Metrics for Energy Efficient Buildings: How Do We Measure Efficiency?

Philip Fairey, Florida Solar Energy Center
David B. Goldstein, Natural Resources Defense Council

ABSTRACT

Several mutually incommensurate metrics have been used to rate building energy efficiency. Metrics are constructed with two different goals: a broad goal of comparing different buildings with respect to their efficiency and the narrow goals of comparing all-electric buildings with those using two or more fuels. This paper discusses the comparative usefulness of broad measures of energy efficiency such as energy use per unit of conditioned floor space (Energy Use Index or EUI), the HERS Index and its commercial analogue the zEPI Index, and percent-better-than-a reference-code. This comparison is performed in a policy context of how efficiency in buildings is defined. It then looks at four possible ways of comparing fuels: normalized modified loads (used in the HERS Index and in the International Energy Conservation Code), site energy, source energy, emissions-weighted energy, and cost-weighted energy.

It finds that the simpler methods of EUI and site energy provide the least useful information. The other methods offer useful answers to different questions, and therefore the user may need more than one of them in order to make justified decisions on energy efficiency.

Introduction: Efficiency and Consumption Metrics and Fuel Conversion Metrics

Energy management is facilitated by the use of quantitative indicators of energy consumption as a function of energy service needs: "You can't manage what you don't measure." This observation is embodied in structured energy management systems standards such as ISO 50001 (International Organization for Standardization, 2011). ISO 50001 requires organizations to define one or more energy performance indicators (EnPIs), set a goal for continual improvement of the EnPIs, and track progress towards meeting the goal over the years.

Experience with energy management system standards has shown that many companies prefer to use only one energy performance indicator (EnPI), and a very simple one at that, even though a suite of EnPIs would make more sense and allow better energy management. (Goldstein and Almaguer 2013)

EnPIs can be used to benchmark the performance of one facility compared to others, as well as tracking the performance of the same facility over time. Thus, if the goal is to improve the energy performance of a building, one needs one or more metrics for energy performance. In addition, if the building uses more than one fuel source, as most buildings do, and one wants to make comparisons, we also need a metric to compare different fuels.

This paper discusses both metrics for benchmarking and metrics for fuel conversion. It points out the shortcomings of relying on a single metric, both in terms of ranking overall energy performance and in terms of comparing fuels.

Types and Purposes of Efficiency Metrics

Overview

At the outset, it is important to distinguish between efficiency metrics and energy performance metrics. 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), and this paper reviews metrics for energy efficiency. The U.S. National Academy study discusses several alternate definitions of efficiency. 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 behaviors that allow the user to be satisfied with lower levels of energy service, and from improvements in O&M procedures.

This observation is consistent with the definition of efficiency for a component of a building, such as COP of a heat pump, which is tested under consistent conditions of indoor and outdoor temperature, times at different levels of load, etc., or of a lamp, which is tested at consistent levels of ballast performance, temperature, electrical inputs, lumen output, etc.

This definition is also consistent with common practice in the building energy efficiency community in the United States. However, as noted in the National Academy study, alternate definitions of efficiency include energy use divided by a measure of output, so that metrics such as energy use per square meter of conditioned floor area can be defined as "efficiency" by some analysts. As will be shown, such a definition is inconsistent with the more nuanced definition suggested here.

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 (Goldstein and Eley 2014). 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.

Efficiency as defined above refers to the performance of the engineered system of a building, holding all other factors constant.

EUI and the definition of efficiency

A very commonly used metric for building energy performance is the energy use index (EUI), defined as the energy consumption per unit conditioned floor area. In the U.S., this metric is expressed in Btu per square foot, and it is the metric used by the widely accepted Energy Star Portfolio Manager.

Experience with Energy Star has found that this one-parameter system is appealing to the affected industries. Benchmarking using this index helps building owners and managers to decrease energy consumption. This appears to be due to two factors: drawing attention to energy performance as an opportunity for improvement, and inducing market competition over lower EUIs, especially when the metrics are disclosed.

