Asset Rating). LEED is the US Green Building Council's "Leadership in Energy and Environmental Design" program, an internationally-recognized green building program (US Green Buildings Council 2012). If there were perfect agreement, the points would fall along an almost straight line (the horizontal axis is nonlinear, especially near 0 and 100, since the ENERGY STAR score is expressed in percentile units.) A higher score means better energy performance. The vertical axis is expressed in percent savings compared to the baseline code-compliant building; thus a higher score means a more efficient building.

While the averages of metered results (ENERGY STAR scores) are consistent with the simulations (LEED percent savings), there is wide variation from building to building as illustrated in Fig. 1 (Johnson 2003; Turner and Frankel 2008). The same pattern is observed for residential buildings: while the average of metered results also is very close to the average simulation for residential buildings in the USA, there is considerable variation from house to house.

These variations are to be expected: they embrace both differences in operations and maintenance and differences in the level of energy service demanded. Asset Ratings are intended to predict the energy use of a building assuming standard conditions, not actual conditions. Some of these variations are controlled (made neutral dependent) by the ENERGY STAR program, but others are not. Data centers and other energy-intensive activities along with variation in thermostat settings are some important reasons for the variation.

Other reasons for variation in the case of this figure include potential changes in the building itself between the design documents used for LEED compliance and the as-built structure.

Note that these large variations between projected energy consumption by a model and individually measured building energy use are found in to be at least as large when the model is statistically based, such as those used for ENERGY STAR and National Australian Built Environment Rating System (NABERS), as when they are engineering simulations. Figure 2 (Bloomfield and Bannister 2010), which is based on retail buildings in Australia, shows the same kind of scatter as Fig. 1; however, the range of departure from the model appears even larger—spanning a range of energy use of between 3:1 and 5:1 for the same predicted energy use.

Note that the vertical scales on the two figures are not the same: The Y-axis in Fig. 2 covers a much wider range of the variable. Note also that the choice of retail buildings in Fig. 2 is based on the fact that this figure is the only analysis of its kind. While retail buildings display somewhat more site-to-site variation than, for example, offices, the extent of variation is less than a factor of 2 larger (7.9:1 as compared to 4.5:1 for the case of New York buildings as reported in Hsu et al. 2012). We doubt that the difference would affect the conclusion we draw.

Many of the papers cited here seem to argue that adding more explanatory variables to the regressions, trying to maximize intercomparability of the buildings, such as stratifying hotels by their star ratings, would reduce this large variability. However, none of them has yet been able to control for the wide variability of

Fig. 1 As-Designed Savings versus ENERGY STAR Scores. Source: New Buildings Institute www.newbuildings.org Report to EPA on Building Performance versus Design Intent, July 2003 (Johnson 2003)
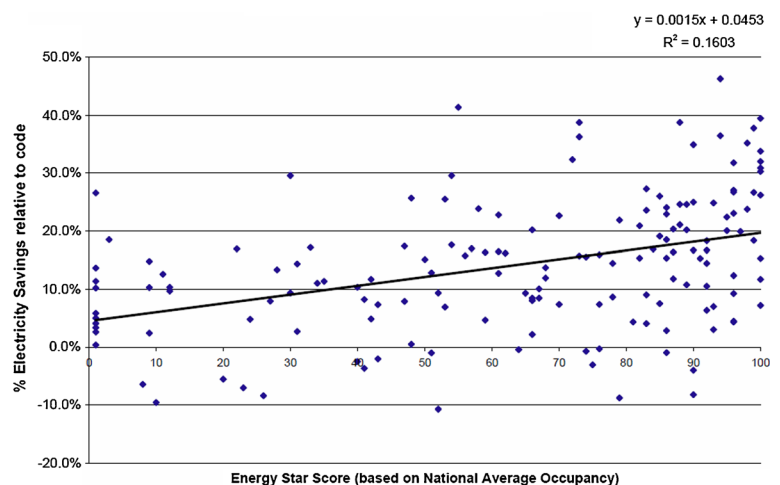
Fig. 2 Metered energy use as a function of statistically predicted energy use for retail buildings. Source: "Energy and Water Benchmarking in the Retail Sector: NABERS Shopping Centres." (Bloomfield and Bannister 2010)
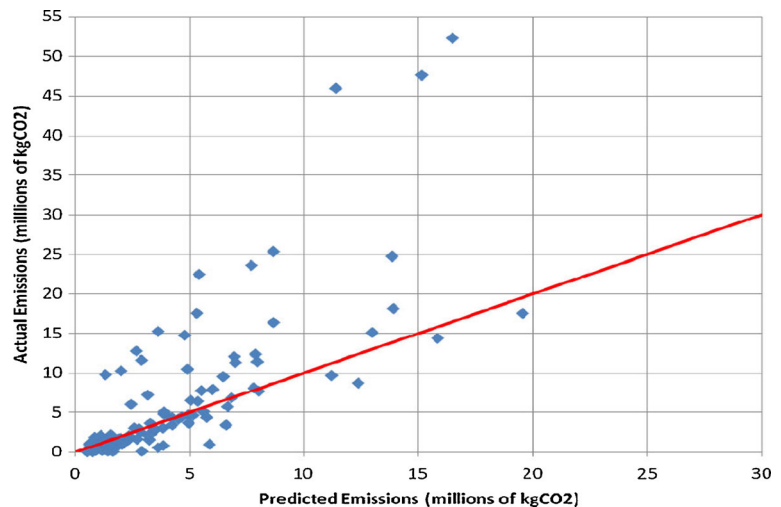
outcomes compared to the statistical predictions. We hypothesize that data on the O&M Indices of the buildings, and especially on the Energy Service Indices, may be helpful in explaining the variability. An organization's energy policy must focus on both the building assets and also on implementing effective operational practices if it aims to comprehensively realize energy saving opportunities. An energy plan with EnPIs for both capital assets (Asset Ratings) and operational effectiveness evidently will provide more useful information than a plan that relies on just one EnPI. However, we demonstrate next that Operational Ratings are not the best way to measure O&M effectiveness, since they combine the effects of variations in operation and maintenance with the effects of variations in energy service levels as well as the efficiency of the building itself.

In summary, Asset Ratings tell you how good the efficiency technology in a building is, but good technology does not assure low energy use: Asset Ratings address only one out of the three dimensions of a building's energy performance. Operational Ratings can tell you how your building's energy consumption compares to itself in previous years or to other similar buildings, but lower-energy-use buildings do not necessarily have better energy performance. This disconnect is a consequence of the impossibility of having a single-parameter rating that embraces all three dimensions of energy performance.

