Available online at www.sciencedirect.com

SCIENCE DIRECT®

Energy and Buildings 36 (2004) 865–880

ENERGY and BUILDINGS

www.elsevier.com/locate/enbuild

Inter-comparison of North American residential energy analysis tools

Evan Mills*

Energy Analysis Department, Lawrence Berkeley National Laboratory, MS 90-4000, Berkeley, CA 94720, USA

Received 1 October 2002; accepted 23 November 2003

Abstract

Energy analysis software is an essential component of efforts to foster increased energy efficiency in buildings. In North America alone, there exist hundreds of web- and disk-based building energy analysis tools, serving a diversity of audiences. Some are specialized while others consider the building as a whole. We evaluated 50 web-based residential tools and 15 disk-based tools. While the state-of-the art in tool design has risen considerably over the past three decades, today's users are faced with an increasing—and often overwhelming—array of choices and, often, conflicting results. A surprising number provide little or no detailed analysis of energy savings options. A number of important building energy issues and efficiency features cannot be sufficiently well evaluated using any of the existing tools. Many factors conspire to confound performance comparisons among tools, and the sources or implications of observed differences in results are difficult to pinpoint. For the tools we tested, predicted whole-house energy bills ranged widely (by nearly a factor of three), and far more so at the end-use level. We also discovered a remarkable number of indications of errors in programming or algorithms. Tool design should be grounded in social science and engineering. Analytical results (e.g., benchmarking) and end-use-specific "what-if" functions are more helpful for many users than rarified engineering outputs. Desirable technical features include modeling of occupant effects, open-ended energy calculations as well as results normalized to actual consumption history, incorporating means for users to grasp the uncertainties embodied in the results, and ensuring quality control to remove errors from the design and programming of tools. More coordinated planning of tool development could help address the fragmentation and dilution of efforts that has historically hampered tool quality and market penetration.

1. Residential energy analysis software

Energy analysis tools are integral to the process of identifying and implementing building energy savings measures. Such tools have many uses, ranging from consumer education to performing detailed design analyses. The scope can vary from a component- or end-use level to the whole building. The intended audiences vary correspondingly, from end users to intermediaries such as contractors or auditors to policy analysts. Our interest here is primarily on whole-building residential tools intended primarily for end-user audiences.

In their ideal form, building energy tools enable users to accurately and cost-effectively evaluate energy use and savings opportunities as well as non-energy issues such as cost, environment, comfort, safety, and aesthetics. Basic tool building blocks include the core simulation engines and algorithms, coupled with user interfaces, and supported with data on weather and component properties. The long-term vision held by many in the building science community is one involving virtual (collaborative) "life-cycle" building tools that simulate actual buildings and their construction coupled with intelligent systems that monitor and archive design intent and performance and feed the results back to the simulation tools, which, in turn, grow more refined through integrating better empirical data [1].

The origins of building energy software trace back to the 1970s. Prior to that time, energy audits were conducted by hand at significant cost. In the 1980s, the first-generation of simulation-based analysis and design tools came into use by researchers and consultants. The 1990s were marked by tool improvements and a rapid proliferation of tools targeted at a broader spectrum of users, including commercial and residential consumers, and the advent of web-based tools. In parallel with these technical developments was a perhaps 500-fold reduction in the cost of delivering tool-based audits.[1]

---

* Tel.: +1-151-486-6784; fax: +1-151-486-6996.
E-mail address: emills@lbl.gov (E. Mills).

[1] According to Michaels [2] the evolution from the early computer-based residential audits to the emerging email-based audits has seen a cost reduction from approximately US$ 250 per home to US$ 0.50 per home. The cost reductions were due to a combination of lower computing costs, reduced human labor, and increased penetration.

0378-7788/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.enbuild.2003.11.005

Persistent barriers to the mainstream adoption of building energy tools include the time required to use them, process the often-extensive outputs, and evaluate strategies for reducing energy use below the performance level predicted for the existing or baseline building. It can require the use of multiple tools and multiple “runs” to evaluate alternate scenarios.

Despite steady improvements over time, residential energy tools have attained very low market penetration. This has been partly ascribed to the extensive fragmentation of development and deployment efforts (as evidenced by the hundreds of tools in existence), resulting in a proliferation of tools each with a low user base and insufficient developer revenues to support continued development. Development teams typically number from one to five people, versus one to two hundred even for considerably simpler mainstream consumer software (e.g., checkbook-balancing tools). Numerous bugs and runtime instabilities evidence a lack of sufficient resources for quality assurance. The argument has been made for unifying the currently disparate development efforts into a more coordinated and collaborative initiative [3]. This is particularly logical given the limitations and volatility of public-sector funding for tool development.

2. Prior reviews

Mills and Ritschard [4] previously evaluated disk-based tools applicable to multifamily buildings. Most of these tools no longer exist, while the remainder evolved considerably since the original review. *Home Energy* magazine has published various review articles, each of which looks at only a handful of tools (e.g. [5]). The Electric Power Research Institute commissioned a proprietary review of four web calculators in 1998 [6].

A particularly thorough prior study was conducted for the California Energy Commission [7]. Although only eight residential tools were evaluated (two disk-based and six web-based), the information collected was more detailed than in prior studies. The study concluded that a tool should provide three kinds of recommendations (1) no-cost options, such as behavioral changes, (2) envelope measures applicable during remodeling, and (3) equipment retrofits. The report lists non-energy benefits and case studies as additional information that tools should offer, as well as multiple user levels, recallable results, comparisons among multiple scenarios, and the ability to evaluate single measures (i.e. without having to do a whole-house survey). The authors emphasize the importance of tools that “educate” the user (i.e. not just generate numbers). The study concluded that no single tool “consisted of all the desirable features and functionality”. Two additional related criticisms of all the tools were that recommendations are often vague and don’t specify the exact efficiency level that consumers should select, and that some give ranges (instead of point values) for results but do not assist the user in understanding the underlying uncertainties.

Few studies have grappled with the question of tool accuracy. In the early 1980s, Wagner [8] compiled measured versus predicted energy consumption in a verification exercise spanning 100 simulations performed by 18 tools. ANSI/ASHRAE Standard 140-2001, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs, specifies test procedures for evaluating the technical capabilities and ranges of applicability of computer programs that calculate the thermal performance of buildings and their HVAC systems [9].

While not a critical review, The US Department of Energy’s Building Energy Software Tools Directory [10] is a rich compilation of tools and developer-provided information per a standard format. Lawrence Berkeley National Laboratory also maintains an on-line list of buildings energy software tools [11].

3. Methodology

To identify candidate residential tools for evaluation, we conducted web and literature searches, including review of the above-mentioned prior work. Our investigation was limited to tools developed and used in the North American

Table 1
List of tools examined

Tool name (web-based)	Developer
Appliance calculator	San Diego Gas and Electric Co.
Appliance Energy Estimator	Southern California Edison
ATCO Energy Sense House	Atco Gas (Canada)
BEACON	Oarsman
Chicopee Electric Light Department	Chicopee Electric Light Department
City of Oxford Electric Energy Calculator	City of Oxford
Comfort Check	Enercom & Nicor
Ecalc	Pacific Gas & Electric Company
ELPC Pollution Calculator	ELPC
Energy Calculator	Niagra Mowhawk
Energy Calculator	Electric Power Research Institute
EnergyCheckup.com	GeoPraxis

Table 1 (*Continued*)

