Numbers, stories, energy efficiency

Mithra Moezzi
School of Urban Studies and Planning
Portland State University
1825 Southwest Broadway
Portland, Oregon 97201
USA
mithra@pdx.edu

Keywords
policy-making, energy behaviour, quantitative study, statistics, political targets, transition, metrics

Abstract
This paper explores the role of quantification in energy efficiency policy and research, and advocates for modernizing quantification practices. Efficiency policy rests on numbers and stories about these numbers, which together shape problems as suitable for human action and guide what solutions seem desirable or possible. This occurs on many levels, from practical definitions of energy efficiency to long-term policy goals. Drawing on policy sciences and the philosophy of statistics, we show why something as basic as quantification merits a closer, social sciences-based, look.

Energy is used in diverse, complex, and dynamic ways, and configurations change constantly over time. Numerical summaries and protocols for judging efficiency always reduce and simplify. They are generally poor at conveying change and relationship. So it is crucial to attend to what numerical summaries and protocols do – what they express, standardize, and hide – and consequences thereof. The paper analyses some current energy metrics to illustrate the difference that "how you measure" makes. In turn, the stories associated with efficiency quantification provide schemas of what efficiency means and how it can (or did) happen. These stories develop to "stick" politically, but often incorporate little of the complexity of relationships and interactions that are the hallmark of society and energy systems within. These over-simplifications are especially problematic in the era of climate change, climate change policy, and internationalism.

The paper suggests improvements to how metrics are used and interpreted, and underscores the need for more intense scrutiny of quantification practices and better recognition of the limits of quantification. There is no perfect metric, but there is a need to more fully understand what they do, and to create policy visions and causal stories that better recognize diversity, interaction, and the role of people in creating and changing energy use, versus just as buyers and users of efficient hardware. These new stories and aesthetics of quantification can help bridge gaps between what policy now aims to control and what would make more difference. This paper explores the role of quantification in energy efficiency policy and research, and advocates for modernizing quantification practices. Efficiency policy rests on numbers and stories about these numbers, which together shape problems as suitable for human action and guide what solutions seem desirable or possible. This occurs on many levels, from practical definitions of energy efficiency to long-term policy goals. Drawing on policy sciences and the philosophy of statistics, we show why something as basic as quantification merits a closer, social sciences-based, look.

Introduction
The problems, solutions, and evaluation of energy efficiency rest in quantitative expressions of energy use. Obviously so: the original heart of energy efficiency policy is engineering, and as Salais (2012) comments, governance is increasingly quantitatively based.

In everyday work in the field of energy efficiency, energy use metrics appear relatively objective. Differences between one form of expression of energy use and another may be noticed but have the flavour of technical details – a by-product of necessary standardization, but not of much general interest. In fact the topic of energy metrics has been called "lethally dull" (Ueno 2010). Despite the ubiquity of quantification in the energy efficiency field – the estimates we derive, claim, and quote – there has not been much social scientific (or even statistical) attention to the social, subjective, and reductive nature of numbers in expressing, potentiating, and evaluating energy efficiency,[1] nor to the statistical quality of these estimates. How are energy efficiency expressions defined and derived? How do they convey and shape activities of a complex world? What are the unintended consequences of this communication – what do these expressions see and what do they miss? How are they accidentally or deliberately misconstrued to favour particular narratives about value and progress? And what is the general scientific quality of the number generating and analytical processes with respect to the goals and claims made for them?

In the earlier years of energy efficiency research and governance, policies focused on increasing the efficiency of relatively isolated elements – insulation, furnaces, etc. – in fairly restricted geographic areas or contexts. This piecewise attention has since evolved to address larger systems, such as of entire houses or commercial buildings, a bigger and more diverse set of end uses and situations, including wildly different ways of doing whether across individuals or across countries. More schematically, it has set sites on addressing crossover between supply and demand (e.g. load curves and renewable supply) and interactions between sectors, such as transportation versus residential buildings. Most importantly, it has also been reoriented to address aggressive policy goals aimed at reducing total societal greenhouse gas emissions versus the more relative nature of energy efficiency *per se*. This moves the game to a more societal rather than end-use platform. The increases in scope, depth, and scale brings the importance of interactions and path-dependency (Labanca and Bertoldi 2013) into view, as well as cross-cultural cross-social challenges for defining, legitimating, and managing the nature of the energy services that energy use delivers. The shift of attention from relative to absolute savings also means coming to better terms with the fact (or experience) that end use efficiency as commonly defined is often associated with higher energy use overall (Harris et al. 2010), and that energy "wants" often appear unlimited. Neither of these are fundamental problems if the goal is efficiency *per se*, but they matter much more if the goal is absolute reductions of energy use or energy-related emissions.

The anthropologist Leslie White wrote that "on all levels of reality … phenomena lend themselves to description and interpretation in terms of energy," and argued that societies with higher and more efficient energy use were more advanced (White 1943). Though White's take on cultural evolution now appears old-fashioned or politically incorrect, it highlights how very "social" energy metrics can be. Statistical descriptions of energy efficiency reflect visions of social good and progress, as Hacking (1983, 1990) argues for statistical descriptions of populations in general. If the purpose of energy efficiency policy is to lower societal energy use with consideration to well-being and other societal values, then metrics are fundamental to capturing this purpose.

The paper first presents a short background on the history of statistics and sociological and anthropological interpretations of quantification, standardization, and the causal framing of policy problems. It then turns to some of the general characteristics of energy intensity metrics, as one major type of energy quantification. The core of the paper presents examples illustrating dilemmas that occur in defining efficiency and describing energy use in the real world. The final sections of the paper generalize across these examples, summarize some current challenges, and offer recommendations that can contribute to progress toward a more modern, more probing, system of quantification and non-quantification suitable to the global multi-disciplinary problems of environmental sustainability.

Statistics, commensuration, and power

The field of statistics originated in state-led efforts to describe and count the populace for purposes of tax collection and estimating military resources. These first such assessments were over two thousand years ago, while modern efforts became widespread in the eighteenth century, closely linked with the bureaucracy of the state (e.g., Hacking 1990). Historians of statistics describe these censuses as far more than a neutral act. Rather they are a way of defining and controlling a population, evoking power through counting and categorizing. From here the field of mathematical statistics developed, first focusing on characterizing people and then covering other topics (e.g. environmental characteristics), devising techniques for inference and compact descriptions of qualities, quantities, relationships, diversity, uncertainty and so on, which otherwise might be amorphous, undefined, or incommunicable. This quantification and accompanying statistical analyses make the abstract and diffuse into something visible, comparable, and evaluable. In so doing it also reifies certain elements of the world, ignores others, and helps define deviance and normality, sometimes contentiously so (Anderson 1990, Desrosières 1991).

