Methodology to Assess No Touch Audit Software Using Simulated Building Utility Data

Howard Cheung and James E. Braun

Purdue University

M. Rois Langner

National Renewable Energy Laboratory

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Technical Report
NREL/TP-5500-66001
October 2016

Contract No. DE-AC36-08GO28308

Methodology to Assess No Touch Audit Software Using Simulated Building Utility Data

Howard Cheung and James E. Braun
Purdue University

M. Rois Langner
National Renewable Energy Laboratory

Prepared under Task No. BE4C.0201

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

Technical Report
NREL/TP-5500-66001
October 2016

Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at SciTech Connect http:/www.osti.gov/scitech

Available for a processing fee to U.S. Department of Energy
and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
OSTI http://www.osti.gov
Phone: 865.576.8401
Fax: 865.576.5728
Email: reports@osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312
NTIS http://www.ntis.gov
Phone: 800.553.6847 or 703.605.6000
Fax: 703.605.6900
Email: orders@ntis.gov

Cover Photos by Dennis Schroeder: (left to right) NREL 26173, NREL 18302, NREL 19758, NREL 29642, NREL 19795.

NREL prints on paper that contains recycled content.

List of Acronyms

ANSI	American National Standards Institute
BESTEST	Building Energy Simulation Test
CAV	constant air volume
COP	coefficient of performance
DOE	U.S. Department of Energy
EIA	U.S. Energy Information Administration
HVAC	heating, ventilating, and air-conditioning
IPMVPC	International Performance Measure and Verification Protocol Committee
LBNL	Lawrence Berkeley National Laboratory
LHS	Latin hypercube sampling
M&V	measurement and verification
NMSC	National Main Street Center, Inc.
NREL	National Renewable Energy Laboratory
NTHP	National Trust for Historic Preservation
PGL	Preservation Green Lab
PNNL	Pacific Northwest National Laboratory
Purdue	Purdue University
RESNET	Residential Energy Services Network
TMY	typical meteorological year
VAV	variable air volume

Nomenclature

α	Classification error (Type I error)
Δx_i	Deviation between correct value of a variable x and its corresponding estimate of the i^{th} data point [unit varies]
N	Number of data points
σ	Sample standard deviation
t	Statistics of Student t-distribution

Acknowledgements

The authors would like to acknowledge the owners and developers of the representative no-touch audit tool evaluated in this study, for their discussion and operation of the tool during the analysis of the building simulation results that made the project successful.

Executive Summary

The National Renewable Energy Laboratory (NREL) has a long history of developing analytical and empirical methods to test and diagnose building energy modeling tools and calibration techniques (Judkoff et al. 1995, 2010, 2016; Neymark et al. 2002, 2008a, 2008b, 2016). Some of these methods have become Standard Methods of Test through the American National Standards Institute (ANSI) and the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) (ANSI/ASHRAE Standard 140; ANS BSR/RESNET 1201-2016). This report describes a set of simulation based tests developed for assessing the performance of no touch building audit tools and presents results for a representative, available tool. This is an analytical approach to assessing the representative tool. A companion report (Cai et al. 2016) describes an empirically based set of tests derived from field data that was applied to the same no touch building audit tool. Analytical and empirical validation methods each have advantages and disadvantages. Using both approaches together provides a more complete testing and diagnostic process (Judkoff et al. 1983/2008, 2006).

Building audits are conducted in many commercial buildings to reduce building energy costs and improve building operation. Because the audits typically require significant input obtained by building engineers, they are usually only affordable for larger commercial building owners. In an effort to help small building and business owners gain the benefits of an audit at a lower cost, no touch building audit tools have been developed to remotely analyze a building's energy consumption. The America Saves project, led by the National Trust for Historic Preservation, uses a no touch building audit tool to analyze small commercial building utility data and provide energy efficiency solutions to small buildings in historic downtown areas across the United States. This tool (which will be referred to as the "Tool" for the remainder of the report) was used as a representative tool for developing a methodology to test performance of no touch building audit tools. The methodology is described in this report. At a high level, this Tool serves small buildings and businesses by:

- Disaggregating different types of energy end uses in buildings without expensive submetering
- Estimating the potential energy savings that can be achieved by retrofitting a building
- Diagnosing building operating issues.