Tool name (web-based)	Developer
EnergyCheckup.com HERS version	GeoPraxis
EnergyGuide	Nexus
Environment—Energy Calculator	BC Hydro
EREN Energy Conversions Calculator	US Department of Energy
Find Out About Your Electricity	Environmental Defense
Home Energy Advisor (EPA/LBNL)	USEPA/Lawrence Berkeley National Lab
Home Energy Analysis (SMUD)	Sacramento Municipal Utility District
Home Energy Audit	Texas-New Mexico Power Company
Home Energy Calculator	Central Main Power Company
Home Energy Checkup	Alliance to Save Energy
Home Energy Saver	USDOE/Lawrence Berkeley National Lab
Home Energy Survey	Electrotek Concepts
Home Energy Survey	Pacific Gas & Electric Company
Home View	Volt VIEWtech
Watts On Schools	American Electric Power
KCPL Electricity Calculator	Kansas City Power & Light
KUA Electricity Calculator	Kissimmee Utility
My Home	Redwire/Green Mountain Power
On-Line Energy Profile	San Diego Gas & Electric
On-Line Home Energy Audit	International Council for Local Environmental Initiatives
On-line Home Energy Survey	Southern California Edison
ORNL Calculators	Oak Ridge National Laboratory
Personal Energy Profiler	United Illuminating
PowerSmart Home; PowerSmart Business	BC Hydro
PSNH Electricity Calculator	Public Service of New Hampshire
PVWatts	National Renewable Energy Laboratory
Radon Project	Columbia University & Lawrence Berkeley National Lab
Residential Calculator	Utilities
Residential Energy Bill Analyzer	Florida Power Corporation
Residential Energy Bill Analyzer	Electrotek Concepts
Residential On-Line Energy Audit	Enercom
Residential Ventilation Calculator	Lawrence Berkeley National Laboratory
RP&L Energy Calculator	Richmond Power & Light
Solar Energy Calculator	Iowa Energy Center
Torchiere energy cost and payback	Lawrence Berkeley National Laboratory
Twenty Percent Solution	Lawrence Berkeley National Laboratory
Western Massachusetts Online Energy Calculator	Western Massachusetts Electric Co.
Your California Home	GeoPraxis
Tool Name (disk-based)	**Developer**
AkWarm (v1.03d)	Alaska Housing Finance Corporation
BTU Analysis REG (v6.1.0)	Enchanted Tree Software
Energy-10 (v1.4.035)	NREL, LBNL
ENERPASS (v4)	Enermodal Engineering
EZDOE (v2.1)[a]	Elite Software
E-Z Heatloss (v6)[a]	Thomas & Associates
HOT2000 (v8.606)	Natural Resources Canada
J-Works (v4.809)	MicroWorks, Inc.
MECcheck (v3.3)	US Department of Energy
Micropas (v6.01)[a]	Enercomp, Inc.
NEAT (v7.1.3)	Oak Ridge National Laboratory
REM/Rate (v10.3)	Architectural Energy Corporation
ResRatePro (v1.26)	Florida Solar Energy Center
TREAT (v0.8.985)[b]	Taitem Engineering
VisualDOE (v3.0.111)[a]	Eley Associates

[a] Demonstration version.

[b] Beta version.

context.[2] The disk-based tools were selected from the on-line version of DOEs directory [10]. The selected tools (Table 1) were then cataloged and reviewed for useful features, methods of presentation, interface design concepts, etc. From this main set, the subset with a “whole-house” orientation were identified and evaluated in considerably more depth.[3] To be included in the “whole-house” detailed review, a tool was required to consider the full range of residential energy end uses and fuels.

Our methodology for comparing tool characteristics expands significantly on that used by Mills and Ritschard [4]. A detailed matrix was constructed to capture, as comprehensively as possible, house and household description, output, user-support features, and analytical methods used by each tool (Table 2). The range of user needs and the corresponding presence of these features in the tools informed decisions about which features to record in the matrix. No one tool possessed all possible features.

As a basis for top-level comparisons, we determined how many inputs were possible for each tool. There are various ways to define an “input”. Some prior studies have used “number of questions” (e.g. [7]), but individual questions often seek multiple pieces of information. We found it more consistent and meaningful to include the actual number of pieces of information that a user might have to enter. We observed that the numbers of inputs and input screens are systematically higher for the disk-based tools, due to their professional target audiences and correspondingly greater sophistication, especially concerning extensive materials and component libraries in which multiple characteristics of multiple items can be specified by the user.

We also separately tallied the numbers of technical features (including house and household description) and decision-support features (including calculation methodology, output, and user-support) for each tool to glean an overall sense of the comprehensiveness and ease of use.[4]

Toward the goal of understanding and comparing the tools’ predictive power, we chose real homes for which we had actual consumption data and a detailed description of physical characteristics and occupant behavior. We compared the tool results to the test houses and to each other. The choice of two test houses (California and Ohio) allowed us to explore different climates. While this is not a comprehensive accuracy evaluation; the results are useful

Table 2
Information tabulated for each tool

General Information
Developer
Program version
Release date
Cost
Computer requirement
Operating system requirement
Hard drive requirement
RAM requirement
Commercial/e-commerce content
Privacy statement
Non-proprietary
User base
Audience(s)
Consumer
Professional
Type of tool
Open-ended calculation
Bill disaggregation
User level(s)
Input screens
Inputs
Technical Features—General Building
Age of house
Building type(s)
Single-family detached
Townhouse
Apartment building
Mobile home
Room additions
House geometry
House orientation
Number of stories
Floor area
Ceiling height
Exterior shading
Technical Features—Building Envelope
Foundation type
Wall exterior/construction type
Doors
Insulation levels
Foundation
Floors
Walls
Ceiling
Roof
Attic radiant barrier
Roof color, reflectance, or absorptance
Window area
Glazing/frame types
Skylights
Leakage (airtightness)
Leakage (blower door data)
Caulking and weatherstripping
Technical Features—HVAC Systems
Heating system type(s)
Cooling system type(s)
Secondary heating
HVAC system efficiency
Duct location/insulation/sealing

---

[2] There is ample room for additional research in this regard, which would have to address differences in energy analysis conventions, language, etc. A recently passed Buildings Directive in the European Union mandates the establishment of energy rating systems. As a result, new initiatives have begun in many countries, usually involving development of simulation tools.

[3] Due to cost or other constraints, test or demonstration versions were used in some cases. Our data tables were provided to the developers for verification.

[4] Evaluations of speed and performance for web-based tools were conducted on DSL or faster connections. The disk-based tools were evaluated using a PC equipped under Windows with an x86Family 6 Model 8 Stepping 10 Intel ∼356 processor.

Table 2 (*Continued*)

- Ceiling fans
- Whole-house fans

Technical Features—Major Appliances

- Water heating
  - Types
  - Fuels
  - Solar water heating
  - Variables (e.g., setpoint, recovery factor)
  - Water conservation options
- Refrigerator
- Freezer
- Refrigerator and freezer sizes
- Multiple refrigerators and/or freezers
- Stove
- Oven
- Dishwasher
- Clothes washer
- Clothes dryer
- Hot tub or spa

Technical features—miscellaneous end uses

- Miscellaneous end uses (gas and electric)
- Usage-driven end uses
- Miscellaneous gas end uses
- Module to describe generic appliances

Technical features—occupant effects

- Number of occupants
- Ages of occupants
- Occupants home during day
- Movable window insulation
- Movable window shades
- Thermostat type(s)
  - Standard
  - Setback option
  - Programmable
- Zone heating/cooling
- Water heater setting
- Tap water consumption
- Use of cooking appliances
- Use of dish- and clothes-washing appliances
- Use of clothes line
- Use of miscellaneous appliances
- Use of lights

Technical Features—IAQ

- Calculations
- Content

Technical features—economic analysis

- Variable energy prices
- Variable energy tariffs (e.g., block rates, TOU rates)
- Cost-effectiveness indicator(s)
  - LCC
  - IRR
  - PBT
  - Other
- Rebates, tax incentives, etc.
- Early appliance retirement

Energy analysis methods and details

- Type of calculation(s)
  - Simulation
  - Engineering estimates
  - Watts X hours
  - Survey data/lookups
- Weather locations
- Solar gains
- Internal gains
  - Occupants
  - Appliances
- Lighting
  - Aggregate analysis
  - Room-by-room or fixture-by-fixture
- Retrofit/savings calculations include interactions
- Calculation time-step
- Transparency of assumptions and methods

Defaults

- Location-dependent defaults
- Pre-defined prototype library
- HVAC-vintage-driven defaults
- Appliance-vintage-dependent defaults

Outputs

- Energy consumption
- Peak electricity demand
- Energy savings
- Energy cost/savings
- Consumption by fuel type
- Cost by fuel type
- End-use breakdowns
- Retrofit recommendations
  - No-cost measures
  - Cost-associated measures
  - Ranking of measures
  - Flexibility of retrofit cost assumptions
- Benchmarking
- Run comparisons
- HVAC system sizing
- Water consumption
- Emissions
- Output time-step
- Graphical outputs
- Stored/retrievable runs

User and Decision-Support Services

- Internal text-based content
- FAQs
- Glossary
- General program help
- Context-sensitive help
- Help search
- Example input and output sets
- Case studies
- Non-energy benefits
- Links to external energy-related web sites
- E-mail support

in demonstrating the variations among tool results and the need for more exhaustive validation efforts. Lacking was sub-metered end-use data to compare against end-use predictions from the various tools.