It is important to note that for systems such as Energy Star and Australian Commercial Buildings Disclosure program, the metric is used as an operational rating: a metric based on metered data normalized for relevant variables such as weather, occupancy, etc. This paper will also discuss metrics that are asset ratings, and therefore based on simulated energy consumption under fixed operating assumptions.

EUIs are often used as metrics of efficiency, but such an interpretation is incorrect if one accepts the definition of efficiency provided here. 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.

(Note that if one accepts the alternate definition of efficiency, then EUIs are a metric of efficiency but the relationship is tautological.)

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 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 by some dozen authors working in different countries and for different building types explained at most only 20% of the variation in energy use.

Figure 1 illustrates this type of variability. Evidently, the variation is several times larger than the amount believed to be possible to save based on O&M improvements. 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). The range of likely variation in efficiency is up to only 30 or 40 percent.

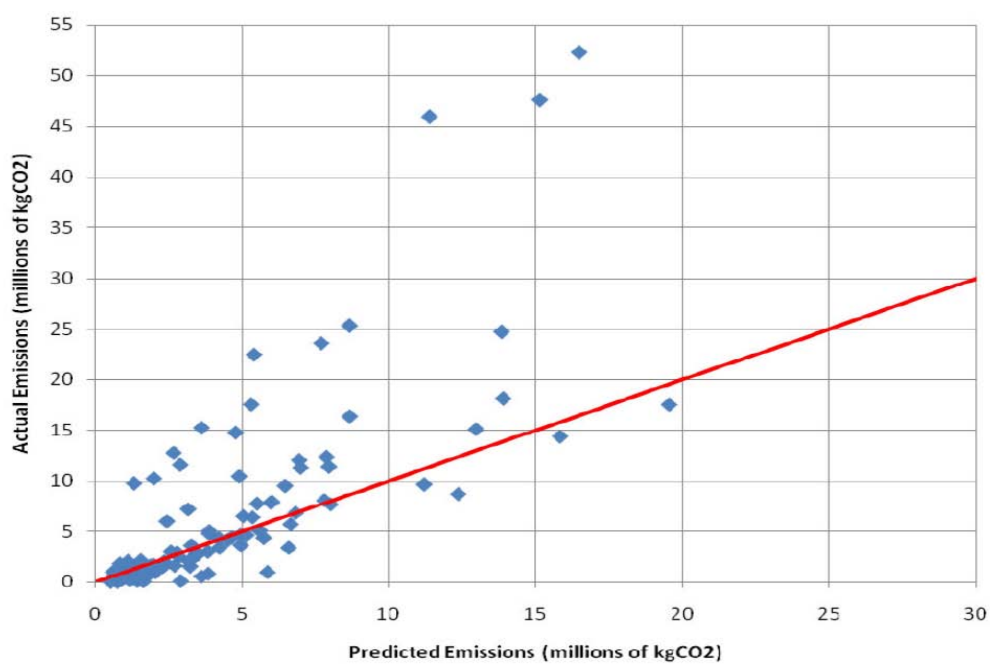

Figure 1. Metered energy use as a function of statistically-predicted energy use for retail buildings. *Source*: (Bloomfield and Bannister 2010)

(The scale of this figure is in CO_2 emissions rather than energy use, but this doesn't affect the conclusions concerning high variation: as noted below, one metric for evaluating building efficiency is essentially emissions. We also note that for current electric grids, emissions tend to be proportional to source energy use or cost-weighted energy use.)

The authors thus must hypothesize that most of the variation is in the level of energy service provision, as the other sources of variation are insufficient to explain the variability seen in Figure 1. There are a wide range of reasons why there is so much variation in EUI that is unrelated to efficiency. Reasons associated with variation in energy service level include:

- Operating hours for some or all parts of the building may vary
- Occupancy may vary due to employee travel or working offsite
- Weather may be different in measured years
- Energy-intensive processes like food preparation, data processing, washing, etc., may occur onsite or offsite
- Reasons for variation in EUI due to O&M differences include:
- Controls and equipment may not work properly, or may be set improperly.
    - Manual controls may not be used as intended
- Tenants may demand more comfort: temperature, humidity, and outside air
    - Or controls may be set incorrectly to create this outcome regardless of tenant demands

This is not to say that EUI is worthless: if one is limited to a single-parameter characterization of energy performance, this is a good one to use, especially in the case of operational rating EUI, as long as one refrains from assigning a normative value to the outcome: a high EUI is not "bad," and a lower EUI is not "better." Reducing EUI is "good" only if we hold everything else equal, and the EUI metric cannot tell us if we have done that.