There is a fine line between having a detailed modeling method that requires time and expertise, and a too-simplistic model that reduces fidelity of the results. This balance of robustness is important, especially for tools like the one assessed in this report whose target audience (i.e., small businesses and small commercial building owners) need quick, streamlined, and trustworthy information without requiring much technical expertise. While the above listed functions do not represent a comprehensive building audit, they help building owners to justify the cost and benefits of a more detailed building audit that may lead to building retrofits. A major concern of using these types of tools is their accuracy in performing an automated building audit. To assess the accuracy, NREL and Purdue University developed a methodology to assess no-touch audit tools by evaluating the Tool's three functions mentioned above. The evaluation methodology was designed to be replicable for other no touch building audit tools on the market, enabling others to evaluate additional tools in the future.

Because it is necessary to assess no-touch audit tools across many combinations of building types and potential causes of high-energy consumption, it would be costly to perform the evaluation through experimentation alone. Therefore, a simulation-based evaluation method was developed to facilitate a broad sampling of the parameter space with thousands of test cases. In a companion report, data was collected from fourteen real buildings to develop a set of empirical tests (Cai et al. 2016). Empirical and analytical test methods each have different advantages and disadvantages, and using them together results in a more complete testing process.

In this work, building simulation results were generated for typical commercial buildings (Deru et al. 2011; Thornton et al. 2011; Goel et al. 2014) under a variety of building types, vintages, climates, levels of technology, and operating issues. Simulated energy uses derived from the model results were then compared to estimates from the Tool to assess its accuracy in disaggregating energy end uses, estimating annual energy use, and estimating annual energy savings. The tool used twelve months of simulated electric and gas utility bill data along with simple building characteristic information as primary inputs to perform these functions. To quantify the accuracy of the energy end use estimates, the bias and uncertainty of the estimated annual energy uses, normalized energy end uses, and cost ratios of energy end uses are presented. The quantification of the accuracy of the energy savings estimates also involves the calculation of bias and uncertainty by comparing a pre- and post-retrofit scenario. True diagnosis and false alarm rates for the Tool's different alarms signaling probable causes of high energy use are also presented to assess the Tool's accuracy in diagnosing operating issues.

The project team created 2,952 sets of simulation results to evaluate the Tool. Each set of simulation results is a test case. The test cases were created based on five different types of buildings (small office, medium office, standalone retail buildings, primary schools, and main-street buildings), four vintages (addressing common construction techniques relevant to pre-1980, post-1980, 20014, and 2013 eras) and four different climate zones (2A, 4B, 4C, and 5A) under ASHRAE (2014). An example of the kinds of insights this testing process can provide is presented below from the representative tool that was tested.

- The results show that the Tool typically estimates electric end uses with a bias (over or under prediction) within ±5% and uncertainty between 2.5-15%, except for medium office buildings where the bias is within ±8% with an uncertainty of 18%.

- The results show that the Tool typically estimates gas end uses with a bias within ±5% and uncertainty between 1-2%, except for primary schools where the bias is closer to +8% with uncertainty of 3% due to more complex heating, ventilation, and air conditioning equipment.

- The end use cost ratios for primary schools and medium office buildings had uncertainties greater than 10%.

- The energy saving estimates had uncertainties greater than 20% for all building types. The Tool tends to underestimate energy savings except for in primary schools, where energy savings were overestimated by more than 10%.