It is important to keep in mind that the tools evaluated, especially those that are web-based, are under continuous development. Only those features available to users at the time of the evaluation (Spring, 2002) were recorded. The review of web-based tools was exhaustive, whereas the disk-based tools represent only a subset of those available.

4. Findings

4.1. Existing tools exhibit considerable range and creativity

Our review shows that there are many approaches to the design of residential energy tools and different levels of detail can be offered to users. More detail (questions asked) does not, however, automatically translate into a "better", more thorough, or more accurate tool. As suggested by a comparison of Fig. 1a–c (for web-based tools) and Fig. 2a–c (for disk-based tools), some require a relatively small number of well-considered inputs while others ask a proliferation of questions and still miss key issues. For example, the Kansas City Power and Light's web-based tool asks 198 questions, but only encompasses 30 of the 124 potential features itemized in Table 2.[5]

The value of detail has a lot to do with the type of answers sought by the user (e.g., the availability of dozens of miscellaneous appliances is immaterial for a user attempting to evaluate their potential for space-heating savings by installing a new heating system).

The tools vary in their usability (e.g. approachability, navigability, wait time, etc.). Some have very elegant and easy-to-navigate interfaces while others were cumbersome (e.g., many screens, poor text legibility). Some are able to collect large amounts of information via a simple interface, while others had elaborate interfaces that did a poor job of collecting information. Several of the tools provide the user the opportunity to compare a base-case house with one outfitted with one or more energy efficiency measures.

Considerable creativity is demonstrated in the design of many existing tools. Even tools that are not particularly comprehensive (e.g. those providing load calculations only) have things to offer. While the diversity of specialized tools offers valuable features to users, it is disadvantageous that they are not interoperable, e.g. similar information must be re-entered for each tool and the results are not coordinated or integrated.

4.2. Users face bewildering choices and often-confusing input requirements

There are today hundreds of web- and disk-based energy tools. Approximately 220 were listed in DOEs Building Energy Tools Directory as of Spring 2002 [10].

The first web-based energy calculator was the Home Energy Saver, developed in the mid-1990s. There ensued a rapid proliferation of web-based energy calculators. There has since been considerable consolidation; many web-based tools have vanished from the Internet. The (often unanticipated) cost of building and maintaining these sites is no doubt a factor in this trend [12].

In the course of this study, we identified 50 web-based residential calculators, 21 of which can be considered "whole-house" tools. Of the whole-house tools, 13 provide open-ended energy calculations, 5 normalize the results to actual costs (a.k.a "bill disaggregation tools"), and 3 provide both options. Across the whole-house tools, we found a range of 5–58 house-descriptive features (68 possible) and 2–41 analytical and decision-support features (55 possible).

We also evaluated 15 disk-based residential calculators. These tools offer ranges of 18–58 technical features (70 possible) and 10–40 user- and decision-support features (56 possible). Of these tools, 11 provide open-ended energy calculations, one normalizes the results to actual costs, and three provide both options.

The disk-based tools contain 21–364 input screens and 45–9870 inputs, far more than the corresponding numbers for web-based tools. Despite the large numbers of potential inputs, limitations in the designs of some of the disk-based tools limit users' abilities to model their homes with the desired level of detail. The limiting of house geometries to a six-surface box shape is an example of this type of shortcoming.

Meta-evaluations of the disk- and web-based tools are presented in Tables 3 and 4, and the complete matrices of features appear in Mills [13].

The tools exhibit a large range in analytical scope. It was surprising how few enable the evaluation of certain key energy issues and opportunities, e.g. the performance of thermal distribution systems, advanced windows, cool roofs, or programmable thermostats. Few address indoor air quality considerations and other non-energy benefits of energy efficiency [14]. Most tools, however, give considerable (and appropriate) attention to miscellaneous energy end uses. Various important buildings energy issues and energy efficiency features cannot be sufficiently well evaluated using any of the existing tools (e.g., peak power demand, IR reflective roofing, high-R perimeter attic insulation, thermal comfort, advanced crawlspace/foundations, advanced thermal distribution modeling, early appliance retirement, time-of-use tariff structures). Few tools offer substantial decision-support content (either local or via links to useful web sites).

Many tools provide estimates of baseline energy bills but no recommendations or estimates of potential savings, and fewer still address cost-effectiveness or emissions analysis (even superficially). Where available, most savings recommendations are spotty, with a large focus on low/no cost measures (often focusing on appliance usage) and less on investments in better equipment or envelopes. Most recommendations are illustrative rather than comprehensive, e.g.

---

[5] The EPRI tool is another example that appears to be very extensive (9 input screens and 79 questions), yet is in fact very inflexible and full of embedded assumptions. For example, the efficiencies of heating systems and many other appliances are fixed, and by having the user enter "number of hours per year use of heating system" the building size, geometry, and envelope characteristics are entirely by passed.

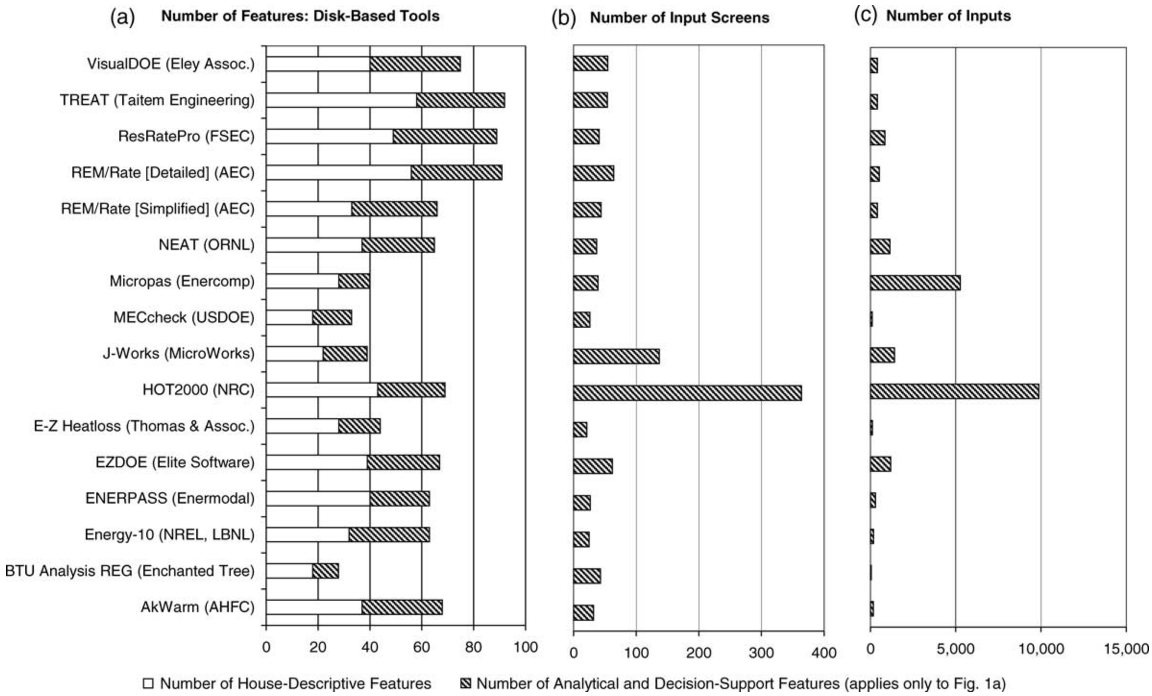


Fig. 1. (a–c) Features, input screens, and inputs vary widely: disk-based tools.

for our test house, SCEs On-Line Home Energy Survey only suggests caulking and weather-stripping, CFLs, and occupancy sensors for outdoor lighting.