Operational Ratings can be a very useful tool to direct an organization's attention at energy management, as seen by the rapid uptake of ENERGY STAR buildings in the market, as well as by the success of the Australian CBD policy. They can be effective at the senior management level as an overview EnPI or to evaluate a portfolio of properties where differences in energy service demands, weather, etc., tend to average out. But they are not a useful tool for accomplishing better energy management at the operational level because they provide no guidance as to which of the three dimensions accounts for the observed energy use. They may nonetheless motivate short-term operational improvements, which are believed to be capable of saving 15 to 30 % of energy use.

There is a very wide variance between the energy use per square meter of different buildings of the same type in the same country or even the same city. In New York City, for example, the range of variation in energy use per unit floor area is 4.5 to 1 between the top and bottom 5 percentile for offices; for hotels the variation is 3.2 to 1, and for retail 7.9 to 1. (Hsu et al. 2012) While many analyses, most notably correlations with building age and size, explain some of this variation, considerable spread still remains after all the variables that the researchers are able to test and find statistically significant have been included in the model. The statistical models used in this report, which is noteworthy for the comprehensiveness of the data base available to it, explained only 20 % of the variation in energy use.

Figure 2 illustrates this type of variability. Evidently, it is several times larger than the amount believed to be possible to save based on O&M improvements, so it has to be dominantly based on other factors. It is also several times larger than the difference in efficiency; a study of the efficiency compared to energy code of new

buildings in California found very few that failed to meet code but even fewer whose Asset Rating was 40 % lower than code (Eley 2000). One would expect a relatively large range in a state that consistently supported above-code buildings through financial incentives. Again, the range of likely variation in efficiency is up to only 30 or 40 %. Note that this is not the range of efficiency that is possible to observe, but rather the range that is widely enough diffused in the marketplace to be visible in a statistical analysis.

We thus must hypothesize that most of the variation is in the level of energy service provision. Perhaps because we have not seen this hypothesis proposed in the literature, there is no solid evidence to corroborate or refute it. The Energy Service Index proposed later in this paper establishes a framework that allows this hypothesis to be tested.

This problem of unexplained variability is typical of all of the studies we have reviewed, as is discussed in the sidebar on Energy Use per Square Meter as an EnPI. Each study attempts to explain variations in energy use by statistical analysis but in every case substantial unexplained variance remains. All of the studies cited show large variability between the mean value of energy use per square meter predicted by the statistical model and the individual meter readings.

More importantly, while the New York City report offers some plausible hypotheses about how energy service demands including tenant comfort demands may account for much of the variation, the analysis is not yet sufficient to explain the physical causes of the variation in a way that would help improve energy management or test whether an energy management plan is on track to achieve its goals.

Another caution about the exclusive reliance on Operational Ratings can be seen through the policy concept of Net-Zero Energy Buildings, a concept that has been increasingly discussed among policymakers. Surveys have found over a hundred net-zero buildings around the world (NBI 2012; NBI 2013). But such surveys routinely overlook the fact that there are hundreds of millions of net-zero buildings that have been around for years: these are rural huts used for commercial as well as residential purposes that are unconditioned and have no access to electricity. An Operational Rating of zero purchased energy use looks very efficient until one considers the deficient level of energy service.

Operational Ratings should work most effectively in the context of an energy management system that looks at all four of the EnPIs discussed here, and may enable such analysis and its consequent action. A retrospective study of the National Australian Built Environment Rating System (NABERS) Australia's labeling and disclosure system (Bannister et al. 2012b) found that the ratings, which are primarily operational, seem to have encouraged greater use both of physical upgrades that result in better Asset Ratings and in greater training and motivation of staff, as well as owners and tenants, to operate the building better. As a goal of an energy management system, better operation can be measured analytically by the use of an explicit O&M Index, which we introduce next.

The O&M Index

We suggest that the problem of responding to user needs to understand both the physical efficiency of a building and the operational effectiveness can be addressed analytically by establishing a new EnPI that combines elements of Asset Ratings and Operational Ratings. The O&M Index is an index of how well the building is operated, given the assumption that the physical assets of the building are fixed but the operations and maintenance could be improved to better its Operational Rating.

It also corrects for differences in the Operational Rating that are based on differences in energy service or in user needs. Thus, a building with a tenant who demands more lighting for special visual tasks (consistent with higher power allowances in lighting standards or energy standards) can compare the energy bills to the energy consumption that should have been needed to provide that higher level of energy service. Other high energy intensity services, such as IT or food service onsite, or a tenant that requires the use of multiple large televisions or computer monitors, are also controlled for.

In the cases above, it is clear to most analysts that, for example, a sports bar that provides TV displays of all of the football games that are being played on a Sunday afternoon is providing a higher level of energy service than a bar that only offers two or three games. It is also clear that an office building that hosts large servers is providing more energy service that an otherwise identical building that outsources these IT needs. A building with a mobile phone tower on the roof evidently is

providing additional energy service. A refrigerated warehouse evidently will provide energy services not offered by a warehouse that is only conditioned to within a range of 10–35 °C. A food store selling upscale products such as ¥20,000-a-piece melons in Japan (~US$ 200) will require more lighting power than a discount food store, while a store with lots of frozen food and refrigerated food displays will consume more energy than one featuring tinned or packaged foods.

Likewise, a building where the tenants are satisfied with lower levels of energy services (for example 18 °C interior temperatures in winter) can have a base building modeled with the same temperature preferences as the actual building so as not to allow inflated estimates of energy use based on more typical thermal preferences in the reference case. Otherwise, the unusual tenant choices would appear spuriously as a sign of exemplary energy management.

A more ambiguous case is presented in the case of a building with a tenant who likes cooler temperatures than 21 °C during the summer. In this case, it can be argued whether this low thermostat setting is an indication of better performance in terms of greater user-perceived comfort or poor energy management behavior, and the user of the index would have to decide what parameters to use in the simulation of expected energy use.

The O&M Index is the ratio of metered energy performance of the rated building to the modeled performance of the rated building. However, the model is calibrated not by inputting the typical or average levels of energy service, which are prescribed by the Asset Rating, but rather with data or estimates for actual level of energy service (ASHRAE 2012), determined retrospectively. These conditions can be input at varying levels of precision depending on the cost of obtaining more accurate input data compared to the benefits of having a more precise simulation result. The technical definition is shown in the equation below:

$$O\&M \, Index = \frac{EP_{RB,EB}}{EP_{RB,ND}} \qquad (3)$$

where

$EP_{RB,EB}$ The energy performance of the rated building determined from the utility bills. Electricity, gas, and other fuels measured at

the meter would be converted to common units.

$EP_{RB,ND}$ The energy performance of the rated building, determined through modeling, but the actual operating conditions of the rated building are used, e.g., neutral dependent or "ND". See explanation below.

The building simulation in the denominator relies on neutral dependent building descriptors derived from the actual (and potentially varying) conditions in the rated building as opposed to being fixed and independent of the description of the rated building (neutral independent). In other words, the inputs to the simulation are calibrated to agree with how the building really is being used, to the extent of accuracy desired by the user.