It has been argued that higher EUIs present the greatest opportunity for energy savings because the building is so inefficient to start out with. While the first statement may be true, the second statement is not the reason. The apparent reasons are, first, that a high EUI may represent a building with high tenant energy service demands, rather than low efficiency. But with high energy service demands, efficiency investments have a higher return, so more of them can be justified. Alternately, the high usage may be due to additional end uses, such as food service or IT, which have additional efficiency opportunities that might not be worth looking at in a typical building. And finally, a given percentage savings from a base of a larger EUI is a larger number.

Asset EUI and Operational EUI. An asset rating controls for these variations in energy service demand and operational behavior by testing buildings at constant conditions. This test must take the form of a simulation model output because controlling for behaviors in a metered energy consumption context would be prohibitively expensive.

Even after such normalization, there is still too much variation in EUI to make it an appealing choice for an efficiency metric. A given asset rating EUI will be achievable with a building with low insulation, inefficient equipment and systems, and subpar lighting design if it encounters favorable conditions, such as a lack of solar exposure due to shading by adjacent buildings of vegetation, if its architecture does not demand large window area, if its occupancy does not require very much outside air, if its choice of foundation (basement, crawl space, slab on grade) is favorable for the climate, if its HVAC system chosen for cost or comfort reasons is lower in energy use than conventional systems, if its footprint requires shorter hot water pipe runs, etc. Conversely, the same EUI will be difficult and expensive to meet for variations of these parameters in the opposite direction.

This problem is further compounded by operational EUIs, which will be affected by occupancy, window management, thermostat settings, number of energy-intensive appliances such as spas and home theaters, etc.

The use of EUIs as the basis for building codes was attempted in several jurisdictions over the past 40 years, and all of them found these problems to be severe (Goldstein and Eley 2014).

However, it is possible to establish an efficiency-only metric that has been tested and found to be acceptable to the affected industries and consumers. This process involves establishing a ratio between A) the simulated performance of the proposed (or constructed) building and B) the energy performance of a reference building of the same size that has a fixed set of efficiency characteristics; . This method, which typically defines a score of 100 for the reference building, has the advantage that a ratio causes errors in the simulation model, in building input description, etc., to cancel to first order.

HERS and zEPI: Figures of Merit for Energy Efficiency

The HERS rating was developed in its current form by RESNET, a nonprofit standard-setting organization for home energy rating, in 2006. The rating establishes a reference house as compliant with the International Energy Conservation Code of 2006, setting its score at 100. A net-zero energy house has a score of 0 by definition, and the scores extend linearly in both dimensions. Almost 2 million homes have HERS ratings as of 2016. The overwhelming majority of rated homes were rated when the homes were constructed.

A HERS rating is based on the as-constructed house, with on-site subjective diagnostics for the quality of installation of insulation and measurement-based diagnostics of air leakage and duct tightness.

An advantage of the HERS system is that the rating scale remains constant over time. Thus, energy code stringency can be defined as a HERS rating target, which may be changed over time. Unlike other widely-used performance based code compliance methods, the test method for compliance does not change over time; instead the numerical target changes.

The zEPI target applies to buildings not within the scope of IECC residential buildings. It establishes reference buildings by assigning a score of 100 to a building with an energy use that is equal to the average energy use of a turn-of-the-millennium building of the same type in the same climate zone. Again, 0 represents zero net energy. ASHRAE recently established an energy standard compliance path for its Standard 90.1 based on zEPI score.

