Accuracy of the Home Energy Saver Energy Calculation Methodology

Danny Parker, Florida Solar Energy Center
Evan Mills, Leo Rainer, Norm Bourassa, and Greg Homan, Lawrence Berkeley National Lab

ABSTRACT

The Home Energy Saver (HES) suite offers popular online simulation tools that enable U.S. homeowners and energy professionals to rigorously evaluate home energy use and develop recommendations on how energy can be saved across all end uses. The underlying analytical system is also available as a web service to power third-party energy analysis tools. Given the system's diverse uses, it is important that the simulation is robust and intrinsically accurate. While the engineering methods and assumptions are extensively documented and subjected to peer review, it is useful to evaluate how well HES predicts energy use in occupied homes. In this paper we compare measured to predicted energy use for 428 occupied homes in Oregon, Florida, and Wisconsin, representing a diversity building types, energy intensities, and occupant behaviors. We show how audit depth, knowledge of operational details, and submetered energy data can be valuable to the process of improving model accuracy—particularly for individual households, where energy use can vary three-fold for homes with virtually identical physical characteristics. Accuracy is strongly proportional to the quality and completeness of inputs, yet audit data are often deficient. Predictions are best—and the tendency of models to over-predict actual consumption is mitigated—when behavioral inputs match actual conditions. We find that Averaged across groups of homes, HES predicts energy use within 1% of actual consumption when physical characteristics and occupant behavior are well accounted for. New research findings are conferring even greater accuracy as they are incorporated into simulation tools.

Introduction

Energy analysis tools are integral to the process of identifying and implementing building energy savings measures. Modeling applications can vary from the fine-grain component- or end-use level to the whole building. User groups include homeowners and renters designers, auditors, home performance contractors, and policy analysts.

The Home Energy Saver (HES) web-based simulation tools provide these diverse audiences with simple, non-proprietary ways of employing state-of-the-art residential energy calculation tools and energy data to support decision-making. The tools integrate a variety of best-practice models, algorithms, and data sources assembled over several decades at Lawrence Berkeley National Laboratory, other DOE National Labs, utilities, and elsewhere within the energy efficiency and energy services communities. Historically, the use of such tools had required greater expertise and knowledge of energy and building technologies and computing (as well as more powerful computers) than possessed by some target audiences. These barriers have been gradually overcome by making these tools available via the internet, thus eliminating the need to run software on the user's local computer; providing user-friendly interfaces; and incorporating extensive "smart" defaults.

Development of HES began in 1994, and the first web-based version of the tool was released in 1996 (Mills 1997). In 2009, the addition of web services enabled third-party software developers to incorporate the underlying models and data into their own user interfaces (Mills and Mathew 2012). The engineering methods and assumptions behind the hourly DOE-2.1E simulation engine and supplementary non-HVAC analysis methods used by HES are extensively documented in the public domain and subjected to peer review (Mills *et al.*, 2007). The HES suite has been expanded to compute asset ratings underpinning DOE's Home Energy Score Program (Bourassa *et al.*, 2012), and multifamily models have been added as well (Mills and Mathew 2012).

It is important that these energy simulations accurately capture the energy use attributed both to the building construction and to occupant behavior. We assessed the HES Consumer and Professional tools (hes.lbl.gov and hespro.lbl.gov), which use identical methods and are jointly referred to in this paper as "HES." Roberts *et al.*, (2012) separately assessed the Home Energy Scoring Tool (homeenergyscore.lbl.gov), also summarized in Bourassa *et al.*, (2012). We illustrate how these validation exercises support continuous refinement of the underlying calculations, and summarize important findings in buildings energy research that should confer even greater confidence in model predictions as refinements are made to calculations procedures.

Elusive Accuracy

The accuracy of energy analysis tools cannot be taken for granted. A comparison of a range of web- and disk-based tools found a three-fold variation in predictions for a given home (Mills 2003). Assessing the accuracy of tools designed for occupied buildings (such as HES or EnergyGauge USA) poses far greater challenges than the case where one model is compared to another using hypothetical buildings with stipulated occupancy and operational conditions (e.g., BESTEST). Analysts have long called for more validation of audit tools against measured data (Pigg 2001). Indeed, accuracy assessments should be an integral part of tool development. For tools that are proprietary, third-party analysts cannot easily understand the underlying causes of observed inaccuracies.

The notion of energy software "accuracy" is deceptively simple. However, in practice the building energy modeling community has yet to develop and agree on a robust definition, let alone methods for identifying and attributing the causes and sources of inaccuracies. In sum, "accuracy science" as applied to energy models of occupied buildings is in its infancy.

A multitude of potential sources of perceived inaccuracy occur, ranging from the establishment of accurate "ground-truth" measured energy use for comparison, successfully collecting audit data and extracting model inputs, and in the intrinsic modeling process itself. Moreover, while it is certainly possible for an energy analysis tool to produce exact agreement with measured data, it can be challenging to determine if the result emerges inadvertently due to fortuitous offsetting errors. And, fixing one of two equal offsetting errors will (temporarily) worsen overall accuracy. The number of variables to be considered is daunting.

Various approaches to validating building energy models have been developed. We focus here on how well a model—given "accurate" and reasonably complete inputs—predicts the measured energy use of occupied homes. This might be referred to as "intrinsic accuracy". We thus attempt to control for errors caused by factors outside the modeling process.

Irrespective of approach, many sorts of "noise" can interfere in accuracy assessments, undercutting usability of the findings. The following questions must thus be considered before a meaningful accuracy assessment can be conducted:

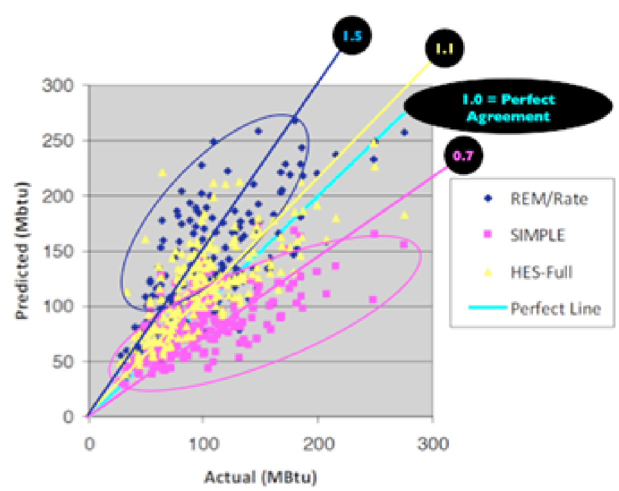
- **What is the intended use of the accuracy assessment?** Accuracy assessments can be conducted on a one-time basis, but are most useful when applied in an iterative and diagnostic fashion to inform software development (Polly *et al.*, 2011). The latter requires high-fidelity data and forensic identification of the underlying reasons for inaccuracies. Accuracy assessments may be narrowly focused on heating and cooling loads under highly standardized operating conditions (e.g., fixed thermostat settings) or broadly defined as whole-building final energy use, including a broader range of physical and operational influences.
- **How is accuracy defined?** A recent study identified 10 metrics (Polly *et al.*, 2011). Accuracy can be expressed in terms of absolute error from the reference point, a proportional error, or a statistical deviation, and applied at the whole-building or end-use level. A given absolute error for a low-energy home has a larger proportionate effect than the same error on an energy-intensive home. Appropriate metrics should be employed to describe the central tendencies, while not obscuring the broader patterns of results across a range of conditions across large cohorts of homes being assessed.
- **What level of precision and accuracy is required for the assessment at hand?** The level of "acceptable" accuracy depends on the intended use of the modeling tool. A tool providing qualitative recommendations versus absolute savings estimates requires less precision and accuracy than one for investment-grade audits. An asset rating and accompanying 10-step score (Roberts *et al.*, 2012), for example, requires less precision than an estimate of end-use-level energy consumption under specific operational conditions.
- **How inclusive is the assessment?** A tool may be accurate in one climate or type of building, but not another. Combining a variety of building types, equipment, fuels, operating conditions, and geographies can require a large number of parametric scenarios. Tools that estimate costs or carbon emissions must account for factors such as hourly load shapes, tariffs, and fuel mix, while those simply estimating equipment loads do not need to do so.
- **How are the home characteristics and 'ground truth" energy use defined and applied**? A model can be compared to other models or to actual buildings, the former at best providing only a pseudo-estimate of accuracy. Just as "bad data" can confound an analysis of billing information, so too can "bad inputs" confound simulation analyses. The same weather records should be used to normalize bills and to underpin simulations. Great care must be taken to understand the implications of comparing billing data (which incorporates behavioral influences) to the results of asset-based simulations (based solely on physical building attributes) that standardize occupant behavior. Rigorous quality control of input data is required to minimize subjective inputs and mischaracterization of the actual home and its use.
- **What types of errors are sought, and how are they to be interpreted?** Sources of errors can include software defects, inaccurate engineering algorithms, non-representative default values or weather data, and errors or gaps in data-collection or user input (Baden 2009). Constrained tool input options, e.g., temperature-bands rather than exact choices for water-heater set-points, can arbitrarily result in more accurate outcomes if the device in question happens to be set at the center of a band than if it is near the edges. Lack of field data on the subject homes (e.g.,

whether or not basements are conditioned) can necessitate the use of default assumptions that are not as "accurate" as would ideally be the case. While stipulating fixed consumption for certain devices or entire end uses in a model (e.g., lighting) may be necessary, it is important that such practices be isolated as reasons for differences between measured and estimated use.

- **Can multiple tools be properly compared to one another?** It is challenging to apply similar inputs into dissimilar tools (or translate inputs collected for one tool another). For example, one tool may offer three choices for insulation levels (e.g. none, moderate, or high) while another accepts an exact R-value. This practice imposes severe limitations on tool comparisons.
- **How can inaccuracies *not* associated with the software be isolated?** Tool developers are generally most interested in the *intrinsic accuracy* of their models, i.e., with fully accurate inputs, and that is the focus of this paper. There are numerous sources of error outside of the modeling process. These include, but are certainly not limited to: actual versus occupant reported thermostat settings, unknown defects in home workmanship, imperfections in weather-normalization techniques applied to the raw measured data, mismatches between weather stations and the home location and microclimate, differences between equipment nameplate and labeled energy use and actual in-situ performance, reliance on default values, and errors in field data collection. In our first data cohort, below, ten homes had identical refrigerators, one of which used 50% (400 kWh/year) more energy than the average of the others due to a door gasket had been torn upon installation. Errors in data entry can also be important. Chapman (1991) observed prediction errors up to 10% due to errors while entering data into the models. Others found variations in results of up to 7.5% even for asset ratings of identical homes (Pigg 2001), a value that could be higher for operational assessments.

Many of the aforementioned methodological hazards are illustrated in a widely cited assessment of three disparate tools: Home Energy Saver, REM/Rate, and SIMPLE (Earth Advantage 2008; Baden 2009). Extensive but sparsely documented reliance on defaults rather than setting inputs to actual known conditions, and use of different weather records for bill normalization and simulation confounded comparisons to actual consumption and likely predetermined the conclusion that increasing the number of inputs did not improve accuracy. Heavy emphasis on the absolute values of errors and consolidated average outcomes obscured distinct differences in the accuracy of the tools: for high-use homes: REM/Rate systematically over-predicted, SIMPLE systematically under-predicted, while HES displayed symmetry around the expected values and lower absolute errors across much of the domain (Figure 1). This and a derivative evaluation (Energy Trust of Oregon 2012) used a vintage-2008 version of HES.

Figure 1. Predicted Annual Energy Use for Three Tools



While not provided in the original study, a visual examination suggests that the slopes of regression lines (added) would be approximately 0.7, 1.5 and 1.1. A slope of 1.0 represents perfect agreement, i.e. where Actual = Predicted (Earth Advantage Institute, 2008).

Comparisons of Model Estimates to Measured Data from Actual Homes

There has been a prevailing view that building energy simulations tend to overestimate consumption in real homes (Polly *et al.*, 2011; Energy Center of Wisconsin 2000). This is important context for the question of the accuracy of HES, and how simulations can be improved by comparing their results to measured data.

The four independent cohorts of field data evaluated here include occupied homes across a diversity of locations (Table 1). The datasets have in common high-quality audits with detailed information on physical characteristics and actual operational conditions for many of the homes. Because our goal is to assess tool accuracy given unambiguous and "accurate" inputs, homes were eliminated from the sample if they contained incomplete or suspect data or major miscellaneous uses (e.g., unmetered solar, swimming pools, portable heaters) not adequately characterized to allow modeling. Homes with minor supplemental wood heating were retained, but those with unmeasured primary wood heat were excluded. We then applied data-completeness filters, ensuring that key inputs were required to be present for a home to be included (LBNL 2012a). This quality assurance process resulted in the elimination of 232 homes from the raw sample of 660 homes received by LBNL.

We then exercised the HES model using four increasingly complete sets of input values (Table 1) in order to systematically identify how the type and completeness of inputs affects accuracy. The key distinction is between Asset analyses ("Rate the home, not the occupant") and Operational analyses, where the physical characteristics of fixed assets and occupant effects plus those of lighting and small miscellaneous appliances are comprehensively considered. The most complete scenario (Operational) required 31 model inputs, with an additional 27 optional inputs. Note that this is far lower than the ~200 possible inputs to HES, and thus does not represent a "best-case" for the accuracy of model estimates.

Table 1. Definitions and Constraints of the Four Modeling Scenarios

Scenario	Description	Required inputs	Optional inputs	Total inputs	Number of occupants*	Behavioral inputs**	Lighting & misc. appliances***
Defaults	Only location is provided (based on ZIP code to assign weather tape)	1	0	1	no	no	defaults
Asset::Visual	Non-intrusive, non-instrumented audit	18	9	27	no	no	defaults
Asset::Full	Instrumented audit: more equipment and envelope characteristics captured	26	16	42	no	no	defaults
Operational	Asset::Full scenario + behavioral inputs	28	29	57	yes	yes	audit data

* This primarily influences the hot-water use calculations.