The purpose of the O&M Index is to eliminate much of the modeling noise in comparisons of the Asset Rating with the Operational Rating (as shown in Figs. 1 and 2). Because the prediction is for custom conditions that may change from year to year, it is not anticipated that the O&M Index would be used externally to the building ownership, tenants, and managers. Nor would it be potentially disclosed, as the EU requires for its Asset Ratings, or as many jurisdictions require for Operational Ratings, or used for certifications, since it would be expected to change as tenants come and go, or even if they stay but their needs or staffing levels change. Instead, it would be used to evaluate the level of energy management success that the operations team and the tenants are able to achieve.

The O&M Index can be used in the process of an energy management plan to focus attention on operational practices that can be continuously improved every year, and to separate them from equipment or facilities retrofits that may only be performed every 5 or 10 years, and from the level of energy service required by the tenants.

Controls can represent a gray area between Asset Ratings and the O&M Index in theory, but in practice they are addressed in the referenced Asset Rating systems in a consistent and clear way to distinguish between the average performance of the building as constructed and the behavior of the occupants. They can be addressed using the Asset Rating's standard conditions as a starting point in the user-directed simulation that is in the denominator of the O&M Index.

The key issue is the presence of controls, and their commissioning. Evidently, if the lighting controls are at a centrally located circuit box that is not accessible to the

occupants, the amount of lighting energy used will greatly exceed that for a building in which dimming controls and vacancy sensors are present at every workstation.

The way that this distinction between asset and O&M is made in practice, both in Asset Rating standards such as COMNET and in energy codes, is to assign fixed credits for the presence of controls based on average (or conservative) assumptions about how much the occupants are likely to use the controls.

Thus, for example, if bi-level lighting controls are provided in hotel corridors, the COMNET Manual provides a credit for 20 % savings compared to the normal assumption for lighting power schedules. Since the simulation used in the denominator of the O&M Index definition of Eq. 3 is customized to the operation of the building in a particular year, these precise control credits need not be used in the simulation—the energy manager may choose different credits that more accurately represent the equipment used in the building or the baseline behaviors.

Any variation in the field from that assumption, in either direction, would be reflected in the O&M index. Similarly, any variation in the effectiveness of the controls compared to the assumptions required for the Asset Rating, or from alternatives used in deriving the O&M Index—a variation that often is a consequence of a failure to commission the controls—would also be reflected in the O&M Index.

A poor O&M index is not necessarily indicative of poor maintenance practices. In some cases, it is an indication that the modeled equipment was not installed, or was poorly installed, or that the system was not commissioned properly (or at all). In this case, it provides a quality check on the validity of the Asset Rating compared to the as-built building. Since Asset Ratings currently do not have strong methods for assuring high-quality commissioning, the O&M Index might be used as the basis to request re-commissioning as part of the energy management plan.

These are both valuable outcomes, as the use of the O&M Index encourages meaningful comparisons of expected energy use based on simulations to metered energy use. More of these comparisons will benefit not only the users at the individual building level but also the authors of standards for Asset Ratings, as they will be able to improve the required input assumptions based on substantially more data. This feedback will also be valuable to simulation software designers.

The O&M Index has two evident advantages: it is tied to the Asset Rating, since they are both based on the same energy simulation model, so it is easier to derive; and neutral dependent variables can be more finely tuned in the modeling process (the EPA ENERGY STAR and Australian regressions are significantly limited by available data).

The set of conditions—neutral dependent variables—that could be controlled for is far broader for a simulation-based analysis than it is for a statistically based analysis such as that which underlies ENERGY STAR. Also the magnitude of changes in energy use with changes in, for example, occupancy hours, can be projected more accurately using a simulation-based approach.

- One example is a multi-tenant office building with a law office with long hours on the 20th floor. This building will produce very different results if the 20th floor has its own separate cooling system than if the building has one chiller-based system for the whole building. In the former case, only the 20th floor system will operate long hours, while in the latter case, the operation of only one floor late at night requires that the whole-building HVAC system be operated.
- A second example is a religious building. One building might be used once a week for 2 h, while another might conduct religious services five times a day and offer educational services 12 h daily.
- Another example is a building with very high internal loads and low heat gains and losses, which will have very little responsiveness to weather variations compared to its peers. Such differences will be averaged and not accurately accounted for in a statistical model, as compared to a calibrated simulation model.
- Similarly, the presence of energy-intensive activities that are outsourced in some buildings but not others, such as printing, server use, workshops in educational buildings, food service, athletic facilities such as heated swimming pools, laundry and dishwashing, etc., can be handed through the O&M Index but are too site-specific to be handled by an empirically based statistical approach.

Note that if the building is operated in the exact manner specified for the asset rating, then the energy model should closely predict the energy bills. In this case, the energy performance of the rated building under neutral independent conditions will be the same as under neutral-dependent conditions. The numerator in Eq. 1 will be the same as the denominator in Eq. 3. But this is unlikely to occur, since real-world controls will not usually produce the exact temperature and air flow conditions as the modeled controls, and real settings may differ from the settings expected (and simulated) by the energy managers.

Just as the Operational Rating alone offers little insight into how efficient a building is, the simulated energy as modeled under actual, individual conditions, also may not offer much guidance on how efficient the building is. But by allowing the simulation to be calibrated to the field conditions in which the building operates, the O&M Index offers considerable management value as an apples-to-apples basis for comparison with metered energy use.

There is a wide range in the level of effort needed to derive the parameters necessary to develop the calibrated model needed for the O&M Index. Some users will want to make adjustments based on very cheap and simple methods, involving look-up tables or occupant surveys or even professional judgment, while others may want to meter key parameters to provide more exact inputs to the models. Even with very elaborate measurements of inputs, one can still expect some amount of noise in the comparison of a simulated result to a measured result. However, we anticipate, as discussed below, that the level of noise will be much, much lower than that displayed in Figs. 1 or 2.

Since the main purpose of the O&M Index is to assist the building's owners, tenants, and managers in measuring their success in energy management, rather than qualifying the building for any benefits, the energy team should make that decision on an individual building basis. The more effort that goes into developing the calibrated model, the less noise or uncertainty will remain in evaluating how well efficient operation of the building is being achieved.

If the O&M Index is 1.0, this is likely to be an indication of reasonably good energy management practices. A value greater than 1.0 suggests that the next step in the energy plan should be to improve operations, or to re-inspect or test the energy features of the building to assure that the efficiency measures were installed as specified or input into the model and that they work as they are supposed to. A value of less than one may be indicative of exemplary energy management—for example, controls that are turning off unneeded energy users at the local level as needed or requested by each worker. But it also may indicate inadequate provision of energy service compared to what the modeler expected.