Critical review of fuel comparisons metrics in common use

Introduction

Whenever a building uses more than one fuel—typically gas and electricity but also in some cases steam or chilled water or oil—or is compared to a building that does so, a metric is needed to compute how the fuels are weighted. This is a value-laden choice: there is no "neutral" or objective way to do so. This paper argues that the choice of fuel metric should correspond to (and promote) desired outcomes for the building occupants or owners and for society. This paper discusses several potentially helpful ways to choose and evaluate the metric, and one that is not so valuable.

Because fuel metrics are value-laden, they have been very controversial over the past 40 years, especially in the U.S. but also globally. But this does not obviate the need to make a choice. Again, this choice has made many researchers uncomfortable, because there is no objective or scientific way to do so: the choice concerns the policy question of what is important. What characteristic should we try to maximize? Because the question concerns normative values and not engineering accuracy, perhaps the best solution is to make two choices and use both.

In the past, most of the methods listed next produced generally comparable outcomes. But as the electric grids decarbonize, they start to diverge.

Fuel metrics in common use and a critique of them

Site energy. Site energy is derived by taking the energy content of different sources and simply adding the numbers: thus one kWh of electricity is considered equal to 3413 Btu of natural gas, or one MJ of electricity to one MJ of gas, and similarly for other fuels.

But this choice is problematic from two perspectives. First, it creates problems with sub-optimization—optimizing the performance of a subsystem at the expense of the performance of the full system. The most obvious demonstration of this is to consider a building with a central chilling plant. The chiller has, for the sake of argument, a COP of 5. If the chiller is in the basement of the building, the energy use of the chiller is the energy input of the electricity used to run it. But if the chiller is off-site, then the energy use is the energy content of the chilled water, which is 5 times lower.

A similar argument applies broadly if one considers an entire utility grid. If the marginal fuel for electricity generation is gas, then the use of one MJ of electricity on site requires the combustion of about 3 MJ of gas (in older generators) or 2 MJ for efficient production, but in any case a value much greater than 1. Thus replacing a use of gas at 1.5 MJ on site with the identical energy service being provided by electricity for 1 MJ implies a reduction in energy use of 1/3 using site energy but causes an increase of energy use of up to 100% in reality. Site energy systematically gives the wrong guidance of fuel switching in such cases.

Note that such cases are nearly universal. Reducing loads from computer equipment seemingly affects only electricity use, but since it reduces internal loads, in fact it also reduces cooling energy and increases heating energy. This also has important elements of fuel switching if the heating fuel is gas.

Another related liability of site energy as a metric is that it depends on a definition of "energy" that is fundamentally unclear. Physics defines energy in a different way than energy planners and managers do. In physics, energy can neither be created nor consumed. ISO 50001 defines energy as: "electricity, fuels, steam, heat, compressed air, and other like media"; with an informative note: "For the purposes of this International Standard, energy refers to the various forms of energy, including renewable, which can be purchased, stored, treated, used in equipment or in a process, or recovered." This implies that energy for the current purpose means a medium that is traded commercially. This definition also aligns with common understanding of energy use, namely that ISO 50001 energy can indeed be consumed.

In the case of chilled water, or to a lesser extent steam or even gas, the definition of energy depends on the standards and specifications of a commercial transaction, and makes assumptions about how the energy is used that may be inapplicable to any particular building.

For example, chilled water may be supplied at a temperature of 10C/50F. This has one energy content if the building is conditioned to 25C/77F but a very different value if conditioned to 20C/68F. And it provides negative energy if the building is a refrigerated warehouse, unless it is used for heating rather than cooling.

Some have argued that site energy is appropriate for a building because the building owner or operator does not control the energy mix that goes into electricity (or steam, chilled water, etc.) and thus cannot influence it. But this argument is misplaced, because the building owner ***can*** control the choice of which fuel source to use when there is a choice. That choice by the building can and should be made using data on what the marginal source (or cost, or emissions) of each energy choice is, rather than the arbitrary and almost-always-incorrect assumption that one MJ of any possible fuel choice is equal.