The point of this diversion is to make strange the now very normal process of quantitatively describing the world and the flows within it. Sociologists have pointed to the "increasing emphasis of quantification in modern science and policy" (Espeland and Stevens 2008:402) and argued quantification has become "the heart of state power" (Espeland and Stevens 2008:417, Salais 2012). The concern in this paper is not the rendition of power *per se* but rather how quantification and statisticalisation operate with respect to pursuing energy efficiency and greenhouse gas emissions reductions. This quantification provides a "grammar" (Espeland and Stevens 2008) of energy efficiency, so to speak, but a simple one in which only certain things can be seen and said, and where variability is typically hidden by averages. This grammar pairs with institutional lexicons of efficiency such as outlined by Lutzenhiser (2014).

The process of quantification of the world is especially interesting in the case of energy use because of energy's ubiquitous nature, the invisibility of energy flows, and the need to evaluate energy use with respect to some measures of social value, be it a square foot of a house, a passenger-kilometre of transportation, or the energy sustainability of a city. That is, energy efficiency metrics and standards require an overall assessment

1. This is not to imply that the details of standards-making are apolitical, as they are clearly of interest to market and other stakeholders, and it is well-recognised that there will be winners and losers in this metricisation. But there is at best narrow attention to the practical effects of these details in terms of what gets built, sold, and done.

of what work was rendered by energy and judgements about whether this use was efficient. Through numbers and associated rules (e.g., test protocols or standards) there are mechanisms to describe what is normal, what is good, what is the minimum acceptable level officially allowed, and what is pathologically "wrong" (and thus needs to be corrected or is instead considered so uncontrollable that it should be ignored). This standardization, in turn, influences the character and form of the homes, buildings, cars, etc. built, in general having a levelling effect (Shove and Moezzi 2002, Timmermans and Epstein 2010).

Deborah Stone (1989) comments that it is a truism of policy sciences that difficulties become problems only when they become seen as amenable to intervention. Thus policy problems and general solutions to these problems go hand in hand. Solutions, Stone argues, rest on some image or narrative of a cause, setting up policy action to pursue proposed repairs and to attribute responsibility. This causal narrative moves explanations from fate to a matter of human agency (Stone 1989). In the case of energy efficiency, "market barriers" and "market failures" have long been posed as the generic cause of economically-inefficient levels of efficiency, legitimating particular actors to fix (Stone 1989) these problems. The rationale for applying energy efficiency to address climate change, however, needs to reach beyond these market explanations, as many of the most important costs (global air pollution, global effects, vulnerability, etc.) are external to individualized cost-benefit analysis.[2]

As noted above, societal energy use is clearly a highly complex system (or system of systems) (Labanca and Bertoldi 2013). From a social scientific point of view, any energy use can be seen to be created or shaped by a tangle of circumstances, interactions, trade-offs, evolutions, etc. These complexities are difficult to see and address through data collection, because there are so many moving parts, so much interaction, so many different ways of doing things, and so many different nodes, with the nature of interactions differing from one node to another. However, complex arguments do not have much leverage in policy-making. Rather, "complex causal explanations are not very useful in politics, precisely because they do not offer a single locus of control, a plausible candidate to take responsibility for a problem, or a point of leverage to fix a problem. Hence, one of the biggest tensions between political science and real world politics" (Stone 1989). It may be increasingly necessary to bring this tension into view, given the nature of climate change and its effects: global, unevenly distributed, unpredictable, and highly path-dependent. This requires sophisticated versions of energy efficiency liaised with efforts that transcend energy efficiency's traditional purview (e.g., societal resilience, land use planning, coordinating supply and demand).

Energy efficiency and policy research is just beginning to take on the importance of stories in shaping what the field does and what it should do. Much of the attention has been on stories to intended to persuade end users to act in different ways, not particularly relevant to the topics covered in this paper. But there has been on how energy efficiency policymakers and researchers use stories to describe and defend (e.g., the "hero stories" described by Janda and Topouzi 2013), what current efficiency metrics and indicators convey, or the using stories to learn about complex processes that standard technical and quantitative lenses on energy use might miss but are difficult to observe first-hand (Janda and Topouzi 2013, Moezzi et al. 2014).

Numbers used in energy efficiency policy generally come along with stories or proto-stories, such as the causal narratives noted by Stone, the hero stories analysed by Janda and Topouzi, general mental models such as *homo economicus* or properly concerned citizens, and most fundamentally, the idea that increasing the technical efficiency of hardware helps solve climate change. Given the challenges faced and an increasing acknowledgment of the need for complexity in describing and intervening in societal energy use, the simple stories currently favoured may be a disservice to achieving the nominal goals of energy efficiency and climate change policy.

One of the aspirations behind this paper is to work toward exploring opportunities for developing pairings of narratives and numbers that do a better job of developing a multi-layered approach to energy sustainability. Ideally this approach could help support more precise and localized improvements, better see important interactions amongst various components of energy systems (e.g., a dark side of standardization; see Shove and Moezzi 2002), expose the multi-dimensionality of changes (e.g., building energy efficiency might also affect air quality, occupant well-being, and longer-term pathways of technology development), and balance the narrowness of numbers with a more realistic social layer that can get at some of the what, why, and how that create both desired change and less-desired unintended consequences. While only scantly developed in this paper, a folkloristic perspective on traditional narrative highlights the importance of story structure and organisation (Dundes 1962), as well as the innate appeal of somewhat complex stories or models (e.g., outcomes differing by actor or situation, path-dependency) as opposed to aphorisms, which are often not very believable or interesting. Eidelson (1997) discusses a variety of examples of complexity theory in social and behavioural sciences, and Baskin (2008) argues draws parallels between storytelling and complex adaptive systems.