When the O&M Index has a value other than 1.0, the difference might also be explained by an inaccurate model. Some factors that might explain the variation are structural errors in the simulation model, input errors in the model such as omitting energy uses such as elevators or occupant-supplied heaters, or conversely assuming uses of energy in the model that are absent from the building. Also, the simulation may not be able to model accurately an advanced feature such as displacement ventilation, natural ventilation, or daylighting. In this case, the modeler probably made an approximation that may be inaccurate.

How can we tell which of these options is most likely? Considerable information can be derived from calculating the O&M index on a monthly basis. This will reveal not only the average performance but the sensitivities. For example, if the index is at 1.3 all of the months of the year, then the problem is likely to be a source of load that was not expected, or a device with large standby loss. Perhaps the energy managers forgot about exterior lighting or parking lot lighting. If the index is 1.3 on average but higher in the hottest months and lower in the coldest, then there could be a source of internal load that is not accounted for. If the index is highest in the spring and fall, then it is possible that the controls allow simultaneous heating and cooling. If the index is close to 1.0 in the spring and fall and grows with heating and cooling degree days, one could suspect that insulation levels in the actual building are not as specified (and included in the model). If the index is high in the winter, low in the transition months, and high in the summer one might suspect that ventilation rates are too high; perhaps the economizer has failed in the open position.

Both for ease of modeling and for ease of interpretation, the calibrated model used for the O&M Index should be calculated assuming the controls work as

intended. Thus, controls failures show up as an O&M index greater than 1.0.

For ease of use, it would be efficient for the building modelers who are calculating the conventional Asset Rating to run several cases that simulate expected variations in operational parameters. This would reduce cost since it might not be necessary to re-engage the energy modeler later in the process. The modeler should work with the building owner to identify likely scenarios that could occur after building operation. For example, the modeling could look at the effect of climate variations and occupant density, and at the number of employees who are at work on an average day (as opposed to traveling or telecommuting), variations in hours of operation, both as a whole building and separately by floor or by wing or by suite, the impact of adding or subtracting internal loads, either concentrated in one location such as a server area or broadly distributed, and the impact of increasing or decreasing lighting power density based on different uses within a single occupancy.

Table 2 presents some examples of cases where differences in operation, or in a few cases building energy decisions during design, that are captured in the O&M index explain the difference between Asset Ratings and Operational Ratings.

Energy service index

A fourth index is the Energy Service Index. This is the ratio of the energy performance of the rated building as it is actually operated (neutral dependent) to the energy performance with standard conditions assumed for the Asset Rating (neutral independent; see Eq. 4). The

Table 2 Differences between asset ratings, operational ratings, and the O&M index

Variable	Asset rating	Operational rating (per ENERGY STAR)	O&M index
Occupant level	Fixed by occupancy type	Adjusted by density	Can vary by schedule as well as density
Ventilation rate	Fixed by occupancy type	No adjustments	Can increase if operated with high OA due to tenant needs; can respond to controls
Schedules of operation	Fixed	Adjustment by hours per week	Can be different by zone or tenant
Thermostat settings	Fixed	No adjustments	As required by tenants, perhaps with limits
Internal loads	Fixed	Adjustable within constraints	As input by user
Process energy	Fixed	Adjustable within constraints	As input by user
weather	Average weather for region	Adjusted by HDD and CDD	Weather can be input or adjusted by user
Drapery use (blinds, curtains, etc.)	Fixed	No adjustments	As input by user
Daylighting savings	As modeled using controls	No adjustments	As input by user
Lighting Power Density (LPD)	Actual in rated building, fixed by area category in rated building	No adjustments; implicitly actual	Actual
HVAC system	Actual in rated building; restricted in base building	No adjustments; implicitly actual	Actual
Fenestration area	Actual in rated building, fixed or constrained in baseline building	No adjustments; implicitly actual	Actual
Many different tenants/occupancy types	As anticipated at permitting	Not handled by the system	Actual for a given year, perhaps retrospective
Residential and commercial mix	As anticipated at permitting	Not capable of being rated by EnergyStar	Actual

authors see value in this index for the management of a specific building if the tenants change their energy service demands from year to year. It could be used to normalize the other EnPIs for the changes. It is also a clear indication of the difference in energy services for the rated building as compared to the standard conditions of the Asset Rating. Thus, it might be useful to a property owner whose tenants have special characteristics that cause the Energy Service Index of their other properties to differ from 1.0. It could be used to forecast the energy use of a new property based on its asset rating and the Energy Service Index of its other properties. Similarly, it might be used by a religious institution or an educational facility that is relocating to a property previously used within the same occupancy category (e.g., religious, educational) but by a user with very different demands.

When the index is less than one, the energy services are less than for the rated building, and when the index is greater than one, the energy services are greater. It represents a measure of how close the standard Asset Rating conditions are to those of the rated building.

$$Energy\ Services\ Index = \frac{EP_{RB,ND}}{EP_{RB,NI}} \qquad (4)$$

where

$EP_{RB,ND}$ The energy performance of the rated building, again determined through modeling, but the actual operating conditions of the rated building are used, e.g., neutral dependent or "ND". Note that this simulation has already been performed to compute the O&M Index.

$EP_{RB,NI}$ The energy performance of the rated building determined from an energy model. The "NI" subscript means that neutral independent modeling assumptions are used. Note that this simulation has already been performed to compute the Asset Rating.

We believe that variations in the Energy Service Index will explain much or most of the variation seen in Figs. 1 and 2. If this conjecture is correct, it can help guide future research on energy performance and allow greater predictability of energy consumption as measured.

Other energy performance indices

This paper focuses on whole-building or whole-tenancy or whole-system EnPIs that we believe often will be more useful to energy management than sole reliance on either Operational Ratings or Asset Ratings. There are also more detailed EnPIs that can be useful to personnel who can track them and use them as the basis of an Energy Management System.

For example, energy managers might want to track large process uses such as data centers directly, and could establish separate EnPIs for the IT use itself and for the HVAC functions associated with it; or they could establish EnPIs based on the effectiveness of daylight controls or the operational patterns of plug loads such as energy usage by time of day (Harris and Higgins 2012). This process is consistent with the general approach suggested in Goldstein and Almaguer 2011.

Accuracy of asset rating systems and of energy models

Asset Ratings, O&M Indices, and Energy Service Indices all rely on energy modeling. These EnPIs will only be useful for the purposes discussed above to the extent that the models correctly predict measured energy consumption. The accuracy is dependent on two types of factors: the appropriateness of the assumptions concerning operating conditions and the accuracy of the energy simulation model itself. We address the latter question first.