** Thermostat setting, hot water temperature setting, loads of clothes washed, loads of clothes dried, cooking hours (stove and oven), zone heating/cooling.

***The Operational scenarios used miscellaneous equipment counts, wattage, and utilization specified in the original audits (or HES defaults where site-specific data were not available). Lighting is modeled as proportional to floor area in the Asset and Operational scenarios, per the Home Energy Scoring tool methodology.

For the definitions of optional and required model inputs for each Scenario, see: https://sites.google.com/a/lbl.gov/hes-public/accuracy/decision-rules

For the two Florida cohorts, input data were obtained from the original audit forms and derivative reports. For the Wisconsin and Oregon cohorts, field data were translated into HES inputs from REM/Rate model inputs in the National Renewable Energy Laboratory's Field Data Repository (Roberts *et al.*, 2012). Measured energy use is weather-normalized in all cases with the same weather data employed in the simulation models.

While our goal is to identify the intrinsic accuracy of the tool, free from noise caused by inaccurate or incomplete inputs, many uncontrolled factors remain. Judgment and approximations used in translating REM/Rate inputs to HES inputs unavoidably introduced error compared to a case where audits gathered inputs expressly for HES. Audit data for lighting and miscellaneous end-use characteristics were limited, which necessitated a degree of reliance on the significant approximation embodied in default values.

Results

Increasing the number of relevant inputs clearly improves Home Energy Saver's predictive power for both electricity and fuel as indicated by declining average error and reduced scatter around the prediction indicated by the CV statistic (Table 2 and Figure 2a-b).

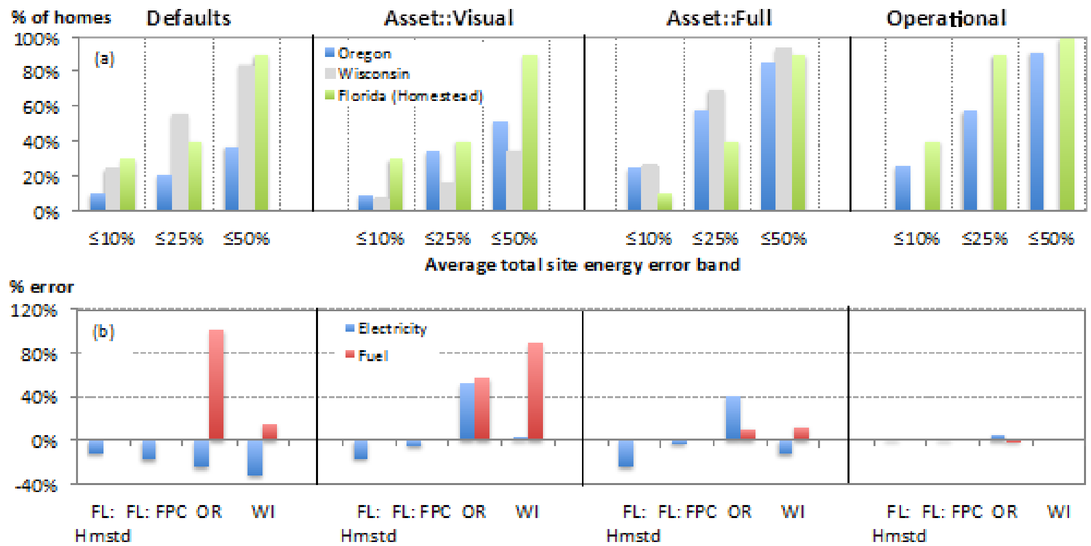
Table 2. Sample Characteristics and HES Summary Results for the Four Cohorts of Homes

	Florida: Homestead		Florida: FPC		Wisconsin		Oregon		ALL
Sample									
Number of auditors	1		4		15		5		25
Representative sample	no		yes		yes		no		
Sample has submetered end-use energy data	yes		yes		no		no		
Original sample size	10		205		299		302		816
Initial sample received by LBNL	10		171		299		180		660
Sample after quality control (Default case)	10		171		139		108		428
Results									
Measured consumption (site MBTU/year)	**53**	CV*	**54**	CV	**136**	CV	**97**	CV	
Prediction errors (%)									
Defaults	**-15%**	31%	**-19%**	n/a	**4%**	31%	**66%**	49%	
Asset::Visual	**-17%**	32%	**-7%**	n/a	**68%**	38%	**56%**	38%	
Asset::Full	**-25%**	30%	**-5%**	n/a	**7%**	27%	**19%**	31%	
Operational	**0.5%**	14%	**1.3%**	n/a	**na****	n/a	**-0.4%**	31%	**0.6%**

* CV (coefficient of variation) is the ratio of the standard deviation of the prediction errors to average measured consumption.

** Lack of sufficiently complete audit data made it not possible to complete the Wisconsin Operational scenario.

Figure 2. HES Accuracy by Inputs (a) and Fuel (b)



Cohort 1: Intensively Metered and Highly Similar Low-income Homes in Florida. The highest-fidelity audit data that we identified are for ten all-electric homes near Homestead, Florida. The characteristics of these low-income, Habitat for Humanity homes are described in an earlier publication (Parker *et al.*, 1996) and supplementary data (LBNL 2012b). We compared the estimates of both HES and Energy Gauge USA (EG) (Parker *et al.*, 1999) to this field data. Both tools use the well-documented and well-validated DOE-2.1E simulation engine (Birdsall *et al.*, 1990) to estimate space-conditioning energy, and different methods for other end uses.