The second critical failure of the site energy concept is that one MJ of electricity is not equal to one MJ of gas, regardless of their sources. This observation is due to the fact that electricity is a zero-entropy source, so you can do more with it. Gas creates usable energy through a combustion process, so the level of entropy depend fundamentally on what process the user chooses. But even the most efficient processes yield much higher entropy than electricity, especially for combustion within a building site.

Electricity can provide low temperature heating and cooling—the dominant (combined) end use in most buildings, with a COP of at least 3-5; whereas gas struggles to achieve COPs of 1. Since electricity can do more, it is worth more, and should be valued more highly when constructing metrics of whole-building performance.

The fact that electricity can do more is reflected in all the other metrics described below as a "source multiplier".

Some have argued that site energy makes sense because all sources of energy are denominated in the same units (which is more obvious for users of SI units). However, adding two quantities simply because they have the same dimensions yields nonsense. One might by this argument add the Reynolds number of an aircraft to its return on investment because both parameters are dimensionless, or rank a car by the adding the contents of the gas tank or battery to the torque on the drive wheels because they both are denominated in Newton-meters (foot pounds). To add two different things, they must be commensurate—of equal value.

A metaphor for this is a traveler who returns from Mexico with a US $100 bill in his wallet and a 100-peso note in his pocket. You might say that he is carrying 200 "desos," but such a statement would not be very useful or meaningful. The reason it is not meaningful is that pesos and dollars are worth different amounts. This is the key principle: to add two different parameters and get a meaningful sum, the two things you are adding must be equivalent. But in the case of fuels, they almost never are.

Note that values are not always positive values: electricity is worth more than gas, but it also has higher environmental impacts and costs a lot more.

What are we trying to accomplish with metrics? The primary problem with site energy is that it adds two or more different things that are measured in the same units but have different value. The evident solution is to weight the things according to their values.

What are the values that people care about, and who is it that cares? This paper argues that what most people care about are energy cost and the pollution emissions associated with energy use. The answers about what we care about may depend to some extent on who "we" refers to—"we" may mean the building owner, or the tenant, or society—but it is hard to identify an individual, a business, or a government that cares about "energy" as delivered to the site.

This critique is not limited to site energy: it is also hard to find to identify a business or government agency that cares about total energy use, in which one MJ of oil is worth the same as one MJ of uranium or coal or wind power.

Source Energy. Source energy is the first approach to try to solve the problems of sub-optimization and adding parameters of different values. It does this by tracing the energy use back to the original fuels extracted from the ground. Thus, if the generation of 1 kWh of electricity requires the production of 3 kWh of coal, the source multiplier is 3. Similarly, if the delivery of 1 therm of gas to the building requires the use of 0.1 therm to run the gas wells and the pipelines the source multiplier is 1.1.

This approach, as noted, elides over the fact that one MJ of coal is not the same as one MJ of gas.

There are several potential flavors of how source energy can be calculated. The source multipliers can be estimated based on marginal energy use or average, and computed on a regional basis or a national basis.

For the purpose of buildings analysis, national source multipliers allow a property owner or a city to compared energy use across regions: a lower number will always signify lower energy use. Another issue is whether energy consumption is determined by average sources or marginal sources. This difference matters in regions such as the Northwest (Oregon, Washington, and British Columbia), where the average source is nearly 100% hydro, but the marginal source is coal. From the perspective of choices made at the building level, the marginal approach makes more sense: it sends the wrong signal to tell building owners that electricity can be wasted in Seattle because the energy comes from hydro when any incremental changes to the building will result in more coal use. It would also provide misleading comparison data for a real estate company that tries to compare its property in Seattle with a comparable one in Atlanta.

The U.S. Environmental Protection Agency uses source energy as the metric for its commercial building Energy Star program, and specifies national source multipliers. Additional flavors of source energy can come about if we look at marginal sources as a function of time of the year. Energy sources at peak may be very different than those during the off peak. Even more significantly, energy sources during the early pre-dawn morning when there is lots of wind capacity and little load will be very different than the sources when renewables supply is lower and demand is higher. Such factors will be increasingly important in the future.