Since Asset Ratings and Energy Service Indices are expressed as ratios, the inaccuracies in both model and the assumptions about operating conditions are canceled out to the first order, since the same model, climate, etc., are used for both the numerator and denominator of the ratios. The O&M Index, however, compares measured energy use in the numerator to modeled energy use in the denominator and the impact of model accuracy is greater.

The literature comparing the predictions of simulation models to metered data is thin. While numerous

publications over the last 40 years have discussed the algorithms used in simulations, the hypothesis that the models correctly predict metered energy use has not been tested thoroughly. The few test results that have been published show good agreement even for an earlier generation of models (Fuehrlein et al. 2000; Schuetter et al. 2013; Goldstein 1978). The credibility and accuracy of rating systems could be improved if more systematic comparisons of simulated and metered data were performed, and one goal of this paper is to encourage such analyses, both for residential buildings and for different types of commercial buildings.

But in the absence of many carefully controlled, published studies it would be incorrect to assert that nothing is known about the accuracy of simulation models. First, practitioners who have modeled hundreds of buildings for which reliable metered date were available say publicly that a model based on observed operating conditions will usually be within 5 % of metered data on a monthly basis and even closer on an annual basis.[6] Second, the main algorithms in the models are based on well-understood laws of physics, or straightforward implementations of accepted engineering[7] (ASHRAE 2013). Third, given the widespread use of models for building code enforcement, voluntary programs, mandatory labels, and building design, it would be hard to believe that deep or fundamental errors could persist without being noticed and corrected. For example, errors in the models used to enforce California's Title 24 energy efficiency standards regularly are discovered by interests whose businesses are adversely affected by the error, or by state or NGO officials who are concerned that the level of energy savings may be

compromised, and the enforcement agency makes regular corrections and improvements.

Thus, the model itself should not often contribute materially to problems of disagreement between simulated energy use and metered energy use. And to the extent that the new research suggested here inspires further analyses comparing predicted energy use to metered use, we can expect that the accuracy of the models, and the consequent reliability of rating systems, will improve continually.

The agreement noted in Section 2 on Asset Ratings between the predictions of absolute energy use (the numerator of Eq. 1) employed in the two Asset Rating systems analyzed and the average of metered results is in large part a consequence of the amount of effort that the designers of the rating systems that have been evaluated have put into selecting the appropriate assumptions to use in the simulations, and which ones should be neutral or not and which should be dependent or independent (RESNET 2013 and COMNET 2012).

The choices made in these systems are intended to simulate typical occupant behavior and energy service demands, and typical performance of controls, such that the simulations are predictive of metered use on average. A poorer choice of input assumptions and requirements evidently would result in poorer agreement or in biased comparisons in which simulated results were consistently lower (or higher) than metered results, on average. Perhaps some jurisdictions have created rating systems that do not work as well as those evaluated here, but we could find no evidence concerning the existence or nonexistence of such problems.

The fact that the rated building energy use employed in Asset Ratings correctly predicts energy use on average is important from a policy perspective, because this fact implies that energy savings from building codes, which either rely directly on Asset Ratings or else rely on prescriptive measures whose resulting energy use can be calculated as an Asset Rating, save about as much energy as the calculations suggest. As noted earlier, Asset Ratings do not attempt to predict the energy use of a specific building in a specific year, because such a measurement re-introduces the effects of energy service demands and O&M behaviors. Usually, the purpose of an Asset Rating is to normalize for those effects.

Further evidence of the reliability of the RESNET system is the observation that one major US Home Warranty company is offering guarantees for up to 5 years that the billed energy use will not exceed the

[6] Personal communications, David Eijadi, Prasad Vaidya, and Lane Burt. These three practitioners described their own experiences in predicting metered energy use by adjusting the values of input variables to an energy simulation model based on building-specific measurements, (as opposed to just adjusting free parameters until the results aligned, which would not validate the simulation). There is some literature on calibration that is intended for the use of predicting (but not necessarily explaining) measured consumption—for example for the purpose of estimating energy savings—rather than analyzing whether defensible choices of input variables lead to a correct prediction of energy use.

[7] The issues of simulation are addressed in ASHRAE 2013 Chapter 19, while issues of algorithms are addressed in Chapter 4, 17, and 18. They have been reported in previous editions of this document dating back at least to the 1970s. However, as noted in the body of this paper, the ASHRAE Handbook does not provide estimates of accuracy of their methods.

rated use by more than 15 %; the guarantee is to pay the home buyer the overage. In Texas, one company has offered such guarantees since 2008. To date, they have issued 4,396 certificates and have had 13 claims.[8] Also, the RESNET system has been demonstrated to predict home mortgage loan performance: lower scores are strongly correlated to lower loan defaults (Quercia et. al. 2013).

The observation that some rating systems can predict average energy use accurately implies that it is worth evaluating a particular rating system to see the extent to which it is indeed accurate. Evidently, any errors observed during this evaluation could be corrected after they are demonstrated. Thus, the observations made here are intended to encourage this sort of research. The contrary observation—namely that rating system predictions always fail to correlate with measurements—would suggest that perhaps they are incapable of doing so and efforts to calibrate them should be abandoned.

Conclusions

We have introduced a taxonomy into which we place four types of energy performance indicators that we believe will be the most useful in making markets work to design, build, and operate energy efficient buildings.

Different EnPIs are most applicable for different uses. For some uses, only one indicator will suffice. This one EnPI will likely be the Operational Rating. Operational Ratings include all dimensions of building energy performance, and are by a large margin the least expensive to implement, and therefore this rating is the EnPI that will continue to see the fastest increase in market uptake in the short term, which is appropriate.

For other uses, two or all three of the proposed indicators for specific dimensions of energy performance, along with the Operational Rating, are best. We argue that most typically, two ratios will be most useful at the building engineering/management level: the ratio implicit in the Asset Rating and the ratio embodied in the O&M Index. For example, in trying to identify technology and design leadership, the first indicator, energy use compared to a reference, is most meaningful. However, even in this case, the O&M Index is also important, as a building designed to save a lot of energy that winds up not really performing better than average will be an embarrassment.

As another example, for a property owner that is trying to show continual improvement in performance of a fixed set of properties operated in ways that are consistent over time, the Operational Rating will be most useful. Asset ratings will only provide added value if the owner plans a capital upgrade as part of its energy plans, or if the usage of the property changes.

As a third example, if the goal is to appraise a building and make lending or pricing decisions based on energy use, then the Asset Rating is most appropriate, as it best describes the likely energy costs of the building when it is operated by a new owner with tenants who may not be fully known at the time of the sales or financing transaction. The asset rating can be combined with the O&M Index and the Energy Service Index of the new owner's other property or properties, if appropriate, to estimate actual energy use and costs ex ante. But if the purpose is to estimate Net Operating Income for the first year after the purchase of an existing property, the Operational Rating may be most appropriate, because the tenancies may not turn over during the first year, and operating conditions may not change much.