Figure 1 is schematic of the four main elements discussed above. Energy use, in all its complexity, is central. This energy use can be described by a combination of numbers and stories about (or interpretations of) these numbers, represented as the two triangles at the figure's base. The policy triangle is on the top. Research floats over all four elements. While research and experimentation may be able to probe complexity and treat energy use with some intricacy, policy is in a harder position, since understanding is not doing, doing is always political, and policy instruments are blunt. This paper, in overview, probes aspects of how stories and numbers about energy use currently work together, and how they might work better, to help improve policy effectiveness in the future.

The concepts of energy efficiency quantification and numbers evoked in the discussions above have been admittedly vague. This paper is not able to do justice to a fuller treatment of numbers in energy efficiency, which include (a) metrics for energy efficiency; (b) metrics of evaluation;(c) the process of

2. Certainly it is possible to frame climate change as an market failure for which market solutions are available (e.g., The Stern Review).

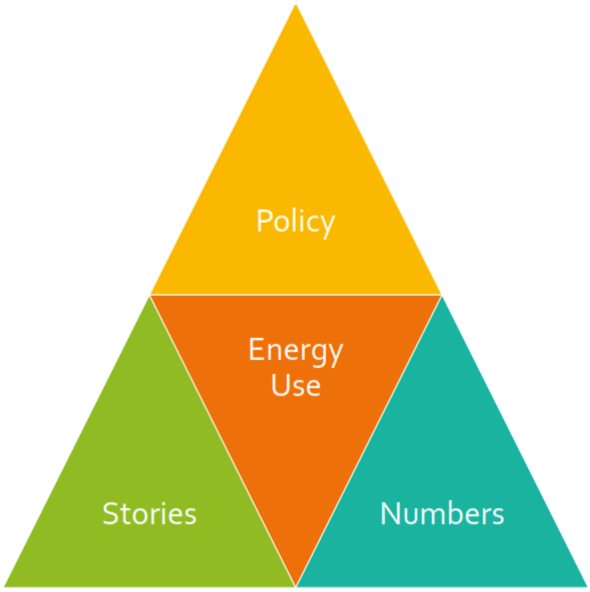

Figure 1. Triangulating energy use for policy applications.

data collection; (d) statistical analysis; and (e) "rhetorical" and false numbers (Lampland 2010). Most of the analysis will focus on metrics for energy efficiency, in particular pertaining to the examples described in the examples below.

Energy intensity metrics

To be used in policy, thermodynamic and other physical definitions of energy efficiency have to be translated to notions of social output (Patterson 1996). Often this is done through an energy intensity metric, which express efficiency as the ratio of energy use relative to a measure of social utility. There are many dimensions of seemingly simple expressions of energy intensity. Table 1 lists some of these dimensions. Energy-based numerators (first row of table) are paired with denominators that reflect some social output (second row of table). These ratios may be aggregated (e.g., per capita) or calculated as individual units (e.g., energy use per occupant in a house). Distributions across individuals can be summarized statistically, e.g., as averages, or benchmarked relative to some percentile of performance. Data are drawn from various sources such as surveys or regulatory filings, and they may often be model outputs rather than direct measurement. Values can also be adjusted for various factors toward comparability, such as weather-normalization, or calculated relative to a counterfactual, such as in technical and behavioural savings potential. Each element in this array of intensity inputs has various types and degrees of uncertainty, often unknown and unexpressed. Any metric on energy efficiency may have a complex pedigree. By nature, people themselves are not often directly included.

Quantifying energy efficiency frames a process and reduces its complexity to fit that frame. Energy use per unit floor area, for example, does not capture house size, total energy use, how a house is used, or air quality and comfort. In standards, some of these aspects are conditions of the definition, e.g., that at least in theory, certain air quality will be achieved. This conditioning of course is part of the elaboration in standards-making and. converting "qualities into quantities" (Salais 2012).

Illustrations of energy efficiency metrics

EXAMPLE 1: IS ENERGY USE TRENDING UP OR DOWN?

Whether energy use is trending up or down over time is a basic question, one that at first seems easy to answer anywhere that supply or consumption are tracked. But in fact it is easy to arrive at *either* answer with nearly the same set of data or with various choices about what fuels to count, what standardization to use (e.g., per capita, weather-adjusted), or how to define sectoral boundaries. Figure 2 shows annual U.S. residential sector energy use between 1960 and 2011 per million people for two different measures of energy use: net energy use (which excludes losses from electricity generation) and total energy use (which includes generation losses and so is higher). These two options for measuring residential energy give very different pictures of consumption trends: *net energy use* per million capita shows a satisfying and continuing decline between the mid-1970s and 2011, while the *total energy use* per capita is nearly the same in 2011 as it was in 1975.

In describing residential energy use trends in a news brief, the U.S. Energy Information Administration featured the declining net energy use trend (per household), attributing this decrease in energy use to improved efficiency despite the increasing size of houses and the proliferation of electronics (EIA 2012). But the increasing total energy use trend is a more complete picture for the energy used by the residential sector. Savings are relative to a counterfactual anyway, so this claim does not mean that efficiency did not save energy. Rather even in this simple case, it is easy to legitimately tell quite different stories with data that appear to be measuring nearly the same thing. Experienced energy data analysts know the difference between net and total energy use, but many observers would not. Rather, authoritatively published data become facts (Espeland and Stevens 2008), sometimes substituting for much more general conclusions (e.g., "energy use is going up."). This typically happens without accounting for the conditions, assumptions, and uncertainties, which are substantial for many of the types of data used in energy efficiency.

The choice of net energy use in the example above isn't wrong. The missing energy between the net and total energy use line can be allocated to the industrial or generation sectors. But net energy use is not a very good diagnostic for trends in total energy use by households. In this case, the major source of conflation is the fact that electricity has become an increasingly important household fuel over these six decades, some of it through changing fuels for space and water heating, and some of it for the proliferation of devices that were uncommon or did not exist at all.[3] On the one hand, the claim that energy use per household has decreased can be seen as defensible opportunism that helps protect the industry from critics. On the other, the fact that these sorts of stories persist with apparently little debate suggests that there is something disquietingly lax about the level of scientific debate, even that challenge may be unwelcome. This non-debate limits the ability of the energy

3. The share of electricity in residential energy use increased from 26 % to 70 % between 1960 and 2011, so the trend lines are not parallel. Also, the ratio between total energy used for electricity and retail electricity sales gradually declined by 12 % over the five decades shown in the figure: from 3.5 in 1960 to 3.1 in 2011 (i.e., generation losses have decreased over time).

Table 1. Dimensions of energy efficiency and consumption metrics.

