Empirical Validation Workshop Pre-Read

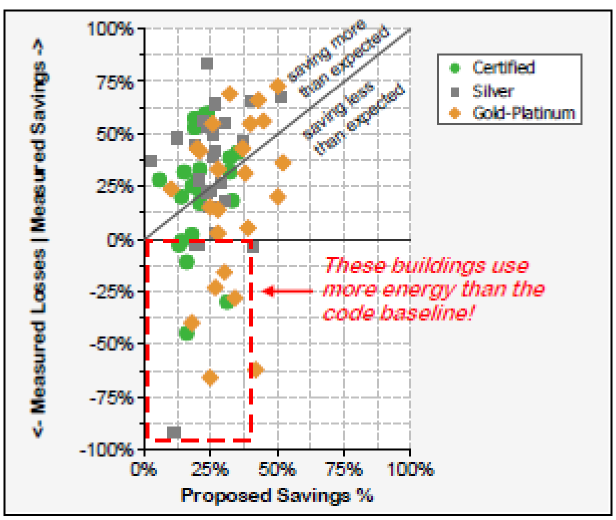
Amir Roth, Ron Judkoff, Philip Haves, Steve Selkowitz

The goal of this workshop is to create a plan for improving the reputation of building energy modeling (BEM) as a scientific and engineering enterprise. Increasing confidence in BEM among BEM practitioners and among those who employ their services—building owners, energy-efficiency program managers, etc., will ultimately expand the (effective) use of BEM in building design and operation.

The George Box quote "all models are wrong, but some are useful" is both apropos and true, but it's a quote only modelers understand and embrace. The constituencies described above often have a difficult time understanding how something wrong can be useful and discount modeling results. However, even though the Federal Low Income Weatherization Program and the California Weatherization Program apply a 0.5 discount factor to modeled savings predictions, they still find modeling useful for customizing retrofit packages for the homes of their clients. Skeptics or those with anti-modeling agendas don't even make it past the "wrong" in the above mentioned quote.

A canonical example of anti-modeling perception is the 2008 paper by Mark Frankel and Cathy Turner of the New Buildings Institute, *How Accurate is Energy Modeling in the Market* (*https://www.aceee.org/files/proceedings/2008/data/papers/3_320.pdf*). Although the median energy use of 90 LEED buildings showed close agreement with modeling, the spread was quite large (Fig. 1). This report cast both energy modeling and LEED in a poor light.

Figure 1. Measured versus Modeled Savings %

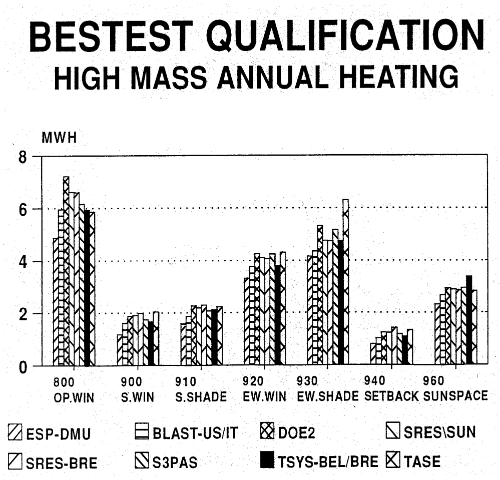


In order to enhance the reputation of modeling, we must change the perception of "wrong" to "mostly right". Preferably, we would be able to quantify and isolate remaining "wrongs"—both to prevent them from contaminating and pulling down the enterprise as a whole and to pinpoint areas of improvement.