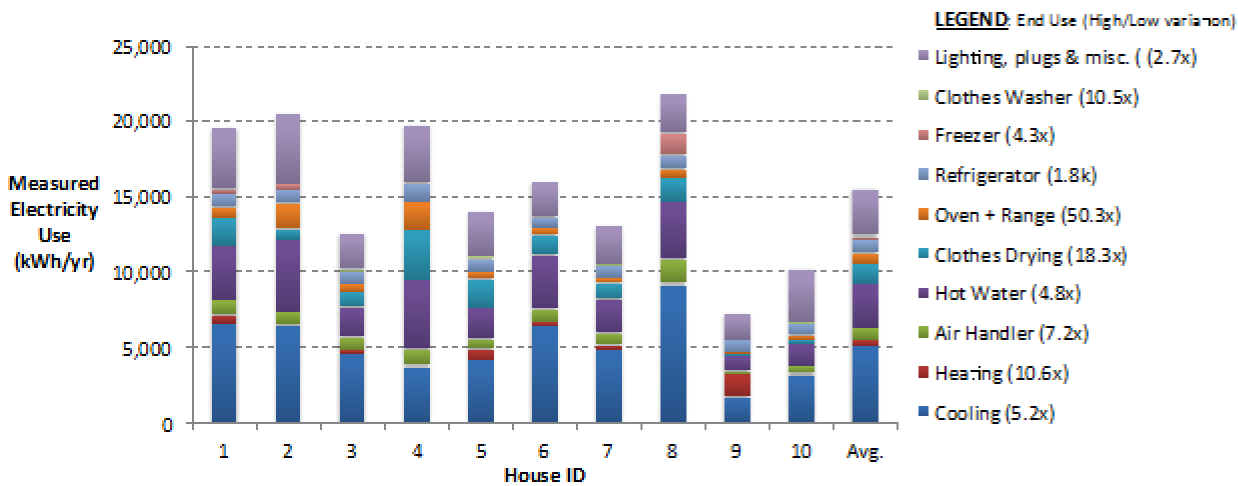
While the influences of occupant behavior have been observed by building energy researchers for more than three decades (Sonderegger 1978), the Homestead dataset is notable in showcasing how energy use can vary substantially among virtually identical homes, in this case all constructed in 1993 by the same builder. Three are of a simple rectangular design (1250 ft^2 of conditioned floor area) with four bedrooms. The other seven are three-bedroom homes, five feet shorter in the long axis (1100 ft^2). All have precisely the same heating and cooling systems (SEER 12 two-ton systems with 5 kW of electric strip heat) and the same windows. Despite these similarities, variance in workmanship had implications for energy model results, as evidenced by the observed variance in air leakage from 1085 to 2257 CFM at 50 pascals pressurization (108%). The homes were identical with respect to major installed appliances (refrigerator, washer, dryer, electric water heater). Three homes added freezers, which were also sub-metered. In addition to the physical similarities of the homes, their location on the same neighborhood block (microclimate) makes them an ideal test case for occupant-related impacts.

Year-long sub-metering at the end-use level (15-minute data) provides an opportunity to quantify behavioral drivers with great precision, understand variance and outliers, and pinpoint sources of inaccuracies. The homes exhibit a 3-fold variation in measured energy use. Moreover, the highest (20,452 kWh/year) and lowest cases (7,257 kWh/year) were both three-bedrooms and units, with identical floor area. Variations at the end-use level were often much higher (Figure 3). House 2 used no heating, while House 9 used 1,467 kWh over the year. Interestingly, House 9 was the least energy-intensive home in other respects.

HES and EG each predicted average whole-house energy use for the cohort within one percent of the measured values (Figure 4). As seen in Figure 5, all but one of the HES predictions for individual homes are within 25% of measured annual energy use (in fact, within 17%), and five are within 10%. Early results helped identify and improve out-of-date default values for water heating, and associated appliances. Detailed sub-metered energy data enabled us to identify a calculation error for air-handling units, which improved accuracy from a 75% under-prediction of that device's energy use to a 1% over-prediction.

Asset scenarios estimated the average energy use of the set of homes within 25%, while lacking any explanatory power for individual virtually identical homes. The estimates average low because they do not consider the relatively high occupancy rates of low-income housing and associated influential behavioral factors. That said, for highly uniform houses such as these, adding even more details on the physical characteristics yields diminishing returns. On the other hand, as seen in Figures 4 and 5b, by incorporating the operational effects of occupant behavior (e.g., thermostat management and appliance utilization) predictive power is enhanced to the point of essentially perfect on-average agreement—at the whole-building level and by end-use.

Figure 3. Measured Total and End-use Annual Energy for the Homestead Cohort

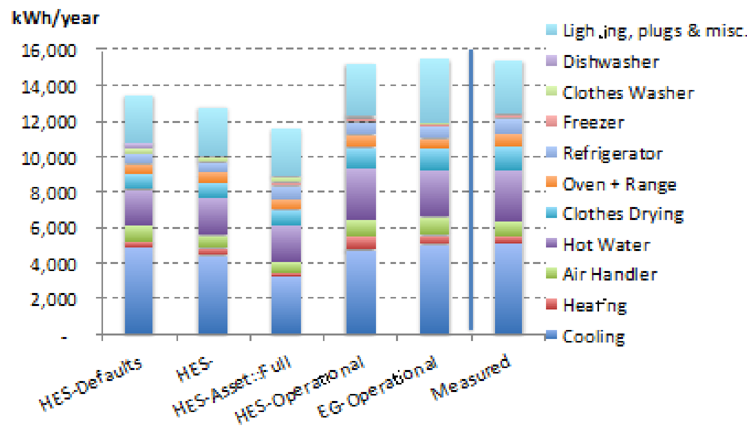


Behavioral Drivers											
Number of occupants	8	6	3	8	3	5	4	5	3	3	4.8
Avg. daytime TSTAT (cooling, F)*	75	77	77	74	74	84	81	76	83	80	78
Avg. daytime TSTAT (heating, F)*	63	71	72	72	67	76	70	72	72	75	71
DHW temperature (F)	123	123	126	119	122	126	122	127	121	122	123
Hot water use (g/day)	82	113	39	110	56	82	48	60	18	28	64
Laundry loads washed per week)	5	5	6	6	9	3	3	8	15	12	7.2
Laundry loads dried (per week)	5	5	6	6	9	1	3	8	15	12	7.0
Stove (hrs/day)	3	0.5	1	1	1	0.25	0.5	1.5	5	0.5	1.4
Oven (hrs/week)	3	1	2	1	3	0	1	2	2	1	1.6
Televisions	5	1	3	1	1	4	1	1	1	1	1.9

* Temperatures measured at the thermostat, averaged over season

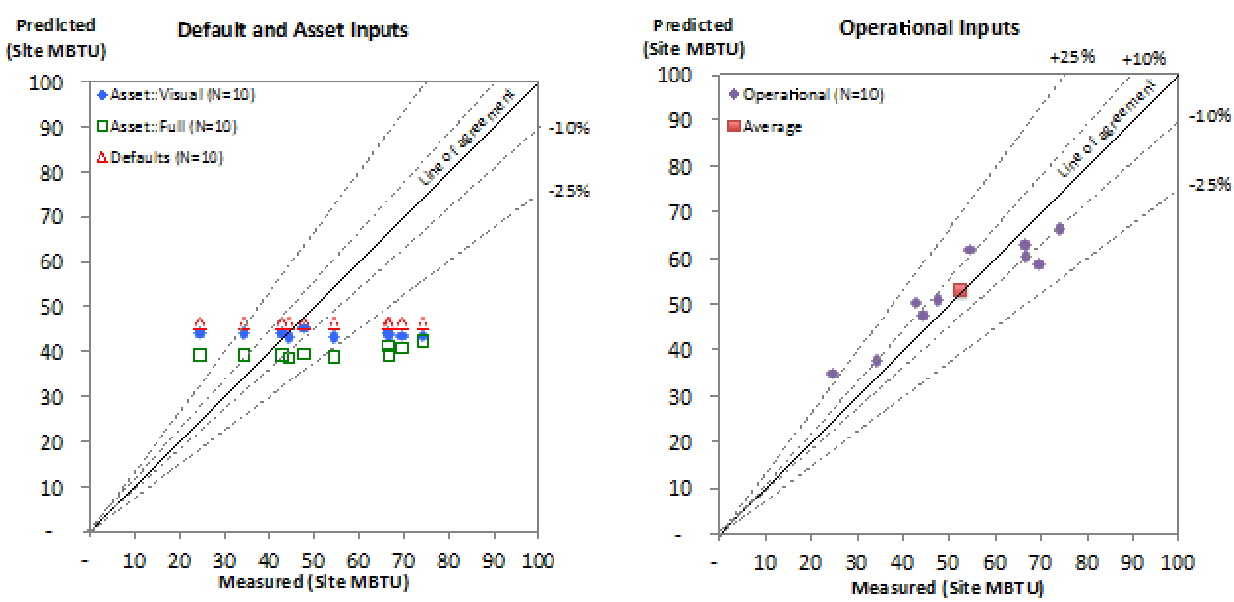
Compensating errors can become visible when end-use data are examined. While the "Defaults" case initially appears to produce better predictions than the "Asset" cases, it achieves this only fortuitously: the fully defaulted house is more than 50% larger than the subject (low-income) houses, while assuming a more efficient AC than in the actual homes. The default assumptions omit an end use (freezers) that is present in some of the homes, while including another that is not present (dishwashers).