Input questions are often formulated in a way that is likely to confuse lay users. In one of many examples, the tool asks for hours per day refrigerator usage, with a default of 24 h, while another tool asks the same question and defaults at 5 h, and yet another asks for hours per month and the default value offered is 335 ($24 \times 30 = 720$). In the EPRI tool, users are asked to enter the number of hours their heating system operates in each year. Even an "energy expert" would not likely be able to make an accurate guess at this value. In yet another example, one tool asks for total lighting hours aggregated by bulb type. This is an unreasonably challenging question for the typical consumer and invites poor estimates and thus inaccurate results.

4.3. *Web- and disk-based tools differ considerably*

Only one web-based tool in our compilation is suitable for professional audiences, as opposed to all of the disk-based tools. This is a somewhat subjective determination. We based our judgment on a tool's technical depth and flexibility, e.g., in modeling specific equipment efficiencies, complex building geometries, a wide range of climates, and providing sufficiently detailed outputs for a professional user's needs. Several of the disk-based tools (and none of the web-based tools) are intended primarily for non-residential applications.

The level of detail varies accordingly, with up to 200 possible inputs among the web-based tools versus a maximum approaching 10,000 for the disk-based tools. The disk-based tools offer correspondingly greater choice and control over building characteristics, system sizing, weather location, outputs, etc. However, the disk-based tools generally offer a narrower end-use coverage and thus there are fewer (in comparison to the web-based tools) that qualified for the "whole-house" designation used in this study. None of the disk-based tools offer recommendations on no-cost energy-saving measures, while most of the web-based tools do so. Few of the disk-based tools offer a cost-effectiveness protocol for evaluating energy retrofit measures, whereas most of the web-based tools do.

Perhaps counter-intuitively, the web-based tools are more sophisticated in some areas. For example, they more frequently provide vintage-dependent defaults for appliance and equipment efficiencies.

The market distribution of disk-based tools is naturally narrower than that of web-based tools. With one exception, the disk-based tools had between 50 and 2300 copies in circulation (MECheck had 25,000 copies). The web-based tools are more accessible to anyone using the Internet, and, among those we evaluated, receive up to 350,000 visitors per year. None of the disk-based tools work on a Macintosh platform, while all of the web-based tools are (by definition) platform-independent.

The web-based tools are free to users, whereas, with a few exceptions, a fee is required to acquire the disk-based ones. In some cases, however, access to web-based tools is restricted to customers of specific utilities (who pay licensing fees to the developers).

With one exception, all disk-based tools we examined provide documentation, making their embedded assumptions

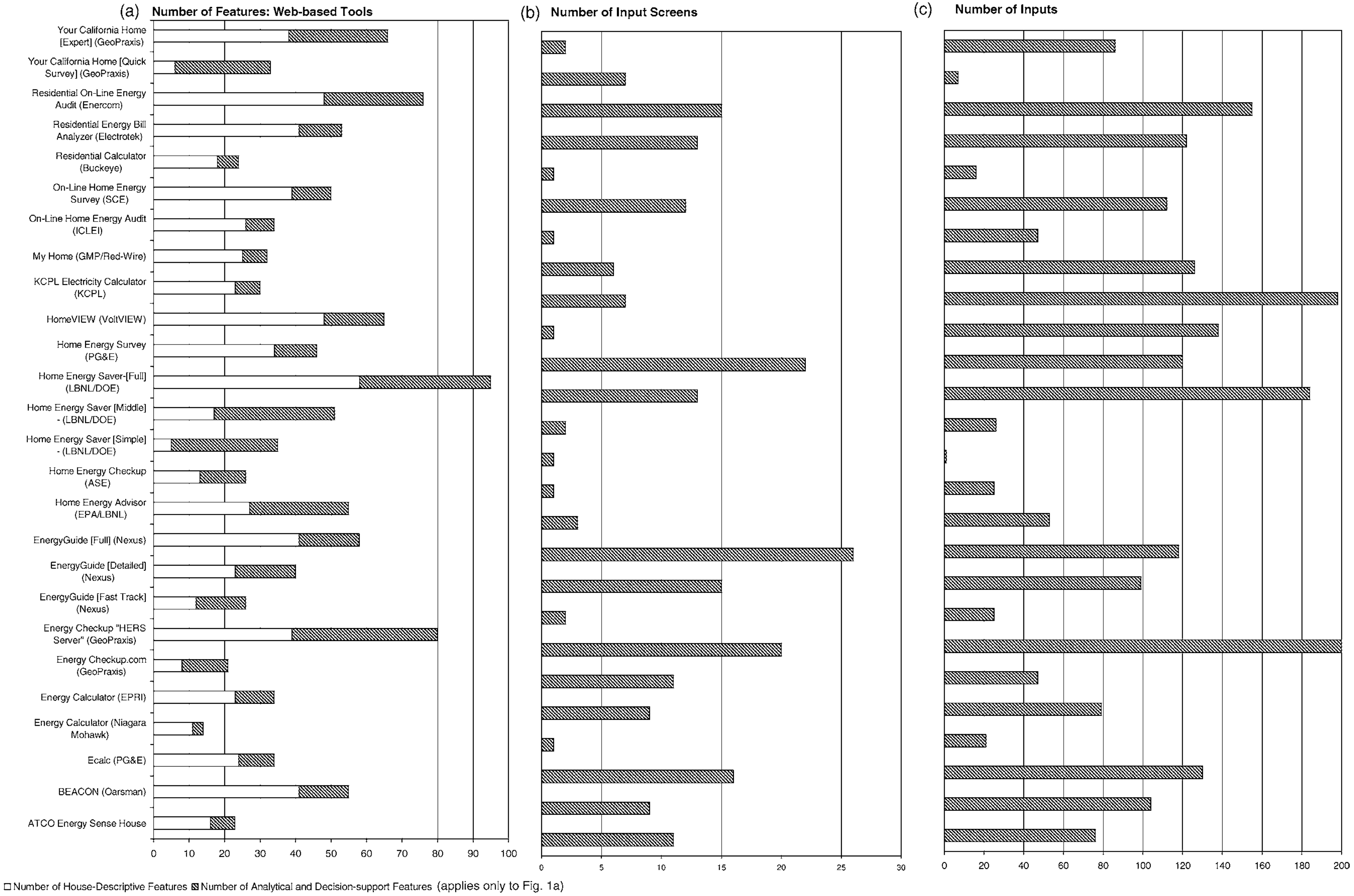


Fig. 2. (a–c) Features, input screens, and inputs vary widely: web-based tools.

Table 3
Meta-evaluation: web-based tools

	ATCO Energy Sense House	BEACON (Oarsman)	Ecalc (PG&E)	Energy Calculator (Niagara Mohawk)	Energy Calculator (EPRI)	Energy Checkup.com (Geopraxis)	Energy Checkup "HERS Server" (Geopraxis)	EnergyGuide [Fast Track] (Nexus)	EnergyGuide [Detailed] (Nexus)	EnergyGuide [Full] (Nexus)	Home Energy Advisor (EPA/LBNL)	Home Energy Checkup (ASE)
• Ease of use/speed of calculations	Somewhat Difficult/Very Fast	Reasonable/Very Slow	Efficient/Very Fast	Efficient/Very Fast	Cumbersome/ Very Fast	Not functioning at time of evaluation	Not made available for evaluation	Reasonable/Slow	Somewhat Difficult/Slow	Cumbersome/Very Slow	Efficient/Fast	Efficient/Very Fast
• Overall suitability for building envelope/HVAC analysis	Very Low	Low	Very Low	Very Low	Very Low	Not functioning at time of evaluation	Not made available for evaluation	Very Low	Very Low	Low	Moderate	Low
• Overall suitability for appliance analysis	Moderate	High	High	Low	Moderate	Not functioning at time of evaluation	Not made available for evaluation	Very Low	Moderate	High	Moderate	Low
• Overall suitability for occupant effect analysis	Moderate	Low	Low	Very low	Moderate	Not functioning at time of evaluation	Not made available for evaluation	Very low	Very low	Good	Moderate	None
• Overall helpfulness of outputs and other information in supporting decisions	Low	None	Low	Very Low	None	Not functioning at time of evaluation	Not made available for evaluation	Moderate	Moderate	Moderate	Moderate	Moderate