An additional complicating factor is how renewable energy sources are treated. The U.S. Department of Energy has looked at the sources of utility generation as being inputs like solar energy or wind, and have used source multipliers based on the efficiency of conversion. Thus, if the efficiency of a solar plant is 25%, the source multiplier is 4. Hydro therefore has a source multiplier of about 1.

This method may serve the purposes of a whole-economy diagram or map of the sources and uses of energy, but it does not provide useful information for buildings. Sunshine and wind are not limited resources, at least not for the foreseeable future. The building owner should not make choices of fuel use based on how efficiently incident sunlight is converted to electricity.

But this begs the question about how renewables *should* be counted. Two obvious candidates are: to count renewable electricity comparably to site energy, with a multiplier of 1; or, to count renewables as free.

Either choice can be problematic. Consider that from a building energy analysis point of view, renewables often ARE considered free. The concept of zero net energy buildings cannot work without this assumption. On the other hand renewables definitely are NOT FREE: they cost about as much as conventional generation sources. They also cost more than would be implied by a source multiplier of 1.

Normalized, modified end use loads (used for residential buildings only). Normalized, modified end use loads is a method used exclusively for homes, and is used in the IECC beginning in 2015 and in ANSI/RESNET/ICC Standard 301-2014 (RESNET 2016). Unlike other fuel weighting metrics, which start by evaluating the worth of different fuels, normalized modified loads starts with a desired outcome when the metric is used to analyze dual-fueled (or multiple-fuel source) buildings: the choice of heating and water heating fuels should not affect the energy score, at either low or high heating efficiency.

The equations that define this method are derived from this goal (Fairey 2000). Note that the logic of the equations works in both directions: not only do the equations derive from the starting assumption that fuel choice for heating does not affect the energy score, but ***any*** system ***other than*** normalized modified loads will inevitably result in fuel preferences, at least at some level of efficiency. In other words, any other system will allow two houses with the same level of efficiency measures to score differently based on fuel choice, and thus when used in a code compliance context will be biased in favor of one fuel or the other.

There is a potentially interesting debate to be had over whether the codes and rating systems ***should*** be biased towards one fuel, and if so which fuel it should be. Perhaps the bias should be different depending on climate or level of efficiency. But this debate has not been conducted in the code development process (outside of California, which has not re-opened the issue or changed course for decades) or in the buildings efficiency literature, so the most neutral and fairest choice for residential buildings is normalized modified end use loads.

Under normalized modified end use loads, renewables generated on site are compared to the marginal source efficiency of generation as of the early part of the 2010s.

Cost-Weighted Energy. If most people care about the cost of energy, then it makes sense to weight energy use at the site by how much it costs. Cost-weighted energy as a metric was introduced in ASHRAE 90.1-1989 both as an analytically justifiable method and as a way of achieving consensus—a mandatory condition for an ANSI-approved standard—in a way that did not include site energy or source energy.

Numerically, cost-weighted energy was nearly identical to source energy. This outcome is not surprising, because low-cost fuels such as coal require complex and expensive generation stations in order to produce electricity, while cheaper fuels such as gas can be turned into electricity more cheaply.

Cost-weighted energy was attractive to gas interests, who had been (and still are) promoting cost-weighted energy, because it yields similar results to source energy, and also to electric utility interests, since cost could include demand charges and peak load pricing, which would cause buildings to reduce peaks.

Cost weighted energy solves the conundrum of how to treat renewables: renewable sources have known costs that can be tracked.

For use as a metric to compare buildings, costs must be uniform across buildings; thus instead of allowing users to set their own costs based on utility tariffs at the building site, the standards-setting agency must establish one set of costs to be used by everyone.

California has the most advanced form of cost-weighted energy, called "time dependent valuation" (TDV), where a file of 8760 hourly values of gas costs and electricity costs are used, derived for each climate zone in parallel to the weather data assure that the highest fuel costs correspond to the largest peaks for both heating and cooling. The variations from hour to hour can exceed 10:1.