Dimension	Examples
Numerators: Energy or emissions units	Site electricity use Total primary energy use Computed GHG emissions
Denominators: Units of output	Per household Per square metre Per person Gross Domestic Product
Statistical qualities	Average Individual Distributional (e.g., median)
Data source	Census, survey, etc. Simulated, modelled Empirical measurement of various sorts
Adjustments	Weather-normalized “Typical” usage (e.g., asset ratings, or "market basket" approach) Factorial decomposition
Realm	Single-family only Program participants only Year

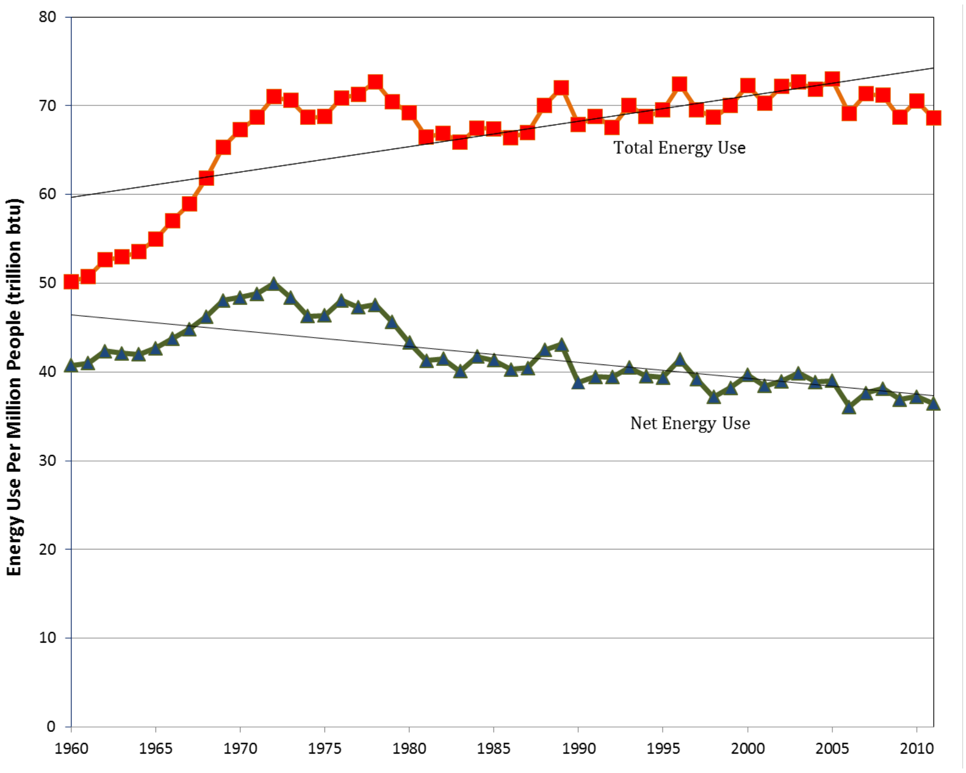

Figure 2. U.S. residential energy use 1960–2011 for total vs. net energy use per million capita (EIA 2014).

efficiency industry to address the level of complexity of the real world in ways that could actually help better shape and pursue policy goals.

As to expressing trends in general, there are all sorts of options for what data to show: expressing trends in one fuel only (e.g., electricity only rather than accounting for most or all fuels), expressing total energy use in a geographic area when say, industrial energy use has been transferred to another geographic location, the choice of certain starting and ending points, etc. The Energy Information Administration warned: "Caution is warranted in the interpretation of residential energy-intensity indicators since the magnitude and direction of the energy-intensity indicator changes are dependent on the choice of the demand indicator" (EIA 1995). It is not uncommon to find conflicting stories about energy use trends, consumption estimates, etc., published by the same institution or by various reputable scholars. For example, a debate about whether a sample of LEED-certified U.S. commercial buildings "saved energy" yielded a number of different answers (Gifford 2008, Newsham et al. 2009, Scofield 2009, Turner and Frankel 2009).

It would be an omission to not mention the political aspects of energy efficiency indicators, trends, and other energy quantifications. There is politics of what numbers are presented and how: in particular, there is a bias toward presenting data in a light that suits desired narratives. Often the data analysts and statisticians who are best acquainted with the conditions, assumptions, and weaknesses of their estimates are not the ones with final authority to see what gets published and how it gets interpreted. Understandably so, as the livelihood of institutions often depend on quantifying past successes, promising future successes, and underscoring the need for them. This process is clearly difficult to control, while the problems it raises may be bigger than imagined (Moezzi and Bartiaux 2007).

EXAMPLE 2: MEASURING RESIDENTIAL ENERGY EFFICIENCY: WHERE ARE THE PEOPLE AND WHAT DO THEY NEED?

Consider the problem of measuring the energy efficiency of a house in order to compare it to other houses. This arises in developing building energy codes, in assessing the efficiency of existing homes, and in benchmarking actual energy consumption of occupied homes. The most basic expression for these comparisons is energy use per house. Houses vary greatly in terms of what they provide, such conditioned floor area, and where they provide it (e.g., in the south of France versus Finland). Various methods are used to adjust for the most obvious of these differences, such as weather normalization. Likely the most common method of standardizing house-level energy use is to figure energy use per unit floor area of total or conditioned space From an engineering perspective, the per-unit-floor-area metric is a straightforward and reasonable approach to comparing space-conditioning use, which is (and this is a major assumption about what houses should do) presumed to be a core service making the entire house habitable at any moment. The per-floor-area metric is ignorant of occupants and occupancy. Simulation modelling routinely predicts energy use in new homes based on assumed occupancy and occupant behaviour, though these assumptions do not necessarily capture actual behaviour even on average (Lutzenhiser et al. 2015).

From the perspective of tracking absolute levels of energy use or greenhouse gas emissions, normalizing by floor area has the weakness of failing to capture total energy use. Furthermore, per-floor-area energy intensity metrics put bigger houses at an advantage with respect to lowering energy intensity (Prahl 2000; Ueno 2010; Wilson & Boehland 2005). This is both because space conditioning energy use is generally lower per square foot due to reduced losses through the envelope, and because (holding other energy-using activities equally), energy consumption from other end uses are distributed over a greater area. In one simulation study, a 2,694-square foot [250 square meters] house had a 30 percent lower energy intensity per square foot than a 1,700-square foot [168 square meters] house – even though the larger house was estimated to use 38 percent more energy (Ueno 2010). The bigger house is thus considerably less energy-intensive but more energy-consuming than the modest-sized house. Total energy consumption per house "overflows" the per-square-foot efficiency metric, an Achilles' Heel from the perspective of reducing total energy use. This also does not mean that efficiency is necessarily the cause of bigger houses, though it is possible that it does encourage increased size (Ueno 2010).