- In terms of its accuracy in diagnosing building operating issues, it was found that the Tool was most accurate in diagnosing:
    - High summer gas use in standalone retail buildings with a true diagnosis rate of 73.6%,
    - Unnecessary reheat in medium offices with a true diagnosis rate of 50%, and
    - High electricity baseloads in office and main-street buildings, with true diagnosis rates ranging from 52.4% to 72.1%.
- False alarm rates for the Tool's message flags were generally higher than desirable for many building types.

Table of Contents

List of Figures

List of Tables

1 Introduction

No touch audit technologies offer a great opportunity to lower transactional costs in prioritizing buildings and businesses with high potential for energy savings, and in determining the best suited efficiency solutions to pursue. They are different from conventional audit processes and technologies in that they only need simple information to conduct a building audit, such as utility bills and floor area, which are readily available to building owners. They do not require the expense of hiring a building auditor or collecting submetered data such as hourly electricity consumption that requires extra instrumentation. The National Renewable Energy Laboratory (NREL) and Purdue University (Purdue) collaborated as part of the America Saves project team to develop a method for assessing no-touch audit tools. The method was used to evaluate a representative no touch audit technology (referred to as the "Tool" in this report).

The America Saves project strives to deliver energy efficiency solutions to small businesses in communities across the nation that aim to save energy, reduce utility costs, and boost profits (America Saves 2015). Led by the National Trust for Historic Preservation (NTHP) and its subsidiary, Preservation Green Lab (PGL), and funded in part by the U.S. Department of Energy (DOE), the America Saves project delivers energy efficiency support and solutions to small business communities by leveraging building and energy use data, remote analytics, and the NTHP's network of National Main Street Centers.

The National Main Street Center, Inc., (NMSC) is a national organization committed to historic preservation-based community revitalization (NMSC 2015). More than 2,000 communities are registered with the NMSC, and the America Saves project aims to leverage this network to expedite small business participation in energy retrofits and efficiency programs (America Saves 2015). As part of this effort, the America Saves project supports NMSC coordinators and community leaders in guiding their local small businesses through energy data collection, audits, feedback, and "no hassle" energy retrofits—stimulating employment growth and improving energy efficiency within their communities (America Saves 2015).

By applying utility data disaggregation and analytical tools (such as no touch audit technologies) to these smaller communities and "hard-to-reach" customers, America Saves helps utilities engage with small businesses and nonprofit customers to deliver cost-effective, scalable energy solutions (America Saves 2015). As part of this process, the America Saves project employs a no touch audit technology (referred to as the "Tool" in this report) that disaggregates energy end uses from monthly electric and gas utility data (Reichmuth and Turner 2010).

The Tool is used by the project to analyze portfolios of buildings and identify buildings that have potential for energy savings within a community. Utilizing expertise in building energy efficiency, energy modeling, and evaluation of fault detection and diagnostics tools, NREL and Purdue were brought on to the project as third-party reviewers to provide an objective, technical review of the Tool and assess its accuracy. Using this opportunity, NREL and Purdue collaborated to develop a robust and replicable process to evaluate this and other no touch audit technologies available on the market under a variety of circumstances.

This report extends previous work on model validation and diagnostics at NREL. The Building Energy Simulation Test (BESTEST) series of reports provided a way to test and diagnose the internal mathematics, physics and algorithms in building energy models and became the basis for ANSI/ASHRAE 140, *Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs* (Judkoff et al. 1995; Neymark et al. 2002, 2008a, 2008b, 2016). The BESTEST method uses a combination of analytical, comparative, and empirical tests (Judkoff et al. 1983/2008, 2006). The Building Energy Simulation Test for Existing Buildings (BESTEST-EX) reports provided a way to test models that use data from existing buildings to improve the disaggregation of energy end uses and the prediction of energy savings (Judkoff et al. 2010, 2011, 2016) That work became the basis for ANSI BSR/Residential Energy Services Network (RESNET) 1201-2016, *Standard Method of Test for the Evaluation of Building Energy Analysis Model Calibration Methods*. This report generally follows the methodological framework of ANS BSR/1201-2016. Simulations were performed to develop 2,952 test cases specifically designed for testing no-touch audit tools. The simulation results act as a surrogate for real building data. There are several advantages to this kind of analytical approach in addition an empirical field testing approach.