	Home Energy Saver [Simple] - (LBNL/DOE)	Home Energy Saver [Middle] - (LBNL/DOE)	Home Energy Saver-[Full] (LBNL/DOE)	Home Energy Survey (PG&E)	HomeVIEW (VoltVIEW)	KCPL Electricity Calculator (KCPL)	My Home (GMP/Red-Wire)	On-Line Home Energy Audit (ICLEI)	On-Line Home Energy Survey (SCE)	Residential Calculator (Buckeye)	Residential Energy Bill Analyzer (Electrotek)	Residential On-Line Energy Audit (Enercom)	Your California Home [Quick Survey] (Geopraxis)	Your California Home [Expert] (Geopraxis)
• Ease of use/speed of calculations	Efficient/Very Fast	Efficient/Fast	Reasonable/Fast	Cumbersome/ Fast	Efficient/Slow	Somewhat Difficult/Very Fast	Cumbersome/ Very Fast	Efficient/Very Fast	Reasonable/ Very Slow	Efficient/Fast	Reasonable/Fast	Reasonable/ Acceptable	Somewhat Difficult/Slow	Reasonable/ Very Slow
• Overall suitability for building envelope/HVAC analysis	Very Low	Moderate	High	Low	High	Very Low	Very Low	Moderate	Low	Very Low	Low	High	Very Low	Moderate
• Overall suitability for appliance analysis	Low	Low	High	High	High	Moderate	Moderate	Very Low	High	Low	High	High	None	Moderate
• Overall suitability for occupant effect analysis	Low	Low	High	Moderate	Moderate	Moderate	Moderate	Low	High	Low	High	Moderate	None	High
• Overall helpfulness of outputs and other information in supporting decisions	Moderate	Moderate	Moderate	Moderate	Low	Low	None	Low	Low	Low	Low	Moderate	Low	Low

Table 4
Meta-evaluation: disk-based tools

	AkWarm	BTU Analysis REG	Energy-10	ENERPASS	EZDOE	E-Z Heatloss	HOT2000	J-Works
• Ease of use/speed of calculations	Reasonable/ Very Fast	Reasonable/ Very Fast	Reasonable/ Acceptable	Cumbersome/Very Slow	Cumbersome/?	Reasonable/Very Fast	Cumbersome/ Fast	Somewhat Difficult/Very Fast
• Overall suitability for building envelope/HVAC analysis	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
• Overall suitability for appliance analysis	Low	Very Low	Very Low	Very Low	Very Low	Very Low	Moderate	None
• Overall suitability for occupant effect analysis	Very Low	Very Low	Moderate	High	Moderate	Low	Moderate	Very Low
• Overall helpfulness of outputs and other information in supporting decisions	Moderate	Moderate	Moderate	Moderate	Moderate	Low	Moderate	Moderate

	MECcheck	Micropas	NEAT	REM/Rate (Simplified)	REM/Rate (Detailed)	ResRatePro	TREAT	VisualDOE
• Ease of use/speed of calculations	Reasonable/ Very Fast	Reasonable/ Acceptable	Somewhat Difficult/Fast	Reasonable/Fast	Reasonable/Fast	Reasonable/Slow	Cumbersome/ Very Slow	Cumbersome/ Slow
• Overall suitability for building envelope/HVAC analysis	Moderate	Moderate	Moderate	Moderate	High	High	High	Moderate
• Overall suitability for appliance analysis	None	None	Moderate	Very Low	High	Moderate	High	Low
• Overall suitability for occupant effect analysis	None	Very Low	Moderate	Low	Moderate	Moderate	High	Moderate
• Overall helpfulness of outputs and other information in supporting decisions	Moderate	Low	Moderate	Moderate	Moderate	High	Low	Moderate

and methods transparent, whereas only one web-based tool does so.

4.4. Evaluating accuracy is an elusive goal

The question of tool “accuracy” is a complex and elusive one. The method of examining accuracy itself (e.g. using occupied versus unoccupied buildings) influences the results [8].

The ability to evaluate accuracy is inherently limited by the availability of measured end-use data, and manipulations of that data (e.g., weather normalization) to facilitate meaningful comparisons to tool outputs. Certain tool outputs can only be measured against “actual” values that are themselves calculated (e.g., HVAC sizing), while others are rarely if ever available (e.g., measured energy use or savings for specific retrofit measures).

Similarly challenging is to understand the *sources* of inaccuracies. As described below, there are many ways in which quantitative errors can occur in tools, ranging from programming errors to problems inherent in a tool’s design.

4.4.1. Types of accuracy problems

Conducting analytical inter-comparisons of residential energy tools raises a number of complicated issues, and the question of “accuracy” has multiple definitions. There are several potential sources of inaccuracy in the results produced by a given tool. The specific illustrations provided below are based on spot checks rather than exhaustive trials of tools included in our review.

4.4.1.1. A tool’s underlying engineering calculations or simulation techniques may contain inaccuracies. Pinpointing the source of such a problem can be virtually impossible for outside reviewers who lack access to technical documentation and the underlying source code and assumptions. See examples below.

4.4.1.2. Even if baseline calculations are accurate, savings calculations may not be. Finding measured data with which to validate savings calculations is far more problematic than finding data to validate baseline bills. Ideally, measured data are needed for savings estimates in each end-use category. Some of the savings estimates we encountered when running our test homes were implausible. One tool estimated the annual savings for a water heater blanket at questionably low values of US$ 2 per year, and at US$ 4 per year for reducing the water heater temperature. Another tool reported only US$ 2 per year annual savings for duct insulation. When testing another tool, going from zero ceiling insulation to “R20–30” resulted in US$ 12 per year HVAC savings, and going from “never” changing the air-conditioner filter to changing “every 3 months” resulted in no change in HVAC costs. When specifying a 10- to 15-year-old standard non-ENERGY STAR model washer in one tool, it predicted only US$ 2 per year savings for upgrading (same answer for hot or cold wash temperature and independent of the number of loads washed). Another tool classified all clothes washers as “energy efficient”, irrespective of the age (up to 27 years) specified by the user.

4.4.1.3. Changes to inputs do not always result in expected changes in predicted energy use. When examining one tool, we noted that energy bills *decreased* when the water-heater thermostat was *increased* from 130 to 140° range to the 140–150° range, and were virtually the same from the “Low: below 120°” setting to the “Very high: over 150°” setting. Similarly, energy use increased with decreasing shower length. Computer energy use increased only $2/year when utilization was changed from “a little” to “a lot”.

We noted several web-based tools in which the results did not always equal the sum of the individual end uses. In another example, the tool did not show any differences in energy bills as a function of house size (we tested a range of 1000–1500 square feet to 2000–2500 square feet). Another tool failed to capture the impact of roof insulation when both roof and attic insulation are specified for an unconditioned attic, and greenhouse-gas emissions calculated by that tool did not always increase when energy use increased.

Bill disaggregation tools provide special challenges. One tool reported increased heating use (US$ 1119 versus US$ 992) when a smaller home size (1000–1499 sq. ft versus 2000–2499 sq. ft) was specified. Also counterintuitive, lighting energy use was identical in the two homes. We observed the same problems in another tool, where in fact lighting energy use increased with decreasing house size. This particular bill-disaggregation tool also computed the same baseline air conditioning use for SEERs 6 to 16, perhaps an artifact of inflexible values for other end uses and an actual energy bill that must be matched.

4.4.1.4. User-specifiable options are often incomplete or not representative of the actual building. Particular issues arise when users attempt to model non-typical homes or usage patterns. Cases involving particularly low- or high-energy-use homes are most likely to exhibit under/overestimation of results (except, of course, when using bill-disaggregation tools). For example, extreme high or low thermostat settings will lead to actual bills that differ from those predicted by tools that do not allow for explicit entry of thermostat settings. Problems can also arise, for example, in tools that specify ranges for inputs, such as a vintage range of “before or later than 1993” for appliance efficiency, implying only two possible “average” efficiency levels based on the user answer, where in fact the user could have an ancient appliance or a brand-new premium-efficiency model.