One alternative to a per unit floor area energy intensity metric for a home's energy efficiency is a "per occupant" energy intensity metric. This is a way of integrating people into the expression of energy efficiency. Figure 3 shows the results for electricity consumption per person (horizontal axis) versus electricity consumption per square foot (vertical axis) for a sample of California single-family homes. The correlation between per-person and per-floor-area measures is 0.36.[4] There is clear positive relationship between the two measures, but the relationship is not very strong. As can be seen from the figure, many houses that are quite efficient with regards to a per square-foot metric are among the least efficient houses from the perspective of a per-occupant metric, and vice versa. To illustrate this point further, the quartile ranking of each household was determined for both the distributions of both the "per-person" and "per-square foot" metrics of annual household energy use. For natural gas, for example (not shown), nearly 10 percent of households were in the top quartile of the per-person distribution, but the bottom quartile of the per-floor-area distribution. Similarly, nearly 10 percent of the top quartile of the per-floor-area distribution were in the bottom quartile of the per-person distribution. This is about what would expect if the relationship between the per-person measure and the per-square-foot measure were statistically random. In short, describing energy efficiency in terms of a per-person energy intensity metric can give a much different picture of a home's "energy performance" than a description based on a per-unit-floor-area energy intensity measure.

For many policy applications, specifying efficiency benchmarks by number of occupants isn't viable. In particular there are no occupants in a house before it is built. But for understanding the energy use of existing houses, a per-occupant metric invites a different perspective for assessing efficiency than a per-floor-area metric. It better reflects the fact that what people do, and what they are assumed to need, matters. We don't know what would happen if a household that is now low-consuming with respect to the per-unit-floor-area metric moved into a

4. The analysis based on the California subsample of the U.S. Energy Information Administration's Residential Energy Consumption Survey (RECS) for 2009.

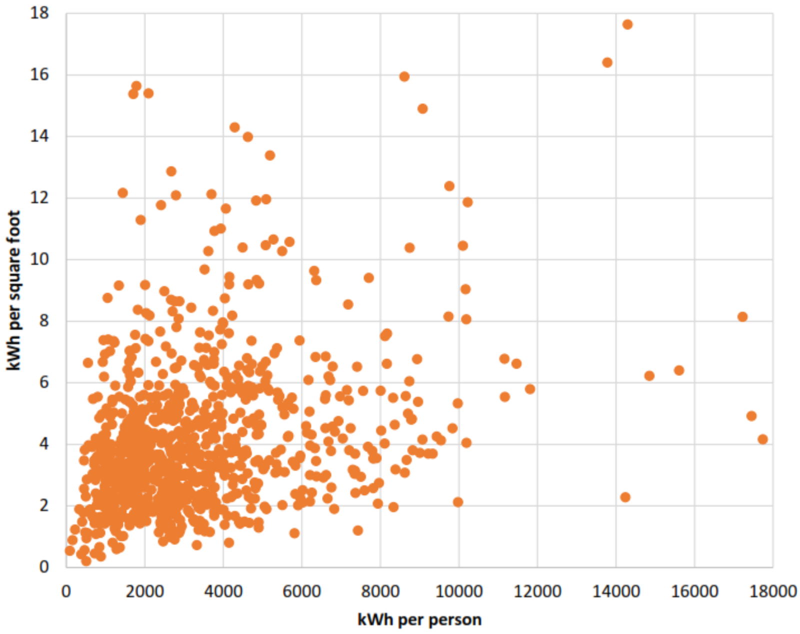

Figure 3. Comparison of electricity use per person vs. electricity use per square foot for single-family California residences (based on EIA RECS 2009 data).

house that was high-consuming by this metric. The point is not to judge whether occupants behave "correctly," but rather to provide a perspective that invites more attention to a broader way of defining and achieving energy efficiency. The basic stories that accompany these two metrics are different, each suggesting somewhat different questions and somewhat different paths forward.

As discussed earlier, one recommendation from applying a systems perspective to energy use is to evaluate actual performance on multiple levels and compare this to predicted energy use. In this case, that would include looking at actual energy use in addition to modelled energy use. To what degree are houses obtaining the levels of energy performance that models predicted for them based on nominal efficiency features and construction characteristics? And if there are systematic biases or large errors in some cases, why? A second level of evaluation could include looking at performance from a different angle. A per-person metric is more aligned with the total energy use of "the housing system" than is a per-unit-floor-area metric, which more closely targets efficiency as an independent matter of technology improvement.

EXAMPLE 3. MODELLING AS AUTHORITY

Models are everywhere in energy analysis, including the development of standards and codes, building design, unit end use consumption estimates, savings potential estimates, evaluation results, and even the fundamental nature of savings (i.e. Negawatts). Models are used because there is often no other choice, but many of the assumptions therein are hidden, and the fact that some modelled data have high levels of uncertainty and little empirical backup may be unrecognized by most observers (e.g., Unit Energy Consumption estimates; see Lutzenhiser et al. 2015) Even when modelling estimation can be roughly compared to empirical data, this is often not done. There has recently more attention to the difference between modelled energy use compared to actual energy use in residential and commercial buildings. Though we know of no systematic review of these comparisons, the general impression is that correlation between modelled and actual energy use is often fairly low (e.g., Halladay 2012, Stein and Meier 2000), modelled energy savings for upgrades to residential buildings may tend to overestimate actual savings (e.g., Brown 2012, Lancaster et al. 2012), and buildings designed to be low-energy may not perform that way (e.g., Karresand 2012). The point here is not to complain that modelling is not good enough, but rather to call attention to the potential importance of paying attention to what accounts for the difference between actual and predicted energy use, even where modelled energy use is correct on average. These gaps may contain a wealth of information, as well as improvements that can be made by "middle actors" (Parag and Janda 2014, Brown et al. 2012). Energy use models might in general be seen as perspectives on problems, rather than surrogates for them, to use terminology introduced by Strauch (1975).