The next necessary step for TDV in a state with a 50% renewables portfolio standard is to develop a method that can credit demand response—both the ability to defer load when supply is tight and the ability to store something, such as hot water or even electricity, when renewable supply is large and might otherwise go to waste. It is not yet clear how this would be done, so additional thought is needed.

A serious liability with TDV, however, is that while it is possible to weight the value of consuming gas or electricity at different hours of the year with considerable accuracy, it is problematic to construct a ratio of gas price to electricity price. Gas, like oil and other similar commodities, is subject to immense and unpredictable fluctuations in cost. While there are forecasts of long-term gas prices that can be used in evaluating TDV, they are not reliable: the models failed to predict the run-up in gas prices a decade ago and equally failed to predict the recent drops.

Thus at the present time (mid-decade 2010s) gas appears to be low in value compared to electricity, and metrics based on TDV will encourage gas use when climate policy would encourage electricity. For example, if the choice for water heating fuel is gas at 90% efficiency or electricity in a heat pump with a COP of 3, electricity is a far better choice from a climate perspective: one kWh of gas at a combined cycle power plant producing ½ KWh of electricity will be matched with another ½ kWh of renewables (due to the state's renewables portfolio standard) to produce 3 kWh of end use hot water, whereas we would have needed about 3.3 kWh of gas to do the same thing. Thus from a source energy or emissions perspective, electricity is far better. But its TDV cost is higher.

But this problem is not easy to fix, because it is impossible to predict the cost of natural gas. While analytic methods can do a good job of weighting the value of consuming a given fuel at a given hour of the year, they cannot estimate a ratio of electricity costs to gas costs that will be valid beyond the short term.

Emissions-weighted energy. In the previous example, one of the things we cared about was CO2 emissions. It is not hard to calculate emissions weightings for energy sources, and these can be used as a metric. ASHRAE Standard 189.1 employs emissions-weighted energy as one of its metrics. As with source and cost, emissions metrics may depend on time of day and of year, are more useful if they are uniform and specified based on incremental outcomes on the grid, and also may be contingent on crediting demand response.

If a region adopts a greenhouse gas emissions cap, as California has done, and as has been done for the power sector in the Regional Greenhouse Gas Initiative, compliance with the cap entails the need to evaluate building emissions.

Observations about the Fuel Comparison Metrics

When ASHRAE introduced cost-weighted energy as a metric, all the options other than site energy produced comparable results: source multipliers for all three options were within a small range of close to 3. Thus, if a code or specification required comparing a real building with a reference building that used the same heating fuel, the multipliers only changed the outcome if the building traded off more heating for less electricity use, or vice versa, and these effects came in only to second order.

But now, as rooftop solar prices are becoming cheap, and the electric grid is becoming decarbonized, and gas prices remain at the low end of their historic range, the methods are beginning to diverge.

Source multipliers, on the margin, are beginning to vary substantially from region to region, and over time. In Western North America, strong renewables portfolio standards combined with more efficient gas plants are lowering the source multiplier towards 1, and the rest of the continent is likely not far behind. But the source multipliers for cost-weighted energy are becoming higher than 3 as gas prices are predicted to remain low while electricity prices trend slowly upward. Time of use cost weights are becoming more relevant as renewables gain in market share both on the utility side of the meter and at the building site.

Emissions-weighted source multipliers are becoming much lower and also more regionally dependent. Incremental source multipliers may change more rapidly over the years than average.

Constancy of Metrics over Time

Since metrics of building energy use are used for comparisons between buildings in the marketplace, and in particular at point of sale, there is clear value in maintaining the calculational structure of metrics over time. If Building A has a rating of 55 this year in 2016 and looks better than Building B with a rating of 70, it could generate confusion if in 2026 Building A's rating were 45 but B's were 40 and neither building had been retrofitted.
On the other hand, since all of the metrics except site energy depend on the characteristic of the incremental supply of gas compared to electricity, the metrics will in fact change over time.