If a tool excludes miscellaneous uses, for example, results can easily be 20–30% lower than utility bills for this reason alone (and all end uses overestimated in the case of bill-disagregation tools). Half of the tools we tested reported miscellaneous energy at less than 10% of total bills, a highly unlikely scenario, and in one case completely excluded it. Other examples include lack of provision for more than one refrigerator, values specified as a range (e.g., floor area), or that otherwise do not fit reality (e.g., different walls have different *R*-values). One tool relies solely on defaulted building descriptions keyed to the user-entered zip code, and thus the resulting defaults will inevitably fail to fully represent the actual home in question (e.g., attribution of cooling energy use where none may exist in fact). Another tool does not allow fractional hours of use for many miscellaneous appliances (e.g., toasters, microwaves), this can lead to over-prediction of energy costs. Another does not allow furnace efficiencies below 78%.

4.4.1.5. Interface design and questions formulated by some of the tools foster input errors or poor house descriptions that adversely affect the results. These potential problems fall into two categories. “Hidden” options—those discretely placed in rather long pull-down menus or activated by the selection of related “lead-in” options—can easily go unnoticed. “Surrogate” inputs can also trigger unnoticed and undesired calculation paths. In one tool, a request for the number of bedrooms, rather than the number of occupants, in a house is an example of such an input.

Wordings of input questions can confuse or mislead users, resulting in inappropriate building description information and thus inaccurate results. For example, as noted above, many tools ask for “hours of operation” for various appliances and it is often unclear whether to provide annual or seasonal averages (in the case of space conditioning questions) or 24 h per day in the case of refrigerators. Several tools ask for annual operating hours for almost every end use including water heaters, furnace fans, and freezers. As another example, prediction of energy costs (bills) requires that the user make an accurate estimate of the weighted-average energy prices where complex tariffs are in effect.

4.4.1.6. Not all tools can be run in all climates. For example, in the case of the tools we examined, 10 of the 22 web-based tools and 5 of the 16 disk-based tools could not be run in the selected test-home cities.

4.4.1.7. The aforementioned factors conspire to confound comparisons among tools. Differences among inputs can range from weather city, to types of HVAC systems, to appliance characteristics, to occupant-driven effects, such as thermostat management. Differences in *results* would thus no doubt emerge from an extensive comparative exercise, but the *sources* or *implications* of these differences for the purposes of accuracy evaluation or tool development would remain largely unidentifiable (especially given the paucity of technical documentation available for most tools).

Another uncertainty associated with accuracy analysis is that different users would arrive at different results, given the many judgments entailed in describing a real home to a necessarily simplified tool.

Further complications apply in the case of bill-disaggregation tools. The question of whole-house “predictive” ability becomes moot, since such tools by definition agree with actual bills. In this case, the accuracy issue shifts to one of end-use predictive power, i.e. the correct allocation of total bills to actual end uses. As noted above, some bill-disaggregation tools exhibited problems when submitted to spot tests. The scarcity of good end-use data makes it difficult to validate such tools.

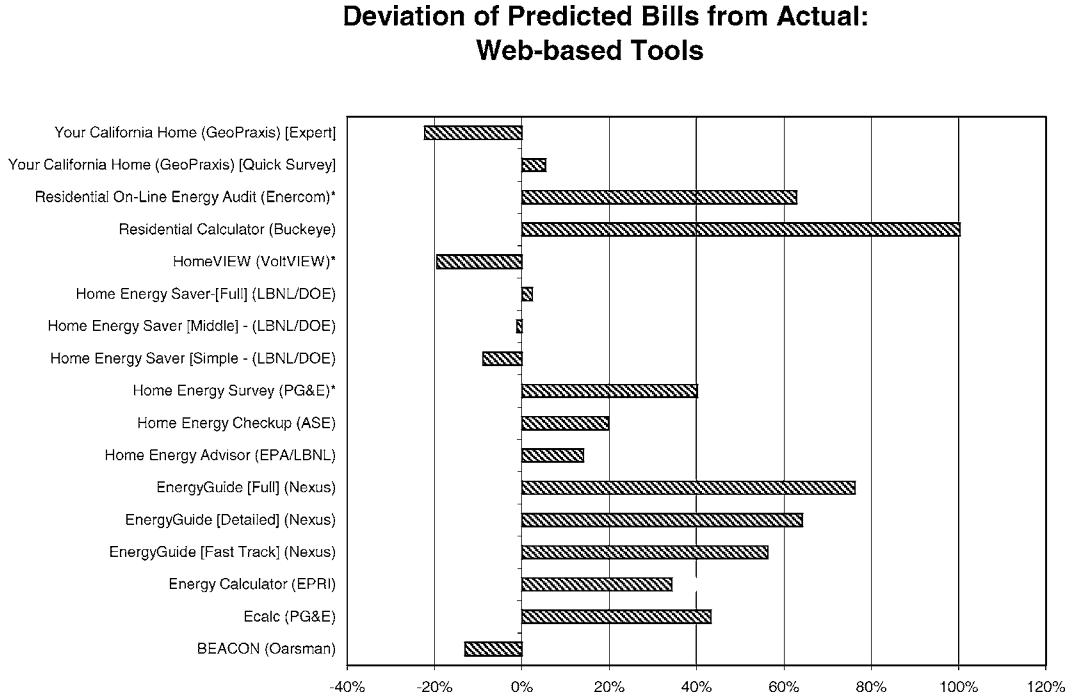


Fig. 3. Predicted vs. actual annual energy bills vary widely: web-based tools (California Test House). Notes: Actual: US$ 1179 per year (8-year weather average). Energy prices specified in the models identical to those in test home. Where applicable, bill disaggregation modules supplied only with August data. EnergyGuide: initial estimates were US$ 2566 (Fastrack) and US$ 3283 (detailed). Subsequent visit yielded lower outputs (shown here) for same inputs. PG&E: Subtotals disagree with grand total by 30%.

4.4.2. *Accuracy evaluation test case: web-based tools*

We evaluated those web-based tools offering an appropriate climate option for our first benchmark home (San Francisco Bay Area). All in all, 12 tools were included in this part of the accuracy evaluation (Fig. 3).

The results demonstrated considerable variability around the actual values, and differences among tools:

- Predicted energy bills varied from 25% below to 100% above the actual (US$ 1179 per year).