For another example of why these differences matter, consider the case of regulatory energy efficiency definitions. These definitions are often based in energy test procedures that rely on assumptions about how the product will be used (Meier and Hill 1997). If actual use is much different than these assumptions, actual energy use of the installed product might be quite different as well. This can bias aggregate savings estimates, but beyond inaccuracy, a product or building that is "energy efficient" under one set of usage conditions may not be so energy efficient under different condition. The precision of the definition of what energy efficiency is and how to achieve it has consequences. For example, some Australian researchers argue that home energy efficiency rating systems can antagonize lower-energy designs for cooling, and make it more difficult to encourage "free-running" buildings (Kordjamshidi and King

2009; Williamson 2010). That is, a house that does not need a mechanical air conditioner can provide lower energy use, and perhaps comfort that is least as good, as one that does not.

Discussion

Energy efficiency numbers are not as solid as we think they are or as we use them. Metrics frame social activities, problems, and solutions in particular ways. Some of what they do not capture, what falls out of their framing, is important, and some of what they count and encourage might only appear to solve problems (Brown 2010). In addition to the subjective and partial nature of how current conventions quantify energy use and energy efficiency, measuring the energy flows in real world processes is difficult statistically. Quantities are uncertain, and these uncertainties are rarely acknowledged or estimated. From a statistical perspective, analysis often proceeds with more confidence, and less critical evaluation, than warranted. This is conventional and pragmatic – acknowledging uncertainty and providing alternative interpretations is counter to the clean, clear, explanations favoured in policy causation and solution stories. As the scope, scale, and geographic and cultural spread of energy efficiency efforts expand from their more limited origin of energy efficiency as a physical property of isolated processes or material, to addressing goals for extraordinary reductions in greenhouse gas emissions the need to interrogate the framings of efficiency metrics has grown. The following list summarizes the main conclusions of this analysis and recommendations that stem from it.

1. *Quantification reduces the complexity of real world processes.* This reduction and standardization is the very point of quantification, to render certain qualities of processes more visible, comparable, diagnosable and manageable. In privileging certain qualities of any process, other considerations fall outside of the frame entirely (e.g. what people are really doing as opposed to what might be assumed or what is average, and non-energy considerations such as vulnerability, resilience, air quality, equity, etc.). Some aspects are so embedded in the frame that they can't be seen or questioned (e.g., assumptions about what services are required). In planning, data analysis, and even in public-facing information such as energy labels,[5] it may be useful to routinely present the results of multiple metrics. For example, a house energy label might include modelled energy use per unit floor area, total "asset" energy use, and observed energy use if available. An analysis of trends in energy use might consider several perspectives (e.g., electricity use, total energy use, interactions with other sectors, etc.). And finally the qualities that are often just assumed to be satisfied (e.g., good air quality, not too much noise) need more attention to be brought into parity. The emphasis here is less on elaborate compact description (such as in giant sustainability indicator sets) but on understanding what's happening and directing attention to developments that might be problematic.
2. *In focusing attention on certain aspects of energy use (and the material configurations that set up this energy use), other aspects or opportunities may be ignored.* Further developing the last point, quantifying processes and setting standards with respect to these quantities has the tendency to reduce diversity of the energy use system. These transformations by nature are often sneaky and difficult to see. With respect to energy use itself, standardization can also make it more difficult to produce alternatives that use even less energy (e.g., smaller houses that meet energy efficiency criteria or "free running" houses, as suggested above). The metrics that are used in policy and research should be analysed for what they capture and what they leave out. For example, as noted above, absolute levels of energy use are not well-captured by the parameters of a per-square foot measure of energy efficiency. So, for any metric, the question should be asked how well the values that inhere in quantitative expressions capture policy intents? Or the things that people (in all their flexibility) value? Any measurement system of energy use is a simplification, and inevitably, some important considerations will be left out. Evaluation methods should pay attention to blindness and to "overflows" (Callon 1998) both structurally and quantitatively.
3. *People are often left out of data collection, energy use metrics, and evaluations, so their role often remains minimized.* Quantitative expressions of energy use can only see what is measured or modelled. Data collection is always selective, and can only see certain things while being completely blind to others (Anable et al. 2014, Lutzenhiser et al. 2015). Because so little data is collected on what people do, and why, their role in shaping energy use often remains invisible, or reduced to whether they buy or act energy efficiently (Lutzenhiser et al. 2015). Modelled energy use is often based on assumptions of ' "average" or "typical" users that have little empirical basis and in any case cannot capture the variety of actions, situations, or ingenuity. So more attention to capturing these actions quantitatively (as well as qualitatively, as argued below) can increase the ability of technologies and policies to negotiate this diversity.
4. *The role of evaluation and learning can be strengthened.* One of the general messages of sustainability science, as well as of a complex adaptive systems view of energy use (Labanca and Bertoldi 2013), is that progress and change need to be assessed from a variety of disciplinary perspectives and at a number of different levels. This should entail empirical observation, and not just numbers but also sufficient meta-information about these numbers, including qualitative information that helps answer "what are people doing?" and "why?" As discussed above, models and model outputs are sometimes imbued with such authority that differences between modelled and observed results, if noticed, might be attributed to people doing something "wrong" – rather than, say, problems with model assumptions, implementation, or technology design. For example, the rebound effect is often evoked when measured energy use exceeds modeled energy use, with the implication that technical efficiency actually improved but that the lower costs of energy services associated with this increase in efficiency were welcomed with

5. See Ueno (2010) for a discussion of how this idea might work for home energy labels. There is reason for scepticism that the public "wants" more elaborate or sophisticated energy information on their purchases and activities; in general the policy emphasis has been on providing simple information. However some of this simple information may not be very helpful.

higher demand for said energy services. This "causation" is indeed possible but is difficult to prove, and rests on the tenuous assumptions that energy services are independent of technical systems and that modeling *ceteris paribus* is by nature quite accurate. A more illuminating approach would include attention to what predictive modeling might misrepresent or how well technological change was implemented. In turn the problem is no longer just correcting what people are doing or ignoring it as a matter of nature, but developing more constructive applications of modeling.