Asset Ratings, Operational Ratings, and the O&M and Energy Service Indices are data points that can be used to support various theories of value or efficiency. None of them is "the truth" in any absolute sense: each in its own way is subject to measurement error, data transcription error, noise in the operating conditions, and inappropriate interpretations. These sources of error are minimized, though, when the analyst can compare all EnPIs.

Each of these four indicators can be conceived of as an answer to a question.

- The Operational Rating answers the question: "how energy intensive is this building compared to its peers?"
- The Asset Rating answers the question: "how efficient is this building?"
- The O&M Index answers the question: "how well is this building being managed?"
- The Energy Service Index answers the question, "How demanding are the tenant energy use requirements as compared to standard conditions?"

This process of selecting the appropriate EnPIs for the job can support a successful energy management system by providing useful feedback to managers, tenants, and owners about the actions that can be taken at the appropriate level to improve energy performance.

---

[8] Personal communication, email, Brett Dillon, 25 October 2013.

Fixed Energy Budgets, or using Energy Use per Square Meter as an EnPI

The idea of ranking buildings in terms of energy use per unit of conditioned space is a simple and intuitively appealing one, and thus it has been advocated in many different locales around the world for the past four decades. But buildings are not that simple, and attempts to implement energy budgets per square meter have led to unattractive side effects that have resulted in many jurisdictions that first tried them ending up abandoning them after feedback from the field. While the reasons for abandoning fixed energy budgets are not well documented in the literature, the actions of California, ASHRAE, and other code development organizations to move in other directions speak volumes.

Beginning with the Building Energy Performance Standards (BEPS) program that the U.S. Department of Energy (DOE) developed in the 1970s but was unable ever to implement due to political opposition, to California's first generation performance-based standards for commercial buildings in the 1980s, to standards in Shanghai in the first decade of the 2000s, to the Russian Federation's performance standards in the 2000 era, the idea of rating energy performance in kWh/m^2 has been a straightforward approach that is seemingly intellectually irresistible. In all of these cases, participants in the standards implementation process observed when the standards were implemented, at least orally, that the standards were unfairly easy to meet for some choices of building size and site constraints, while they were overly burdensome for other circumstances. The concept of fairness was introduced by participants in these processes.

To solve the perceived inequities of fixed energy budgets, California adopted what was termed the custom budget approach for residences in the mid-1980s and this fundamental approach has survived and has been expanded to commercial buildings, not just in California, but in ASHRAE Standard 90.1 as well. The approach is also echoed in the International Energy Conservation Code and in the specifications for programs such as ENERGY STAR new homes and the U.S. Internal Revenue Code Section 45L tax credit for new home efficiency and the Section 179 tax deduction for efficient commercial buildings enacted by the U.S. Congress in 2005. The idea is that the energy budget is customized for each building and is equal to the energy used by a reference building that has the same geometry and other properties as the proposed design or rated building, but is upgraded or downgraded to be in exact compliance with the prescriptive standards. In other words, the energy budget is derived from the prescriptive standards. This approach requires that two models be developed, one for the budget building and one for the proposed design, but the budget building model can be automatically generated by software, as demonstrated in California and as required by COMNET and RESNET software specifications. This approach is also followed in most Asset Rating standards.

The custom budget approach has been so easy to use and so successful that practically all new residences in California use this method, as well as an overwhelming majority of nonresidential buildings. The approach was incorporated into ASHRAE Standard 90.1 in 1989 and expanded in 2004 to address beyond-code programs such as LEED (the U.S. Green Building Council's green building rating program).

The comparison of a proposed design to a reference building reduces modeling errors and errors in model inputs. Since both buildings are required to be simulated under the same conditions with the same simulation engine and weather files, most errors cancel out to first order. Only second-order errors affect code compliance or the efficiency rating. This is one reason why the indices we recommend in this paper are all ratios.

The custom budget approach also has greater industry acceptance in part because it is more transparent. A review of the underlying prescriptive standards shows what is required for compliance. Interestingly, some of the biggest supporters of the custom budget concept in California were the homebuilder associations, which at the time vehemently objected to energy code stringency overall but were more willing to accept stringency when they could understand the process and see how to make trade-offs, e.g. optimize their design to achieve compliance at the lowest cost.

A fundamental problem is that fixed energy performance targets are not economically optimal (Marshall and Peterson 1979). This issue is compounded by the fact that building performance is not simple, and that a number of different variables have an important influence. These variables are not completely independent, and even after controlling for many of them, the variation in energy use per unit floor area is still large (Schipper 1983; Xiao et al. 2012; Wang 2012; Monts and Blissett 1982; Bloomfield and Bannister 2010; Xin et al 2012; Chung 2011; Yamaguchi et al. 2012; Kontokosta 2012; Panayiotou et. al. 2011). All of these reports identify large unexplained variability in the energy use per square meter of building, even after controlling for as many factors as the authors could identify. And there are second-order effects as well: several of the studies find that energy use per square meter is itself a function of the number of square meters of conditioned area.

The hypothesis that energy consumption can be usefully normalized by floor area is not presented in any of these papers as a falsifiable hypothesis that can be tested against data. Instead, the theory papers critique the choice of normalization by floor area on an abstract level while the empirical papers just assume that the hypothesis that normalization by conditioned area is correct and proceed directly to the analysis of energy intensities that are defined this way.

The failure to frame and test such a hypothesis is an important weakness in the current literature. Most of the studies seem to brush aside the unexplained variation as if the authors expected that a future more detailed specification of the regression equations with more variables would reduce the unexplained variation in energy use per square meter, and did not note the consistency between all of the other studies in finding such large variations in different countries and different building types.

Setting a fixed budget thus implies that buildings whose variation from the mean is in the low energy use direction will find the target easy to meet, and those whose variation is in the opposite direction will find the target expensive and difficult. This is the reason for the departure from economic optimality.

Fixed energy budgets are like rations; the same amount of energy is allocated to a particular class of building in a particular climate, based solely on the size of the conditioned area. Rations are one of the methods of resource allocation used in command and control economies. Custom energy budgets or ratios based on such baselines are more market oriented, and markets avoid the economic inefficiencies of rationing. Each designer or builder can find the set of solutions that works best for them or minimizes their overall cost. With rationing, cases will arise where the cost of exceeding the ration (using less energy than the ration) for one building may be a lot lower than the cost of failing to meet it in another building by the same amount of energy. A more optimal outcome would be achieved if the rations were set to equalize the marginal costs of additional energy savings. This observation elaborates on the more detailed mechanism by which fixed targets are not economically optimal.