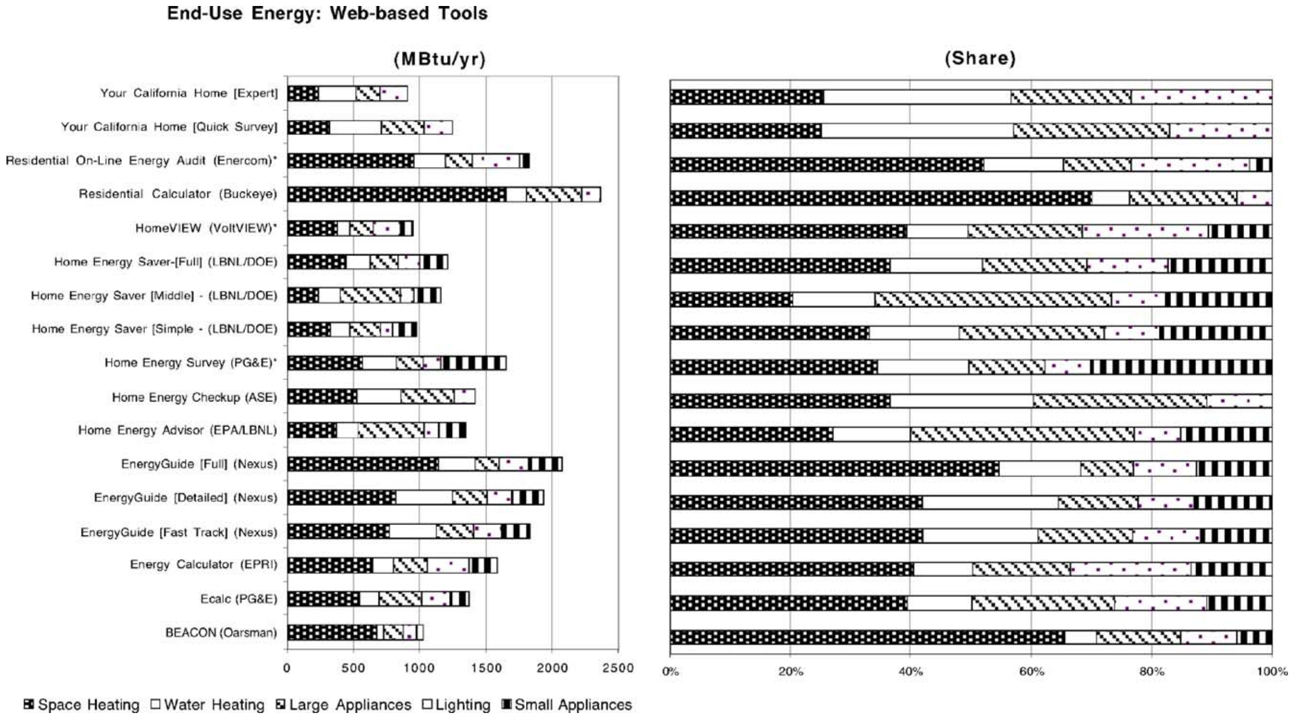


Fig. 4. Predicted energy use and end-use breakdowns vary widely: web-based tools.

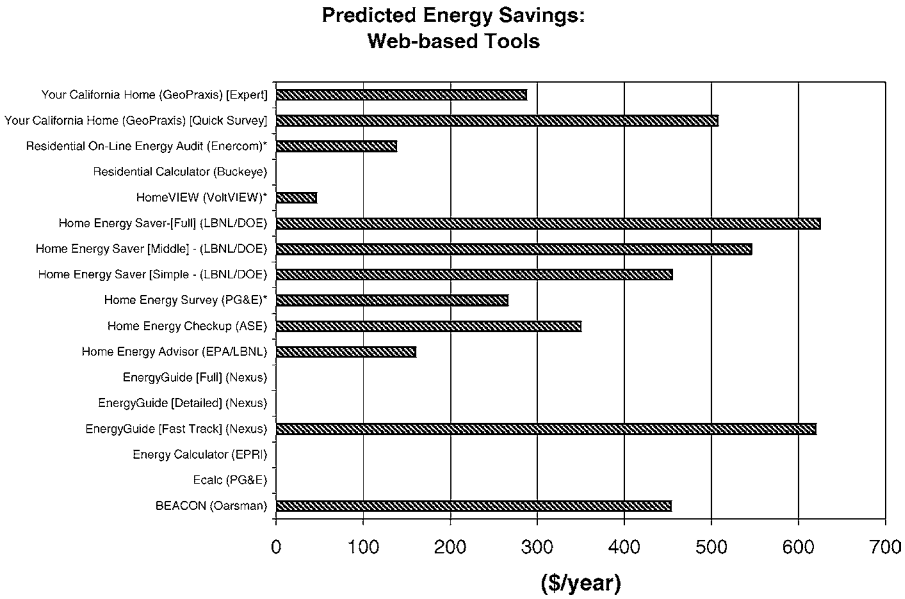


Fig. 5. Predicted annual energy savings defaults vary widely among the web-based tools.

- All tools over-predicted energy use by a significant margin (by up to a factor of 2.4). The variability was higher when examined at the end-use level, e.g. a factor of 8 in water heating energy and a factor of 7 for space heating energy (Fig. 4).
- Energy savings estimates automatically generated by the tools varied from US$ 46 per year (5% of predicted use) to US$ 625 per year (50% of predicted use) (Fig. 5). Each tool has a different set of decision rules for developing recommendations (often non-systematic and non-comprehensive), and thus the issue here is not one of accuracy as much as conveying vastly different information to consumers.

4.4.3. Accuracy evaluation test case: disk-based tools

Because of the limitations of demonstration versions and appropriate weather data, only six of the disk-based tools could be test run meaningfully with the second benchmark house, which was located in Ohio.

The results showed similar variability as seen for the web-based tools:

- Predicted energy bills varied from 2.1 to 2.4-fold above the actual (US$ 969 per year) (Fig. 6).
- All tools over-predicted energy use by a significant margin (by up to a factor of 2.8). The variability was higher when examined at the end-use level, e.g., a factor of 5.4 in air conditioning energy and a factor of 3.8 for water heating energy (Fig. 7).
- Design load predictions varied by factors of 1.5 for both heating and cooling (none of the web tools produce design load recommendations).
- None of the disk-based tools generate automatic retrofit recommendations.

Although sub-metered heating and cooling energy use was not available for the Ohio test house, detailed estimates of end-use energy consumption can be compared to the disaggregated utility data, and the results are somewhat disturbing (Fig. 7). In particular, the space-heating consumption is over-predicted by a factor of 4 or 5 across the board.

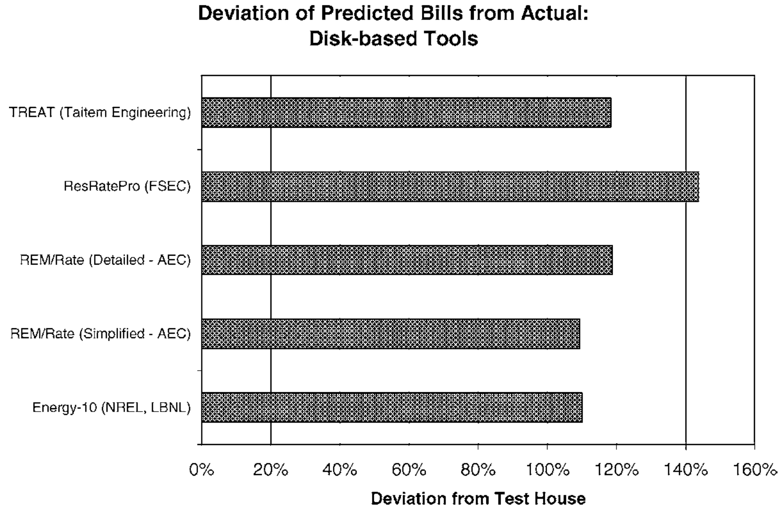
4.4.4. Caveats

Limitations of this exercise include the fact that only two buildings were studied, and without the full spectrum of potential end uses (one test house was located in a non-air-conditioning climate). Also, the analysis was performed by experienced modelers. Results for lay users are likely to exhibit even wider variability.

While Figs. 3 and 6 suggest that some tools appear to be more “accurate” than others, the many above-mentioned caveats apply. A readily apparent question is that of fortuitous agreement with actual bills as opposed to genuine accuracy. For example, the “middle’ version of the Home Energy Saver provides slightly “better” results than the “detailed” version. This is not because the former provides better modeling than the detailed tool, but rather that inaccuracies fortuitously cancelled out. Similarly, the Home Energy Checkup provides results relatively close to actual, however, this is highly fortuitous given that this tool is based on a very approximate “lookup” process using national survey data and highly aggregated climate zones. The test house, for example, has electricity prices a full 50% higher than the Home Energy Checkup’s (invariable) energy prices.

Note also that most results are above those of the actual test house bills. One would expect a more random distribution of over- and under-prediction.

Some web-based tools were not very stable, i.e. they delivered different results when the homes were rerun without



Note: Actual = $ 969/year Agreement = 0% deviation

Fig. 6. Predicted vs. actual annual energy bills vary widely: disk-based tools (Ohio Test House).