5. *The quality of statistical and data analyses in energy efficiency is often lax.* Large data sets and analysis tools are now widely available, and figures, regressions, p-values, and tests named for obscure statisticians can be generated with little training. Big promises are being made for big data. Doing high-quality data analysis is difficult and requires education, experience, good quality tools, and institutional and managerial appreciation of this expertise. At present, the state of this expertise, for high quality policy-facing statistical analysis, seems poor in the United States. In combination with often intense pressures to tell a good and simple story rather than muddy waters, and the limited quality and availability of data anyway, many of the numbers now used may be suspect. Put another away, carefully controlled physical experiments aside, it is not in the nature of statistical analysis to tell simple unequivocal stories involving certainty, causation, and uniformity, while it is in the nature of policy-support funding, reporting, and advocacy to tell them. There are limits to how well statistical vagaries can be made intelligible to non-statisticians.

   One way to improve this situation would be to deliberately increase the presence of statisticians and well-trained data analysts in the energy efficiency workforce. Another (more difficult) improvement would be to encourage debate about the quality of statistical expressions and alternative interpretations. Both of these changes could help strengthen the translation of raw data and publicly-expressed "facts." Here more statistical leadership and some relatively open study about how various institutions currently manage quantification and its interpretation would be most welcome. Such work could bring insights both in statistical and social scientific practices as well as in the improved acknowledgment and management of uncertainty and careful realism as to data limitations (see Morgan et al. 2009) which are now largely ignored. Finally, good statistical analysis can help identify interesting instances of variability, and attention to these cases may provide insight into what can be learned and uses from these non-average situations.

6. *Value complex knowledge and investigation that is not quantifiable or has not been quantified.* Some of the best statisticians who have done applied work have stressed that numbers and common statistical analyses cannot see everything, and in fact can highly misleading especially when they are not accompanied by careful on-the-ground investigation and analysis (Freedman 1991, 1999, 2009). An insistence on quantification (e.g., as required to defend funding, or to legitimize more qualitative observations) can lock out attention to this kind of knowledge. It is commonly said that "you can't manage what you can't measure." This scientific proverb is often attributed to the statistician and quality management pioneer W. Edwards Deming, who in fact said the opposite: that it's a mistake to think that everything important can be measured (Deming 2000). In the field of energy sustainability indicators, one way around this has been to combine indicators with "stories" – narratives and explanations from experts in the field that explain what's behind the indicator – what it shows and what it misses (HELIO International, Williamson et al. 2009). This constructive approach is quite different from reading an indicator in its simple form and may be enough to seed better approaches.

Final Words

There is room for much more attention to practices of quantification and story-telling in energy analysis, research, and policy, and to find ways that quantitative and qualitative information might complement each other to raise the game of energy policy making to better suit the challenges of global climate change. Drawing across the observations made in the previous section, here is the elevator pitch:

- Numbers about energy use may be far less reliable and more labile than they are generally considered to be.
- We need much more attention to how numbers are generated, what they can mean, and what they miss. This will require leadership and perhaps institutional change. It will also require better recognition of the limits of energy use quantification, given the complexity, diversity, and dynamism of the world it tries to capture, and the nature of data available, which is selective, uncertain, and often atemporal.
- We also need more attention to the relationship between research and policy, and the two way bridge between them. Developing appropriately complex stories to accompany numbers may aid communication between these two realms and ultimately the effectiveness of both.

These issues can be studied as matters of organisational and industry practice, and are ripe for open discussion.

References

Anable, Jillian, Ben Anderson, Elizabeth Shove, and Jacopo Torriti. "Categories, concepts and units: representing energy demand in and through time." 2014. http://www.demand.ac.uk/29/07/2014/wp-3-categories-concepts-and-units/ Accessed January 2015.

Anderson, Margo J. *The American census: a social history*. Yale University Press. 1988.

Baskin, Ken. "Storied spaces: The human equivalent of complex adaptive systems." *Emergence: complexity and organization* 10(2): 1–12. 2008.

Brown, Michael F. "A tale of three buildings: Certifying virtue in the new moral economy." *American Ethnologist* 37 (4): 741–752. 2010.

Brown, Rebecca. *Modeled vs.actual energy savings for Energy Upgrade California Homes.* PIER Project PIR-08_108, Working Paper 3.1 2012.

Brown, Karl and Edward Arens."Broken information feedback loops prevent good building energy performance – integrated technological and sociological fixes are needed." In *Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings. American Council for an Energy Efficient Economy.*

Callon, Michel. "An essay on framing and overflowing: economic externalities revisited by sociology." In *The Laws of the Market*, Michael Callon, ed. Oxford, UK. Blackwell Publishing. 1998.

Deming, William Edwards. *The new economics: for industry, government, education*. MIT press, 2000.

Desrosières, Alain. "How to make things which hold together: social sciences, statistics, and the state," *In Discourses on society; the shaping of the social sciences disciplines,* edited by Peter Wagner, Bjorn Wittrock, and Richard P. Whitley. 1991.

[EIA] U.S. Energy Information Administration. "RECS data show decreased energy consumption per household." RECS 2009. 6 June 2012.

[EIA] U.S. Energy Information Administration. 2014. U.*S. States. State Profiles and Energy Estimates. Table CT4: Residential Sector Energy Consumption Estimates, 1960–2012, United States*. (http://www.eia.gov/state/seds/seds-data-complete.cfm., accessed January 2015).

Dundes, Alan. "From etic to emic units in the structural study of folktales."*Journal of American Folklore* 75 (296): 95–105. 1962.

Eidelson, Roy J. "Complex adaptive systems in the behavioral and social sciences." *Review of General Psychology* 1 (1): 42–71. 1997.

Espeland, Wendy Nelson, and Mitchell L. Stevens. "A sociology of quantification." *European Journal of Sociology* 49 (3): 401–436. 2008.Freedman, David A. "Statistical models and shoe leather." *Sociological Methodology* 21 (1): 201–313. 1991.

Freedman, David A. "Ecological inference and the ecological fallacy." *International Encyclopedia of the Social & Behavioral Sciences* 6: 4027–4030. 1999.

Freedman, David. *Statistical models: theory and practice*. Cambridge University Press, 2009.

Gifford, Henry. "A Better Way to Rate Green Buildings: LEED sets the standard for green buildings, but do green buildings actually save any energy? http://www.energysavingscience.com/articles/henrysarticles/BuildingRatingSystems.pdf. 2008.

Gifford, Henry. "A better way to rate green buildings." EnergySavingsScience.com. 2008.

Hacking, Ian. *Representing and intervening: introductory topics in the philosophy of natural science*. Cambridge: Cambridge University Press. 1983.