Figure 4. Average Measured and HES-predicted Annual End-use Breakouts: Homestead Cohort



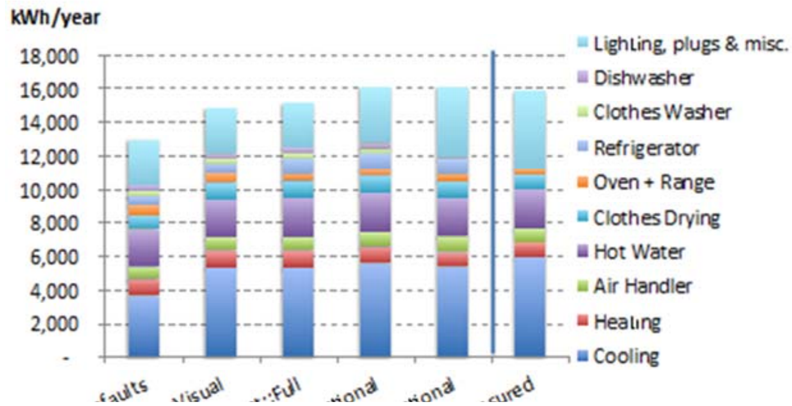
The outliers in Figure 5 stem from pronounced differences in home operation. For instance one high-use home had problems both with the refrigerator and air conditioning systems and the lowest-use house was often vacant while the single mother sometimes stayed with her children at her mother's home, as reflected in the measured hot water use (18 gal/day, ~1/3 of the cohort average). Hot water use data are generally not available to auditors, and is not an input for HES, although we could identify the reasons for deviations in prediction.

Figure 5a-b. Measured vs. HES-predicted Annual Energy Use: Homestead Cohort



Cohort 2: Aggregate of a Large Representative Sample of Homes: Central Florida. We compared predictions from HES and EG against measured energy use and detailed field audit data for a large, statistically drawn sample of 171 all-electric homes in Central Florida (Parker 2002). In each home, 15-minute electric demand data was obtained for total power, space heating, cooling, water heating, refrigerators, laundry, dishwashing, and cooking. Interior and exterior temperatures were also recorded. Figure 6 illustrates how the tools predicted total and measured end-uses when applied to a "composite" home (individual HES runs were not available) that was typical of the audited predominant characteristics in the monitored sample (rather than the average of characteristics for each home). The relative similarity in outcomes between Asset and Operational cases is expected, as these are averages across a large number of homes. Accuracy was, however, highest for the Operational case.

Figure 6. Average Measured and Predicted Annual End-use Breakouts: Central Florida Cohort

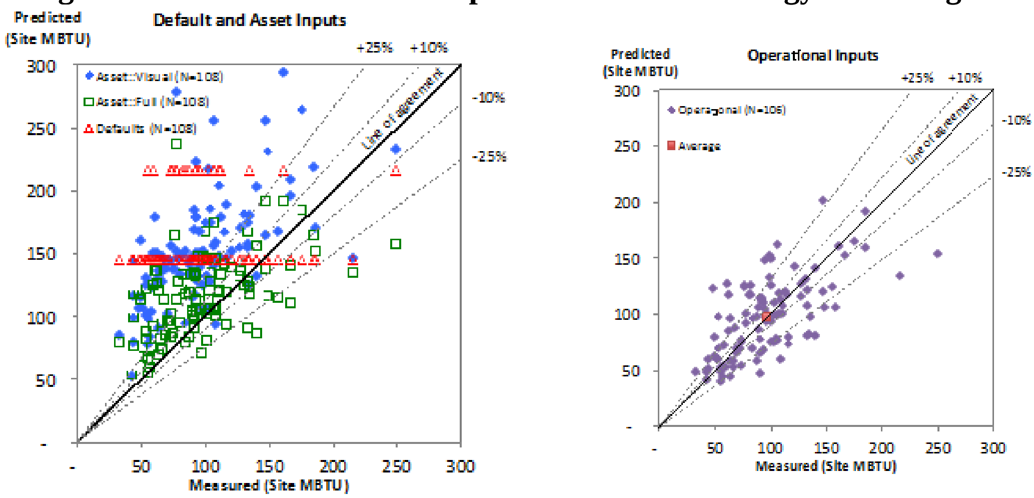


Refrigerators, dishwashers, and clothes washers were not individually submetered and are included under the "Lighting, plugs, and misc." category for the Measured case; thus shadings are not comparable with the simulation cases for those categories

Both the HES and EG models exhibit excellent operational correspondence (within about 1% of actual measurements) to the average measured total energy use as well as end-use level detail. The models did not adjust for an approximately 2% seasonal vacancy rate in the cohort among "snowbird" occupants with primary homes in the north. Thus vacancy rates may be another operational feature to be accounted for within future modeling of large-scale samples.

Cohort 3: A Non-Random Sample of Homes from Portland and Bend, Oregon. The third cohort is drawn from a set of detailed audits for 180 mixed-fuel homes in Oregon, the same dataset used in the Energy Performance Score study noted previously (Earth Advantage 2008). Detailed audits were performed for each home using the RESNET rating protocol. The sample was not representative of the population and did not provide end-use sub-metering, but did include a wide range of house ages, multiple heating fuels, and two localities within the state (Portland and Bend areas). Accounting for operational factors minimized scatter and improved predictions for specific houses, but it was not possible to convert REM/Rate lighting data to HES inputs, necessitating reliance on defaults. The Default and Asset cases resulted in substantial overestimates of actual energy use. Full operational inputs yielded average estimates within 1% of actual consumption across the cohort (Figure 7), with superior fit to that shown in Figure 3.