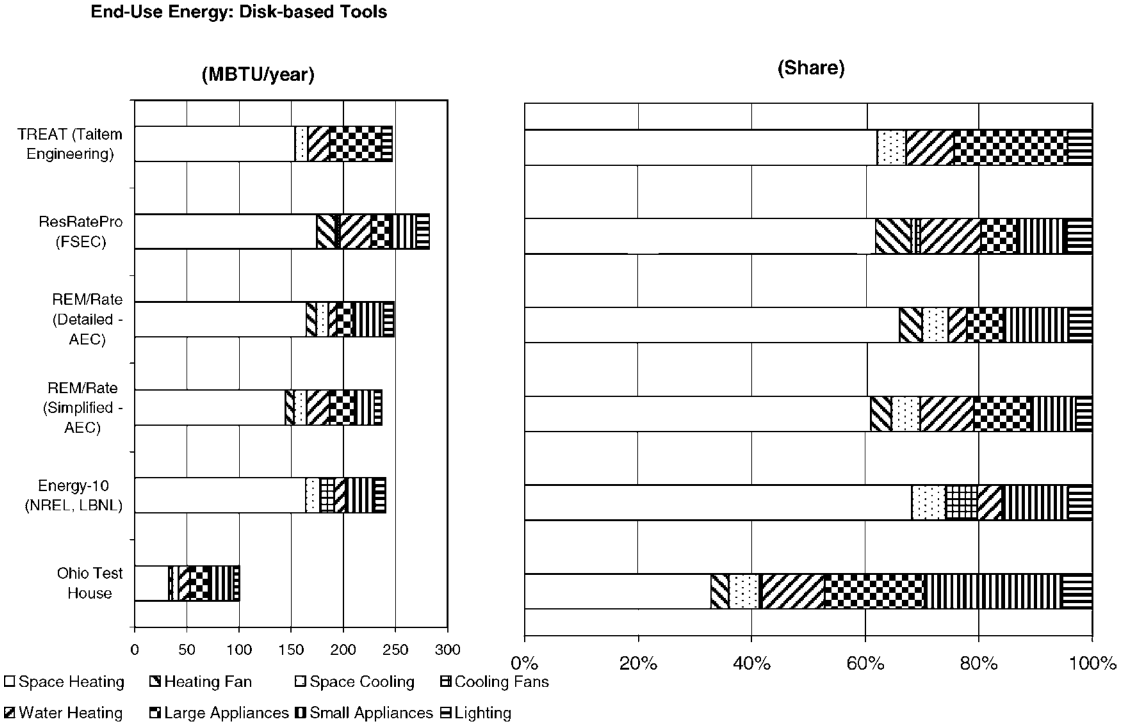


Fig. 7. Predicted energy use and end-use breakdowns vary widely: disk-based tools.

changing the building description or when saved runs were recalled subsequent to the initial session.

More comprehensive accuracy evaluations would require a statistically representative sampling of homes and climates, detailed measured end-use data (baseline and savings for a range of measures), highly flexible inputs (house size, window types, utilization patterns, etc.), relevant outputs. Very large numbers of runs would need to be conducted to examine an adequate array of combinations. Furthermore, complete fulfillment of the preceding list would make most of the existing tools ineligible for evaluation. Conversely, limiting such an experiment to the least common denominator required for all tools to qualify would result in such a highly "denatured" analysis.

5. Conclusions

The design of residential energy analysis tools should be grounded in social science as well as engineering, with close attention given to the intended use and audience. There are many potential avenues for improving the existing tools. Based on our review, we offer the following "best practices" design recommendations for consideration by tool developers.

5.1. *Targeting and usability*

We suggest carefully defining and addressing diverse audiences and their equally diverse needs, providing qualitative decision-support information (in addition to calculations), keeping underlying information and data current, fostering linkages among an every-growing proliferation of tools, and focusing on usability and convenience. Analytical results (e.g., benchmarking) and "what-if" capabilities are more helpful for many users than conventional engineering outputs. Web-based tools are of greatest use if user-entered data and results are saved for future sessions.

5.2. *Technical features and rigor*

Surprisingly, many of the tools only provide estimates of existing energy bills and no recommendations or estimates of potential savings, and fewer still address cost-effectiveness or emissions analysis (even superficially). Few tools offer substantial decision-support content. We also observed that energy analysis tools rarely keep pace with the forefront of building science research (e.g. thermal distribution modeling), and a greater effort should be made to do so. We suggest maximizing the applicable geographic range of tools (weather conditions), ensuring technical rigor (e.g., modeling of HVAC-appliance interactions) while providing for the modeling of occupant effects, open-ended energy calculations as well as results normalized to actual billing history, incorporating means for users to appreciate the uncertainties embodied in the results, and ensuring quality control to remove errors from the design and programming of tools. A comprehensive validation protocol is needed. The BestTest method, for example, is valuable but the focus is limited to building envelope and HVAC modeling [15].

5.3. *Platform*

Web-based tools offer considerable advantages over disk-based tools. Among these are platform independence, lower cost of distribution, ease of updates, and the ability to implement links to a growing array of related resources elsewhere on the internet. Powerful simulations can be located on a central server, lifting any requirement that the user's CPU can handle computation-intensive modeling.

5.4. *Strategic considerations*

Future efforts could encourage heightened objectivity, technical inclusiveness, and accuracy, and improved transparency and documentation of assumptions. Tremendous fragmentation and redundancy (as well as inconsistent analytical results) prevail among tools currently in use. Efforts should be made to unify existing disparate public and private development initiatives in order to focus scarce development resources into higher-quality and better-validated tools.

Acknowledgements

The research described in this article was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Program, US Department of Energy, under Contract No. DE-AC03-76SF00098. We also recognize the comments of an anonymous referee.

References

[1] V. Bazjanac, D.B. Crawley, in: Proceedings of Building Simulation '99 on Industry Foundation Classes and Interoperable Commercial Software in Support of Design of Energy-Efficient Buildings, vol. 21, 1999, pp. 661–667.

[2] H. Michaels, Comparing Approaches to Reducing Efficiency Market Barriers through Interactive Energy Audit Software, in: Proceedings of the 2002 ACEEE Summer Study on Energy Efficiency in Buildings, American Council for an Energy-Efficient Economy, Washington, 2002, p. 7.159.

[3] K. Papamichael, V. Pal, Barriers in Developing and Using Simulation-Based Decision-Support Software, in: Proceedings of the 2002 ACEEE Summer Study on Energy Efficiency in Buildings, American Council for an Energy-Efficient Economy, Washington, 2002, p. 7.179.

[4] E. Mills, R. Ritschard, "DOE-Sponsored Microcomputer Tools for Buildings Energy Analysis: Applications to Multifamily Retrofit Evaluation," Lawrence Berkeley Laboratory Report No. 24307, 1987.

[5] H. Hunter, Residential Energy Audits: Only A Website Away, Home Energy Magazine http://www.homeenergy.org/webaudit.intro.html, 1998.

[6] EPRI, An Evaluation of Web-Based Residential Energy Bill Disaggregation Software, Report TR-111192, 1998, 90 pp.

[7] J. Westerman, Home Energy Analysis Software Study: Final Report, SAIC, San Diego, CA, 2001, 95 pp.

[8] B.S. Wagner, "Verifications of Building Energy Use Models": A Compilation and Review, in: Proceedings of the ACEEE 1984 Summer Study on Energy Efficiency in Buildings, 14–22 August 1984, American Council for an Energy-Efficient Economy.

[9] ASHRAE, "Standard 140-2001—Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs (ANSI approved)", American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE), 2001, 91 pp.

[10] D.B. Crawley, Building Energy Software Tools Directory, World Wide Web site: http://www.eren.doe.gov/buildings/tools_directory/, US Department of Energy, Washington, DC. (The on-line version has been updated extensively since first published in 1998.), 1998.

[11] LBNL, Energy Crossroads Website, Energy Software Category. http://eetd.lbl.gov/cbs/eXroads/soft.html, 2002.

[12] Primen, "Utilities Hard-Pressed to Justify Investments in Residential-Customer Websites," http://www.primen.com/about/pr_utilities.asp, 2002).

[13] E. Mills, Review and Comparison of Web- and Disk-based Tools for Residential Energy Analysis, Lawrence Berkeley National Laboratory Report No. 50950, 2002.

[14] E. Mills, A. Rosenfeld, Consumer Non-Energy Benefits as a Motivation for Making Energy-Efficiency Improvements, in: Proceedings of the 1994 ACEEE Summer Study on Energy Efficiency in Buildings (also in *Energy—The International Journal*, 21 (7/8) (1994) 707–720), pp. 4.201–4.213.

[15] US Department of Energy, BESTEST Software. http://www.eren.doe.gov/buildings/tools_directory/software/bestest.htm, 2002.