The authors suggest that the best compromise could be to calculate metrics based on 10 year or 20 year forward projections, and to recalculate every decade or so. The factors could be adjusted so that a building at typical efficiency receives the same score under the new system as under the old one (both for gas and for electricity separately).

Concluding Observations

This growing divergence between the outcomes and implicit recommendations from the different fuel comparison metrics suggests that it may not be possible to rely on just one. This problem occurs in part because the metric of cost may be of most interest to the building owner or occupant, the metric of emissions may be of equal or even greater value to society. The conundrum is that while most people surveyed in public opinion polls support renewables and efficiency, a metric that underweights renewables by focusing only on emissions suggests that it is acceptable to waste renewable energy.

One problem with the choice of metrics that we would like to avoid, both in the case of metrics for building efficiency and metrics for comparing fuels, is what economists call "Goodhart's Law." (Note that in economist jargon, a "law" is what physicists would call a "conjecture:" an observation that has been found to be correct in some cases and is believed to be likely true more broadly.)

Goodhart's law can be stated as "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes." In this case, the concern is that a metric that is used as a standard or a normative benchmark will induce behaviors that negate the underlying goals of defining the metric in a particular way. This could happen with fuel comparison metrics in several ways. For example, a cost metric that shows very high costs for the time period 6pm to 10 pm will cause building designers to construct or retrofit buildings to pre-cool during earlier periods when there is ample solar generation, and to focus on well-controlled efficient lighting, thus reducing the electricity costs for that hour.

Such problems can be minimized by developing a policy context for the use of metrics that minimizes adverse effects from errors in the metric. An example of failing to do this would be to establish a numerical limit on EUI as measured by time-weighted costs, or emissions, or source energy. Small changes in the weights could lead to non-optimal choices of fuels or tradeoffs between fuels. A better approach would be to use ratios, such as zEPI or HERS, that allow errors in the fuel metrics to cancel to first order.

In sum:

- The metrics most related to what most people care about are cost and emissions;
- Until recently all these metrics (except site energy) led to similar conclusions, but;
- Now they are diverging: as regional grids decarbonize, electricity is favored by the emissions metric and disfavored by the cost metric;
- going forward it is more important to compare them and select the tool best suited to the job;
- One metric may not be enough.

Thus, the authors suggest the following approach may be best, noting that implementing it will not be simple. The two metrics of most interest are cost-weighted energy and emissions-weighted energy. When the two metrics suggest contradictory choices, there is no broad rule about which choice is best, but rather it is an issue of business or professional judgment. Implementing this approach on a broad basis will involve setting a policy choice for the ratio of average (over the year) cost of electricity compared to gas, which will yield a smaller value than current cost government-sponsored projections. It will also require adjustment factors that make for a smooth transition between current rating schemes—both for asset and operational ratings and both for residential and nonresidential buildings—that results in no change for a typical building.

References

Bloomfield, C. and P. Bannister. 2010. "Energy and Water Benchmarking in the Retail Sector: NABERS Shopping Centres." In *Proceedings of the ACEEE 2010 Summer Study on Energy Efficiency in Buildings*, 3:1-13. Washington, DC: ACEEE.

Eley, C. "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.

Fairey, P., J. Tait, D. Goldstein, D. Tracey, M. Holtz, and R. Judkoff, "The HERS Rating Method and the Derivation of the Normalized Modified Loads Method." Research Report No. FSEC-RR-54-00, Florida Solar Energy Center, Cocoa, FL, October 11, 2000.

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

Goldstein, D. and C. Eley. "A Classification of Building Energy Performance Indices." *Energy Efficiency Journal*, 7:353-375 (February 2014).

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

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

National Academy of Sciences (2010). *Real Prospects for Energy Efficiency in the United States*. Washington, D.C, 2010.

RESNET (Residential Energy Services Network, Inc.) and the International Code Council. "ANSI/RESNET/ICC Standard 301-2014 Standard for the Calculation and Labeling of the Energy Performance of Low-Rise Residential Buildings using an Energy Rating Index". January 15, 2016.
.