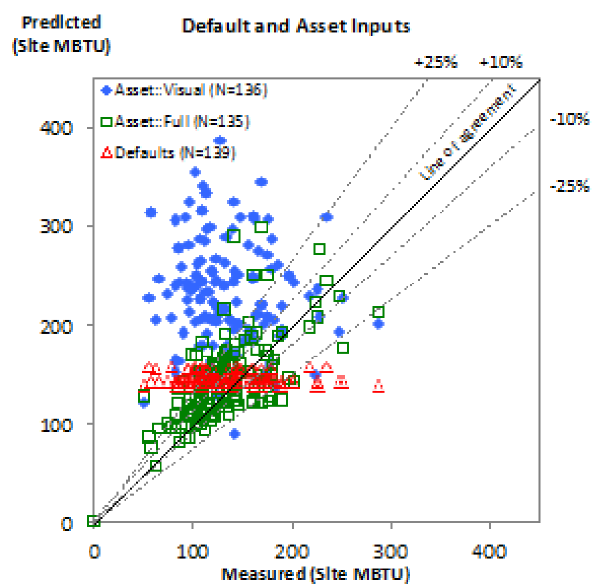
Figure 7a-b. Measured vs. HES-predicted Annual Energy Use: Oregon Cohort



Horizontal bands for the Defaults case represent two disparate weather locations, the only input variable.

Cohort 4: A Representative Sample of Owner-occupied Homes from across Wisconsin. The fourth cohort is drawn from a collection of 299 HERS audits conducted using REM/Rate in Wisconsin in 1998 on a random sample of mixed-fuel, single-family, owner-occupied homes ranging from "mobile homes to mansions" (Energy Center of Wisconsin 2000; Pigg 2001). The original analysts found that heating energy use was over-predicted, attributed in part to heavy reliance on 68°F default thermostat settings. Behavioral factors were identified as key drivers of energy use, as indicated by user-reported winter thermostat settings in the homes ranging from 59 to 74°F, and loads of laundry washed ranged from 1 to 15 loads per week per household (3 and 8 occupants, respectively), zero to half of which utilize a hot-water cycle.

Figure 8. Measured vs. HES-predicted Defaults and Asset Annual Energy Use: Wisconsin Cohort



The Defaults case shows negligible predictive power for individual homes, but good on average (Figure 8), although errors for each fuel (-34% electricity and +16% for gas) fortuitously offset one another. The Asset::Visual case predicts high. Increasing the number of physical characteristics inputs yielded considerable improvement in energy use estimates, with significant offsetting errors (-13% and +12%, respectively). Lack of sufficient audit data precluded an Operational scenario.

Summary Findings

Our evaluations showed that it is possible with comprehensive physical and operational data to obtain prediction accuracy with the Home Energy Saver to within 1% of energy billing records for groups of homes. It is also possible to generally obtain results within $\pm$25% for well-documented individual homes. On the other hand, evaluations using more simplistic "drive-by" Asset::Visual audits were unable to provide an unbiased estimate of the sample average much less any accounting for house-to-house variation in energy use, sometimes with large errors. The "Asset::Full" cases often came close to collective averages—thanks in part to offsetting errors—but the addition of operational detail minimized scatter and maximized accuracy. In fact, we found that proper specification of occupants' thermostat set points, use intensity of hot water, laundry, and cooking equipment were fully as important to prediction accuracy for individual homes as were physical characteristics, particularly for similar homes. Including operational data resolves a common problem of model over-prediction

The Critical Nature of Model Specification and User Inputs

As revealed in the earlier discussion, the accuracy of any given energy simulation model output is dependent on the availability of appropriate and accurate inputs, and the ability of a model to utilize those inputs. While it is incumbent on model developers to employ appropriate default values, inadequate tuning of inputs to actual conditions should not be construed as simulation inaccuracy, but rather an incomplete information or attention on the part of the modeler. Following are key considerations for default and user-input choices.

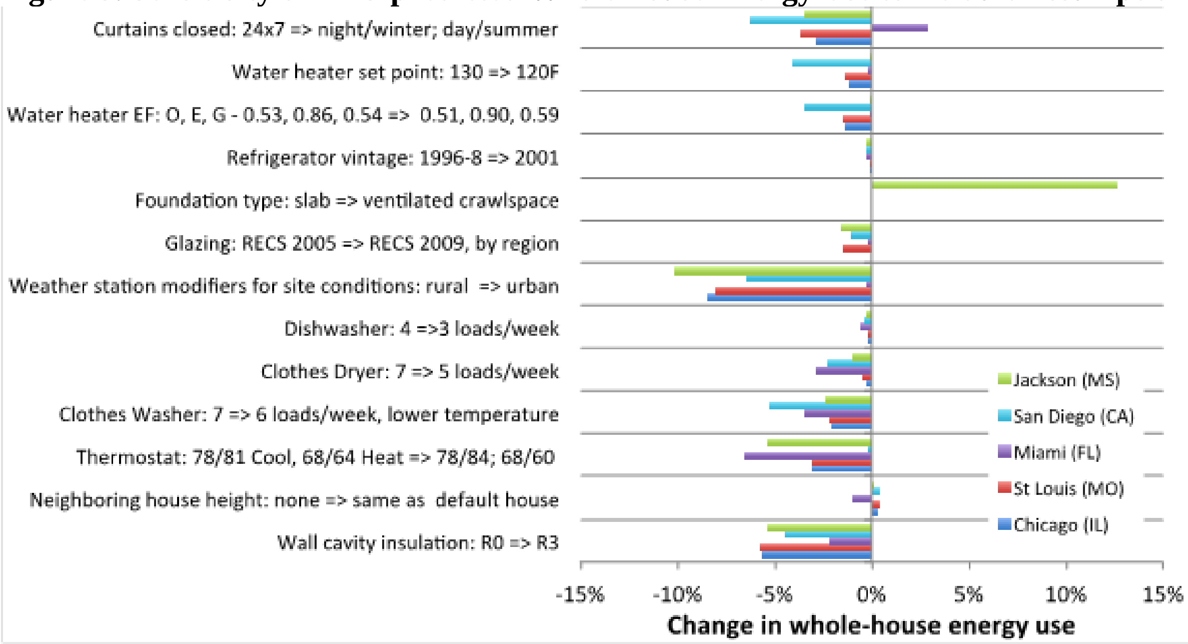
- Disparities in assumed temperatures and thermostat settings were a major factor attributed to the shortfall of actual versus predicted savings estimated by ORNL a quarter of a century ago (Hirst and Goeltz 1985) and later in the Pacific Northwest. These researchers also observed higher post-retrofit temperatures, but it was impossible for them to determine whether this was behavioral *takeback* or the physical fact that well-insulated buildings remain warmer with a given level of internal and solar gains. Thermostat settings were often defaulted in the original Wisconsin audit data, yielding a 22% over-prediction of heating energy on average, with the largest percentage errors for the least-efficient homes (Pigg 2001). Simulation models should default to a thermostat setback and setup unless there are specific data to the contrary. The 2005 RECS data show that approximately half of households report lowering their thermostat during winter sleeping hours. If occupants indicate there is zoning in summer and or winter, modelers should relax thermostat settings if they cannot model it directly. This strategy is commonly used in older and poorly insulated homes to reduce energy costs.
- We learned that if an auditor measures temperatures at the thermostat it is important to distinguish between that value and the value at which the heating or cooling system is switched on. To obtain the appropriate temperature for input into simulation, this often means lower temperatures than the average measured in winter and higher temperatures in summer. This observation also has important implications for occupant-reported thermostat settings that tend to be biased high for heating and low for cooling. Moreover, the

assumption of fully mixed air within a building may be a fundamental error—particularly in poorly insulated buildings—and can lead to over-estimates of energy use.