Incorporating market principles is the reason that energy codes are as long and complicated as they are. It would be simple to require that all walls have the same composite U-value regardless of construction type, but the market does not like this. It prefers, and codes reflect this preference through the consensus-seeking process that both government-adopted and private-sector codes go through, to differentiate U-value requirements between windows and solid walls, and between types of solid walls, such as mass walls or metal framed buildings. Similarly, it would be easy to require that a cooling system meet a level of COP that is the same for everyone, but the stakeholders prefer the type of standards set in ASHRAE 90.1, California Title 24, and other codes that differentiate based on size, air cooled vs. water cooled condensers, and other factors. Lighting codes have requirements for more than 100 distinct building and space uses.

These differentiations add complexity, and the complexity is not factored into fixed energy budgets or the statistical analyses that may underlie them. But the code development forums nearly everywhere in the world come (somewhat) independently to the same conclusion: that such differentiation is an important benefit in the market and makes the code fairer and more acceptable.

We can also see this issue beginning to emerge in the context of Net-Zero energy buildings. Zero is just a special case of a fixed energy budget. What some research is beginning to show is that net zero is much easier to achieve e in some building types than in others.

This is not to say that energy use ratings per unit of floor area have no purpose at all. To the extent that market failures are overcome and energy use is a factor in determining property values, a property with high energy use intensity will have higher operating costs and thus lower value. It does not and should not matter to the market that there are good reasons for this high use, such as the fact that the property houses IT-intensive tenants or retail operations requiring high levels of lighting or product display energy: the market can trade off these obvious benefits of the space with the higher energy cost rating.

Annex—comparison of common asset and operational ratings

All Asset Ratings as well as Operational Ratings use both neutral independent and dependent variables. See the following table as an example. Building descriptors (or broad classifications of building descriptors) are listed as rows. Various types of ratings are listed as columns. The first column has code compliance options, which are essentially binary ratings. The second column is the ASHRAE bEQ asset rating method. The third column is the ASHRAE 90.1 Performance Rating Method (PRM), which is used for US tax deduction for energy efficient buildings in Section 179 of the US Internal Revenue Code (IRS179) and LEED. The last two columns are the ENERGY STAR programs (Table 3).

Table 3 Comparison of common asset and operational ratings

Classification of building descriptor	Title 24 and 90.1 ECB	90.1 PRM (Appendix G)	ENERGY STAR Target Finder	ENERGY STAR Portfolio Manager
Approach	Rated building and comparator energy performance are both determined through modeling		Simulations are used for rated building, but these results are compared to empirical data	Rated building and comparator are both determined empirically
Envelope geometry, building mass and siting	ND—the area and configuration of the baseline building and rated building are identical	ND—The area and configuration of the baseline building and rated building are the same, but the baseline building is rotated and an average is used as the comparator	A—With this program, simulation results are compared to adjusted energy bills	A—With this program, statistical simulation results are compared to adjusted energy bills
Fenestration Area	ND—when WWR is less than 40 % A—when WWR is more than 40 %	A—A window wall ratio is fixed in the baseline building by building type A—when WWR is more than 40 %	A—the rated building is what it is but the comparator is average area	
Fenestration Orientation	ND—baseline building has the same distribution and orientation as the rated building	A—The baseline building has the same fenestration orientation as the rated building, but it is modeled in four rotations and the average is used as the comparator	A—the rated building is what it is, but the comparator is the CBECS average	
Insulation and Infiltration	A—baseline building is set by standard	A—baseline building is set by standard	A—the rated building is what it is, but the comparator is the CBECS average	
Heating source	NI—for T24 ND—for 90.1 ECB	ND—baseline building is the same as the rated building	A—the rated building is what it is, but the comparator is the CBECS average	
HVAC System Type	NI—for T24, baseline building system is different, but depends on the rated building to some extent	A—System type and fuel type is fixed for the baseline building, depending on building type, size and climate	A—the rated building is what it is, but the comparator is the CBECS average	
Lighting Power	A—baseline building is defined by the standard	A—baseline building is defined by the standard	A—the rated building is what it is, but the comparator is the CBECS average	
Schedules of operation	NI—for T24 ND—for 90.1 ECB	ND—the modeler can use the schedules they want, but the same schedules are used for the baseline building	ND—the EPA target is adjusted for simple things like operating hours. The rated building is what it is	ND—the target is adjusted somewhat to match the conditions of the rated building

Table 3 (continued)

Classification of building descriptor	Title 24 and 90.1 ECB	90.1 PRM (Appendix G)	ENERGY STAR Target Finder	ENERGY STAR Portfolio Manager
Plug loads	NI—for T24 ND—for 90.1 ECB	ND—rated building values are used for baseline building	ND—the EPA target is adjusted for number of computers, occupants, etc. to match the assumptions in the model	ND—the EPA target is adjusted for the conditions of the rated building
Climate	NI—for T24 ND—for ECB	ND—can use any climate you want, but it must be the same for the baseline building	ND—crude adjustments are made to the EPA target based on heating and cooling degree days	

A asset, *NI* neutral independent, *ND* neutral dependent, *N* not accounted for

References

ASHRAE. (2012). American Society of Heating, Refrigerating, and Air-Conditioning Engineers, ASHRAE Guideline 14-2002: Measurement of Energy and Demand Savings, Atlanta, GA

ASHRAE. (2013). *American Society of Heating, Refrigerating, and Air-Conditioning Engineers*. Atlanta, GA: ASHRAE Handbook of Fundamentals.

Australian Government. (2013). Commercial building disclosure. http://cbd.gov.au/overview-of-the-program. . Accessed 12 September 201

Bannister, P. (2012a). *NABERS—lessons from 12 years of performance based ratings in Australia*. Manchester UK: International Conference for Enhancing Building Operation. October 2012.

Bannister, P. et al. (2012b). What really makes buildings efficient: results from the low energy high rise project.

Bloomfield, C., & Bannister, P. (2010). Energy and Water Benchmarking in the Retail Sector: NABERS Shopping Centres, Proceedings of the 2010 ACEEE Summer Study on Energy Efficiency in Buildings, Washington, D.C., American Council for an Energy-Efficient Economy.

Bureau of Energy Efficiency, India. (2013). ECOnirman Whole Building Performance—ECBC Conformance Check. http://eco3.org/wp-content/plugins/downloads-manager/upload/ECOnirman%20Two-pager.pdf. Accessed 12 September 2013.

California Energy Commission. (2012). 2013 Nonresidential Alternative Calculation Method (ACM) Approval Manual. Revised. http://energy.ca.gov/2012publications/CEC-400-2012-006/CEC-400-2012-006-CMF-REV.pdf. Accessed 4 November 2013.

DOE. (2011). Department of Energy. Builders Challenge. http://www1.eere.energy.gov/buildings/challenge/builders.html. Accessed 12 September 2013

Chung W. (2011). Review of building energy-use performance benchmarking methodologies (Review), Applied Energy 88(5) (2011) 1470–1479.