A “right” energy model is a predictive one, i.e., one whose predictions align well with measurement. For an energy model to be right, both the calculations and the inputs must be right. However, this workshop will focus entirely on the calculations. This is the more tractable part of the problem and also the more important part in terms of the overarching message. From a tractability standpoint, addressing overall correctness requires decomposition and staging. In such a setup validating calculations assuming correct inputs must precede validation of the inputs themselves; it is not possible to validate inputs assuming correct calculations because that assumption is simply not possible to make. Calculations are also easier to validate than inputs because the set of calculations is smaller and more centralized and because calculations are deterministic and amenable to detailed controlled experimentation whereas inputs are stochastic. From a message standpoint, the calculations are crafted by experts and so they carry a greater expectation of correctness. On the other hand, wrong inputs are either the fault of inexperienced users, or only knowable within a wide uncertainty band. Wrong calculations raise questions about the enterprise as a whole. From a practical standpoint, it may also be more important to validate the calculations. The majority of BEM use cases are set up as model-to-model comparisons—this is what makes BEM “useful” despite being “wrong”—and this is done at least in part to control for uncertain inputs. It is less clear that model-to-model comparisons can control for the effects of wrong calculations. Although important in their own right, model inputs and their correctness—and calibration—are excluded from the proceedings.

There are a number of complementary ways to test and validate model calculations. Analytical tests cover scenarios for which there are theoretical, closed form solutions—these of course assume that the theory is correct. Empirical tests compare model results to measured data. Comparative tests compare model results to the results of other models, presumably ones that have passed the analytical and empirical tests. Comparative tests have significant value in that they cover arbitrarily more ground than analytical tests, are more cost-effective to conduct in large numbers than empirical tests, and can also be more diagnostic than empirical tests. Comparative tests exploit the “Anna Karenina principle” that there is only one right answer but many wrong ones—if a large number of different calculation methods and implementations agree, the answer is likely to be right. ASHRAE Standard 140 combines these three kinds of tests to create a robust framework. As with modeling as a whole, however, only modelers understand and appreciate the Standard 140 structure and methodology. The fact that the standard consists largely of comparative tests bothers some constituencies and feeds the perception that modeling is self-referential with only tenuous connection to the real world. The only tests non-modelers view as legitimate are empirical tests. And it is easy to argue that ASHRAE Standard 140 would benefit from a healthy infusion of empirical test cases—not only to quiet skeptics, but to adjudicate the many simple, well-defined comparative test cases in which existing engines often disagree by 25-30% (Fig. 2). This workshop will focus on validation via empirical testing. A goal of the workshop is to propose a set of important empirical tests that, once conducted, would be codified into the standard.

Fig 2.



A focus on empirical testing is timely because of the recent availability of well-characterized, highly-instrumented test facilities like LBNL's FLEXLAB (Facility for Low-Energy eXperiments) and, potentially, ORNL's FRP (Flexible Research Platform) and others around the country and the world. FLEXLAB, for one, was explicitly designed and constructed with empirical energy model validation in mind. Another goal of this workshop is to learn about the capabilities of FLEXLAB and FRP and which experiments they can support.

A final goal of this workshop deals with communications and marketing. Our first job is to convince ourselves, the modeling community, of the validity of our tools and the soundness of our methods. The proposed, and ultimately executed, empirical validation tests will be driven by modelers and their needs. However, it is also necessary to properly project, package, and message this confidence to non-modelers.

The following are questions that will be discussed during the workshop. It will be useful to think about these and come primed with answers. We would like each participant to prepare a ½ to 1 pager considering these questions. Thanks in advance for participating in the workshop.

- What are the current and future types of simulation analysis for which there is concern about accuracy and robustness? One potential entry here is modeling of old and inefficient existing buildings whose particular inefficiencies—empty wall cavities, old degraded equipment, etc.—are a poor match for the modeling assumptions made by the current generation of building energy simulation tools which target newer buildings.

- For which systems and components is there real cause for concern, or lack of confidence, regarding the accuracy of current models and simulation results? For instance, modelers may lack confidence in models for new, low-energy systems including radiant systems, underfloor-air distribution, natural ventilation, etc.

- What constitutes “success” for a validation exercise? 5%? 10%? 25%? What output metrics determine success and at what spatial and temporal granularity?
- What facilities exist for controlled empirical validation (US and Worldwide)?
- What high quality definitive studies have been done (US and Worldwide)?