- Water heating set-points and hot water usage rates are important. The difference between a rule-of-thumb of 64 gallons per day of use, and 54 gallons per day for more homes with water-efficient plumbing can result in significant differences in predicted energy use. We found hot water use to vary from 18 to 113 gallons per day in the Homestead households. Hot water consumption is considerably lower in modern households (EPA 2005).
- Infiltration models should assume a shielding and terrain class IV unless other site-specific information is available. Houses are typically densely packed, and interspersed with trees and surrounding vegetation. Our tests found a 6% heating energy impact for terrain class choice.
- In cases of unknown cavity insulation levels, R-values should be defaulted to a non-zero value such as R-3, reducing the bias of assuming (in lieu of inspections) no insulation.
- Foundation types must be accurately characterized. We found data in our sample (Wisconsin and Oregon) where basements were specified as conditioned when in fact they were not or where crawlspaces were incorrectly defined as basements. This problem has been observed in Minnesota (Quaid and Anderson 1988), and can create significant prediction errors, both in space-conditioning energy and in other values estimated on a conditioned-floor-area basis.
- Specific, actual counts of large appliances should always be used in the simulation model. Omitting a refrigerator can bias the results by 500 kWh or more annually, more so if it is a second older unit. Models should also allow for custom specification of lighting and miscellaneous electric uses (MELs), as does HES. In a sensitivity comparison of results for fixed *vs.* floor-area-dependent defaults for lighting and MELs on the Oregon sample, total whole-house predictive error decreased from -9% to less than 1%.
- Dish- and clothes-washing operational defaults should reflect expected practices and current demographic information. Surveys show definite trends in these behaviors (e.g., fractions of clothes washing loads done with cold water) (Korn and Dimetrosky 2010).

Model default settings are clearly important—particularly when field data on home characteristics and operation are missing—and can contribute to inaccuracies if not reflective of the actual house being modeled. Figure 9 illustrates the impact of recently updated default values in the HES system at the whole-house level for a test suite of typical home constructions in a broader range of climates than covered by the field data evaluated above.

Figure 9. Sensitivity of HES-predicted Whole-house Energy Use to Default Assumptions



Due to extensive interactive effects, results do not sum. The default change in foundation type was applicable only for Jackson (MS), representing changes between the 2005 and 2009 RECS surveys

Toward "Accuracy 2.0"

More can be done to improve energy simulation models. Recent careful evaluations of simulation versus metered data for a subset of homes in Rocklin, California (Backman *et al.*, 2010) illustrate how current hourly simulation models, specifically DOE 2.2 in NREL's BEopt are generally accurate, but prone to over-predict heating energy, and cooling energy to a lesser extent. Based on observable model sensitivities and the building-science literature, we suggest some factors that may account for a portion of such disparity.

<u>(1) Improved modeling of natural air infiltration:</u> Simulations using the Sherman-Grimsrud infiltration model with defaulted terrain and shielding class III may over-estimate the energy impact of infiltration in envelope-dominated building types by neglecting to account for the typical high level of terrain influence and localized shielding in typical suburban environments (Francisco and Palmiter 1994). This is being addressed within ASHRAE Standard 136/119. Modifications have been incorporated into the models being used in HES to better predict infiltration. Evidence also suggests that a small amount of sensible heat recovery occurs through the building envelope as the air flows into or out of cracks and holes. Conversely conditioned air exfiltrating from the building slows down the conduction process. This beneficial heat transfer may be on the order of 5% in uninsulated walls and has been estimated at perhaps 1% for insulated envelopes (Buchanan and Sherman 1998; Akerman *et al.*, 2006).

<u>(2) Window heat transfer with curtains and insect screening</u>: Window insect screening and/or interior curtains impact thermal resistance when in place. Curtains reduce window heat transfer rates by about 17% (Fang 2001). This is rarely, if ever, accounted for in simulations.

(3) Partition walls: We postulate that in poorly insulated buildings the influence of partition walls between the interior and exterior as well as cabinetry, furnishings, and wall hangings collectively exert a significant influence on overall building thermal conductance that has heretofore been overlooked in simulation models (Purdy and Beausolei-Morrison 2001). When central air delivery systems aren't in operation (typically more than two thirds of the time) these walls reduce heat transfer by providing resistance in series to the exterior of the building from the main zone thermostat, in effect increasing the R-value of the envelope. Our parametric evaluation showed that approximating this effect by adding R-2 to window and wall assemblies results in a 25% reduction in heating use in otherwise uninsulated homes.

(4) Software user interfaces are often not considered in accuracy assessments, but they can serve to compound or mitigate the risk of user input error. Well-designed user interfaces can also help the analyst focus attention on inputs that are most important, while supporting "smart defaults" to best estimate the effects of unmeasured behavioral physical characteristics.

Conclusions

Increasingly comprehensive characterization of a home and its operation substantially improves the accuracy of energy use estimates made with the Home Energy Saver, and presumably other energy modeling tools. When allowed to consider a full complement of physical characteristics and occupant behavior, HES predicts actual energy use within 1% of actual consumption, on average for large samples across a diversity of climates and housing types. Predictions for well-characterized individual homes are generally accurate within ± 25%.

We find strong indications that simplified, heavily defaulted asset-only models (disregarding behavioral effects) have less predictive power than models well-tuned to a home's actual characteristics and operation. Indeed, the physical attributes of a house may be less useful for explaining energy use for a particular household than knowledge of how the systems and equipment are used. On the other hand, well-designed asset methods have good predictive power for the central tendencies of a population. Even so, studies that rely heavily on sparse operational data risk reaching specious conclusions by ascribing inaccuracies to the models rather than to deficient inputs or lack of information or skill on the part of the modeler. The rigor and accuracy of energy audit and data-handling processes is crucial to achieving improved model estimates.

We learned that some inputs—particularly specification of thermostat set points and estimates of hot water consumption are critical. Given observed three-fold differences in energy use across otherwise identical homes, accounting for occupancy and behavioral effects is strongly indicated for any evaluations beyond pure-asset rating.

In any model, fortuitous agreement on total energy does not constitute true accuracy if compensating errors are involved. This also undercuts a model's ability to estimate savings for individual measures or end-uses. Sub-metered field data are necessary for conclusive validation.

Accuracy assessments are most beneficial when integrated into the software development process, rather than performed as one-time applications after the fact. Accuracy trials and comparison with well-characterized data sets can help identify programming errors, inappropriate default values, and user-interface defects, thereby enabling continuous improvement of energy modeling tools.