COMNET. (2012). Commercial Energy Services Network. Modeling Guidelines and Procedures, http://www.comnet.org/mgp. Accessed 12 September 2013.

Crowe E., & Falletta K. (2012). California's Commercial Building Energy Asset Rating System (BEARS): Technical approach and design considerations, Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, Washington, D.C., American Council for an Energy-Efficient Economy.

Energy Information Administration. (2012). Commercial Buildings Energy Consumption Survey (CBECS). http://www.eia.gov/consumption/commercial/. Accessed 12 September 2013.

ENERGY STAR. (2013). http://www.energystar.gov/buildings/facility-owners-and-managers/existing-buildings/use-portfolio-manager/new-energy-star-portfolio-manager. Accessed 13 September 2013.

Eley, C. (2000). Tax credit study: an analysis of ASHRAE/IES 90.1–1999 and California Title 24–1998. Prepared for Pacific Gas and Electric Company. San Francisco, California, USA, 5 June 2000.

Fuehrlein, B., Chandra, S., Beal,D., Parker, D., and Vieira,R. "Evaluation of EnergyGauge® USA, A Residential Energy Design Software, Against Monitored Data", Proceedings of ACEEE 2000 Summer Study, Washington, DC, American Council for an Energy Efficient Economy, August 2000.http://www.fsec.ucf.edu/en/publications/html/FSEC-CR-1670-00/index.htm. Accessed 24 October 2013.

ISO. (2011). ISO International Standard 50001: 2011, Energy Management System. http://www.iso.org/iso/home/standards/management-standards/iso50001.htm. Accessed 12 September 2013.

Goldstein, D., McKane A., & Desai, D. (2011). ISO 50001: Energy management systems: a driver for continual

improvement of energy performance, Proceedings of the 2011 ACEEE Summer Study on Energy Efficiency in Industry, Washington, D.C., American Council for an Energy-Efficient Economy.

Goldstein, D., & Almaguer, J. A., (2013). Developing a Suite of Energy Performance Indicators (EnPIs) to Optimize Outcomes, Proceedings of the 2013 ACEEE Summer Study on Energy Efficiency in Industry, Washington, D.C., American Council for an Energy-Efficient Economy.

Goldstein, D., (1978). Some analytic models of passive solar building performance, Berkeley, California, Lawrence Berkeley Laboratory, LBL-7811, November 1978.

Graden, L., et al. (2008). Minnesota B3 Benchmarking Results: prioritizing the energy savings opportunity in Minnesota public buildings, Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings, Washington, D.C., American Council for an Energy-Efficient Economy.

Harris, D., and Higgins, C. (2012). Key performance indicators, Vancouver, WA, New Buildings Institute, Report to the California Energy Commission Public Interest Energy Research (PIER) Program.

Hassel, S., Blasnik, M., & Hannas, B., (2009). Houston Home Energy Efficiency Study. December 2009. Also by same authors, a summary published in *Home Energy*, January/February 2011.

Hsu, D., Kontokosta, C., Kerr, L. Beber, H. & Hope, D., (2012). PlaNYC New York City Local Law 84 Benchmarking Report 2012.

Johnson J. (2003). Is What They Want What They Get? Examining Field Evidence for Links between Design Intent and As-Built Performance of Commercial Buildings, Vancouver, WA, New Buildings Institute 2003. http://www.newbuildings.org/sites/default/files/IsWhatTheyWantWhatTheyGet200307.pdf. Accessed 12 September 2013.

Kontokosta, C.E. (2012). Predicting Building Energy Efficiency Using New York City Benchmarking Data, Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, Washington, D.C., American Council for an Energy-Efficient Economy.

Maldonado E. (2011). Implementing the energy performance in buildings directive, Brussels, Belgium, The European Union, EA-30–11–026-EN-C, 2011.

Marshall, H. E., & Petersen, S. R. (1979). Economics and the selection and development of energy standards for buildings. *Energy and Buildings, 2*(2), 89–99.

Monts, J. K., & Blissett, M. (1982). Assessing energy efficiency and energy conservation potential among commercial buildings: A statistical approach. *Energy, 7*(10), 861–869.

National Academy of Sciences (2010). *Real prospects for energy efficiency in the United States*, Washington, D.C.

NBI. (2012). New Buildings Institute. Advanced buildings. Vancouver, WA, http://www.newbuildings.org/advanced-design/getting-50-beyond. Accessed 12 September 2013.

NBI. (2013). New Buildings Institute. Net Zero Energy, a set of resources accessible through http://newbuildings.org/zero-energy. Accessed 12 September 2013.

Panayiotou, G. P., et al. (2011). The characteristics and the energy behaviour of the residential building stock of Cyprus in view of Directive 2002/91/EC. *Energy and Buildings, 42*, 2083–2089.

Roberto Quercia, R., et al., Home Energy Efficiency and Mortgage Risks, University of North Carolina at Chapel Hill, NC, March 2013.

RESNET. (2012). Residential energy service network. What is a home energy rating? http://www.resnet.us/home-energy-ratings. Accessed 12 September 2013.

RESNET. (2013). Residential Energy Service Network, BSR/RESNET 301-2013 Standard for the Calculation and Labeling of the Energy Performance of Low-Rise Residential Buildings using the HERS Index, Final Draft Standard v.2, 22 August 2013. http://www.resnet.us/standards/BSR-RESNET_301-2013_FDS-v2_08-22-2013.pdf. Accessed 11 September 2013.

Schipper, L. (1983). Energy efficiency and conservation in buildings: the use of indicators, *Energy* 8(1).

Schuetter, S., Hackel, S., Gunasingh, S., Searching for savings opportunities in commercial new construction programs, Madison, WI, Energy Center of Wisconsin, ECW Report Number 267-1, July 2013.

Turner, C., & Frankel, M. (2008). *Energy performance of LEED for new construction buildings*. Vancouver, WA: New Buildings Institute.

U.S. Green Buildings Council. (2012) https://new.usgbc.org/leed. Accessed 12 September 2013.

Wang, J. C. (2012). A study on the energy performance of hotel buildings in Taiwan (Review). *Energy and Buildings, 49*(2012), 268–275.

Xiao, H., Wei, Q., & Jiang, Y. (2012). The reality and statistical distribution of energy consumption in office buildings in China. *Energy and Buildings, 50*, 259–265.

Xin, Y., Lu, S., Zhu, N., & Wu, W. (2012). Energy consumption quota of four and five star luxury hotel buildings in Hainan province. *China, Energy and Buildings, 45*, 250–256.

Yamaguchi, Y., et al. (2012). Classification of Japanese retail facilities to establish benchmark energy consumption, Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, Washington, D.C., American Council for an Energy-Efficient Economy.