Hacking, Ian. *The taming of chance*. Vol. 17. Cambridge University Press, 1990.

Halladay, Martin. 2012. "Energy modelling isn't very accurate." Green Building Advisor. 2012.

Harris, Jeffrey, Rick Diamond Carl Blumstein, Chris Calwell, Maithili Iyer, Christopher Payne, and Hans-Paul Siderius. "Toward a policy of progressive efficiency," In Karen Ehrhardt-Martinez and Skip Laitner, eds., *People-Centered Initiatives for Increasing Energy Savings*. American Council for an Energy Efficient Economy. 2010.

Holling, C. S. "Understanding the complexity of economic, ecological, and social systems." *Ecosystems*, *4* (5), 390–405.

Janda, Kathryn B. and Marina Topouzi. "Closing the loop: using hero stories and learning stories to remake energy policy." In *Proceedings of 2013 eceee Summer Study.* European Council for an Energy Efficient Economy.

Karresand, Helena. "The wrench in the works: household behavior and why energy efficient buildings are not enough." In Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings: American Council for an Energy Efficient Economy. 2012.

Kordjamshidi, Maria, and Steve King. "Overcoming problems in house energy ratings in temperate climates: a proposed new rating framework." *Energy and Buildings* 41 (1): 125–132. 2009.

Labanca, Nicola and Paolo Bertoldi. "First steps towards a deeper understanding of energy efficiency impacts in the age of systems," in the *Proceedings of the 2013 eceee Summer Study on Energy Efficiency*.

Lampland, Martha. "False numbers as formalizing practices." *Social Studies of Science* 40 (3): 377–404. 2010.

Lancaster, R, S. Widder, L. Lutzenhiser, S. Chandra, and M. Moezzi. *Post-retrofit residential assessments.* PNNL-21362. 2012.

Lutzenhiser, Loren. "Through the energy efficiency looking glass." *Energy Research & Social Science* 1 (1): 141–151. 2014.

Lutzenhiser, Loren, Mithra Moezzi, Aaron Ingle, and James Woods. *Advanced Residential Energy and Behavior Analysis Project*. Forthcoming 2015. Prepared for the California Energy Commission, PIER Natural Gas.

Meier, Alan K., and James E. Hill. "Energy test procedures for appliances." *Energy and Buildings* 26 (1): 23–33. 1997.

Moezzi, Mithra and Françoise Bartiaux. 2007. "Liberating energy analysis." In the *Proceedings of the 2007 eceee Summer Study on Energy Efficiency in Buildings.*

Moezzi, Mithra, Christine Hammer, John Goins, Alan Meier, and Loren Lutzenhiser. *Behavioral strategies to bridge the gap between potential and actual savings in commercial buildings.* Prepared for the California Air Resources Board and the California Environmental Protection Agency. 2014.

Morgan, M. Granger, Hadi Dowlatabadi, David Keith, Robert Lempert, Sandra McBride. *Best practice approaches for characterizing, communicating and incorporating scientific uncertainty in climate decision making*. U.S. Climate Change Science Program. Lead Agency: National Oceanic and Atmospheric Administration. DIANE Publishing. 2009.

Newsham, Guy R., Sandra Mancini, and Benjamin J. Birt. "Do LEED-certified buildings save energy? yes, but …" *Energy and Buildings* 41 (8): 897–905. 2009.

Neylan, Julian. "Social policy and the authority of evidence." *Australian Journal of Public Administration* 67 (1): 12–19. 2008.

Parag, Yael, and Kathryn B. Janda. "More than filler: middle actors and socio-technical change in the energy system

from the 'middle-out'." *Energy Research & Social Science* 1 (3): 102–112. 2014.
Patterson, Murray G. "What is energy efficiency?: concepts, indicators and methodological issues." *Energy Policy* 24 (5): 377–390. 1996.
Prahl, Duncan. "Analysis of energy consumption, rating score, and house size." Washington, D.C.: U.S. Green Building Council. http://www.usgbc.org/Docs/Archive/MediaArchive/310_Prahl_PA772.pdf. 2000.
Salais, Robert. "Quantification and the economics of convention." *Historical Social Research/Historische Sozialforschung* 37 (4): 55–63. 2012.
Shove, Elizabeth and Mithra Moezzi. "What do standards standardize?" in *Proceedings of the 2002 ACEEE Summer Study on Energy Efficiency in Buildings.* American Council for an Energy Efficient Economy. Washington, DC. 2002.
Stein, Jeff Ross, and Alan Meier. "Accuracy of home energy rating systems." *Energy* 25 (4) 339–354. 2000.
Timmermans, Stefan, and Steven Epstein. "A world of standards but not a standard world: toward a sociology of standards and standardization." *Annual Review of Sociology* 36: 69–89. 2010.
Turner, Cathy and Mark Frankel. *Energy performance of LEED for new construction buildings*. Vancouver: New Buildings Institute. 2008.
Scofield, John H. "Do LEED-certified buildings save energy? Not really …" *Energy and Buildings* 41 (12): 1386–1390. 2009.
Stone, Deborah A. "Causal stories and the formation of policy agendas. "*Political Science Quarterly* 104 (2): 281–300. 1989.
Strauch, Ralph E. "'Squishy' problems and quantitative methods." *Policy Sciences* 6 (2): 175–184. 1975.
Ueno, Kohta. "Building energy performance metrics" *Building Science Digest* BSD-152. 2010.
White, Leslie A. "Energy and the evolution of culture." *American Anthropologist* 45 (3): 335–356. 1943.
Williamson, Laura E., Hélène Connor, and Mithra Moezzi. *Climate-proofing energy systems*. HELIO International. 2009.
Wilson, Alex, and Jessica Boehland. "Small is beautiful: US house size, resource use, and the environment." *Journal of Industrial Ecology* 9 (1-2): 277–287. 2005.

Acknowledgments

Elements of this paper were derived from work on the AREBA (Advanced Energy Research and Behavior Analysis) project conducted by Portland State University for the California Energy Commission. This project was supported by Public Interest Energy Research (PIER) Project Natural Gas funds. The work also benefitted from participation in a workshop on social practice theory and complex adaptive systems at the Joint Recent Centre in Ispra (February 2015). I also thank eceee Summer Study reviewers and Aaron Ingle, who gave many useful comments on a draft of this paper.