Acknowledgments

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Program and Weatherization Assistance Program, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Scott Pigg and David Heslam facilitated our use of the Wisconsin and Oregon data via Noel Merket and David Roberts of NREL's Field Data Repository. Iain Walker and Rengie Chan at LBNL, Philip Fairey at FSEC, and two anonymous reviewers provided helpful comments.

References

Akerman, M.Y., D. Wilson, and R. Bailey. 2006. "Infiltration Heat Recovery," *ASHRAE Transactions*, QC-06-056.

Backman, C., B. Dakin and D. Springer. 2010. "A Case Study in Reconciling Modeling Projections to Actual Usage," *Proceedings of the 2010 Summer Study on Energy Efficiency in Buildings*, ACEEE, Washington D.C., Vol. 1, p. 36.

Baden, S. 2009. Comments to Diane Ferington (Energy Trust of Oregon), "RESNET Comments on EPS Pilot Draft Study," 27 July.

Birdsall, B., W.F. Buhl, K.L. Ellington, A.E. Erdem, and F.C. Winkelmann. 1990. "Overview of the DOE-2 building energy analysis program, version 2.1D." LBL-19735, Rev. 1. Berkeley.

Bourassa, N.J., L. Rainer, E. Mills, and J. Glickman. 2012. "The Home Energy Scoring Tool: A Simplified Asset Rating for Single Family Homes," *Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings*, ACEEE: Washington, D.C. [Paper 524]

Buchanan, C.R. and M.H. Sherman. 1998. "Simulation of Infiltration Heat Recovery." *19th AIVC Annual Conference*, Oslo, Norway, September.

Chapman, J. 1991. "Data Accuracy and Model Reliability. *Proceedings of the Building Environmental Performance*, Canterbury Conference.

Earth Advantage Institute and Conservation Services Group. 2009. "EPS Performance Score: 2008 Pilot: Findings and Recommendations Report," for the Energy Trust of Oregon.

Energy Center of Wisconsin. 2000. "Energy and Housing: A Study of Single-Family Owner-Occupied Homes." Energy Center of Wisconsin Reports 199-1 and 199-2.

Energy Trust of Oregon. 2012. "Home Energy Performance Scores: Efforts to Date with Modeling Tool Comparison and Summary of Key Issues."

Environmental Protection Agency. 2005. *Water and Energy Savings from High Efficiency Fixtures and Appliances in Single Family Homes.* US EPA, Washington D.C. March 8.

Fang, X. 2001. "A Study of the U-factor of a Window with a Cloth Curtain." *Applied Thermal Engineering* 21:549-559.

Francisco, P.W. and L. Palmiter. 1994. "Modeled and Measured Infiltration in Ten Single Family Homes," *Proc. of the 1994 ACEEE Summer Study on Energy Efficiency in Buildings,* ACEEE: Washington, D.C.

Hendron, R. 2008. "Building America Research Benchmark Definition," NREL/TP-550-44816.

Hirst, E. and R. Goeltz. 1985. "Comparison of Actual Energy Savings with Audit Predictions for Homes in the North Central Region of the U.S.A," *Building and Environment*, 20:1.

Korn, D. and S. Dimetrosky. 2010. "Do the Savings Come out in the Wash? A Large Scale Study of Residential Laundry Systems," 2010 Summer Study on Energy Efficiency in Buildings, Vol. 9, p. 143, ACEEE, Washington D.C.

LBNLa. 2012. "Rules for filtering the audit data for actual homes for the purposes of Home Energy Saver model validation studies." https://sites.google.com/a/lbl.gov/hes-public/accuracy/decision-rules Berkeley, CA: Lawrence Berkeley National Laboratory.

LBNLb. 2012. "FSEC high-fidelity sub-metered residential energy data used for model validation." https://sites.google.com/a/lbl.gov/hes-public/accuracy/submetered-data Berkeley, CA: Lawrence Berkeley National Laboratory.

Mills, E. 1997. "The Home Energy Saver: Interactive Energy Information and Calculations on the Web." *Center for Building Science News* 16(4):1-2. LBL/PUB-731.

Mills, E. 2003. "Inter-comparison of North American Web- and Disk-based Tools for Residential Energy Analysis." *Energy and Buildings* 36:865-880.

Mills, E., *et al.* 2007. "Home Energy Saver: Documentation of Calculation Methodology, Input Data, and Infrastructure". Lawrence Berkeley National Laboratory Report No. 51938. Current version maintained online at https://sites.google.com/a/lbl.gov/hes-public/

Mills, E. and P. Mathew. 2012. "Web Services that Foster Innovation in Buildings Energy Analysis Tools," *Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings*, ACEEE: Washington, D.C. [Paper 818]

Parker, D. S. 2002. "Research Highlights from a Large Scale Residential Monitoring Study in a Hot Climate." Proceeding of International Symposium on Highly Efficient Use of Energy and Reduction of its Environmental Impact, pp. 108-116, Japan Society for the Promotion of Science Research for the Future Program, JPS-RFTF97P01002, Osaka, Japan, January.

Parker, D., M. Mazzara, and Sherwin, J. 1996. "Monitored Energy Use Patterns In Low-Income Housing In A Hot And Humid Climate," *Tenth Symposium on Improving Building Systems in Hot Humid Climates*, Ft. Worth, TX, p. 316

Parker, D., P. Fairey, and R. Hendron. 2010. "Updated Miscellaneous Electricity Loads and Appliance Energy Usage Profiles for use in Home Energy Ratings, Building America Benchmark Procedures and Related Calculations." FSEC-CR-1937-10.

Parker, D.S., P.A. Broman, J.B. Grant, L. Gu, M.T. Anello, and R.K. Viera. 1999. "EnergyGauge USA: A Residential Building Energy Simulation Design Tool", *Proc. Building Simulation '99*, Kyoto, Japan, September 13-15th. Texas A&M, College Station,

Pigg, S. 2001. "A Rating Tale," Home Energy, http://www.homeenergy.org/show/article/id/235

Polly, B., N. Kruis, and D. Roberts. 2011. "Assessing and Improving the Accuracy of Energy Analysis for Residential Buildings." U.S. Department of Energy, Building America Program.

Purdy, J. and I. Beausolei-Morrison, 2001. "The Significant Factors in Modeling Residential Buildings," Canmet Center for Technology, *7th International IBPSA Conf.*, Rio de Janiero.

Quaid, M.A. and M.O. Anderson Jr., 1988. "Measured Energy Savings from Foundation Insulation in Minneapolis Single Family Homes," *Proc. American Council for an Energy Efficient Economy 1988 Summer Study*, Vol. 1, PP. 118-129. ACEEE, Washington DC.

Roberts, D., N. Merket, B. Polly, M. Heaney, S. Casey, and J. Robertson. 2012. "Assessment of the U.S. Department of Energy's Home Energy Scoring Tool," NREL Report (in draft).

Sonderegger, R.C., 1978. "Movers and Stayers: The Resident's Contribution to Variation Across Houses in Energy Consumption for Space Heating," *Energy and Buildings*, Vol. 1, No. 3.