- The simulated data is by definition perfectly true and complete, so if there is disagreement between the tool outputs and the simulated data, it has to be because of the tool, not the data.
- The simulated data can be easily generated for a broad array of test cases providing a comprehensive test set.
- The simulated data can produce highly diagnostic test cases.

Empirical field testing is also important because it provides a bottom line test under conditions that are most similar to how a tool will actually be used in practice. Ideally, both kinds of tests will be available because each has pros and cons and together they provide a more complete testing process. Field data from 14 actual buildings was collected and used to further assess the Tool in a companion report (Cai et al. 2016).

Efforts were conducted by Lawrence Berkeley National Laboratory (LBNL) to develop a testing procedure and metrics to assess the performance of automated measurement and verification (M&V) tools (Granderson et al. 2015). However, automated M&V tools serve a different purpose and audience than no touch audit technologies; they predict energy savings after a retrofit, rather than identify probable causes of high energy use. Thus, the two assessment methods, although similar, are both necessary in assessing the different types of tools. Additionally, Lee et al. (2015) compared a number of no touch audit technologies based on their interface, requirements from users, target audience, and other features. However, the study did not develop a method to assess the accuracy of the tools. The suite of test cases presented in this report is unique and was developed in a way that can be adopted by other entities to assess additional tools on the market.

The no touch audit tool estimates building energy end uses and diagnoses probable causes of high energy use in commercial buildings. The building energy end uses include:

- Electric heating, cooling, and baseload energy consumption (baseload energy includes miscellaneous energy consumed by the building that is not associated with heating or cooling energy)
- Gas heating and baseload energy consumption.

Inputs to the Tool include monthly utility bill data and simple descriptive information that a nontechnical building owner could provide, such as the building floor area and occupancy. The Tool flags energy end uses that are higher than expected and offers potential solutions to reduce energy end use.

The evaluation was conducted by comparing the energy end use breakdown calculated by the Tool from simulated monthly utility data to an equivalent dataset that NREL and Purdue produced using advanced modeling techniques. Simulated utility bill data sets were provided as inputs to the Tool using EnergyPlus and OpenStudio modeling software for small and medium sized office buildings, primary schools, retail buildings, and main-street buildings (mixed use, with retail on the first floor and office on the second floor). The simulated data was developed using DOE's Commercial Reference Building Models (Deru et al. 2011) and the Pacific Northwest National Laboratory's (PNNL) Commercial Prototype Building Models (Thornton et al. 2011; Goel et al. 2014). These models account for different vintages (addressing common construction techniques relevant to pre-1980, post-1980, 2004, and 2013 eras), and four American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) defined climate zones (ASHRAE 2013a):

- 2A, hot-humid climate, represented by Houston, Texas
- 4B, mixed-dry climate, represented by Albuquerque, New Mexico
- 4C, mixed-marine, represented by Seattle, Washington
- 5A, cool-humid, represented by Chicago, Illinois.

OpenStudio Commercial Reference and Prototype Building Models were used to automate the model generation process. These models were developed by national laboratories using data from the Commercial Building Energy Consumption dataset, to represent typical commercial building performance. This technical report provides a detailed discussion of the replicable modeling and evaluation process used to assess the accuracy of the tool's ability to disaggregate energy end uses, to calculate potential energy savings, and identify probable causes of high energy use.

The remainder of this report is divided into eight sections with appendices. Those sections are summarized below:

- Section 2: An overview of the Tool
- Section 3: A general description of the tool assessment method
- Section 4: Creation of building simulation results
- Section 5: Evaluation methods and results of the accuracy of energy end use disaggregation
- Section 6: Evaluation methods and results of the accuracy of potential energy saving estimates
- Section 7: Evaluation methods and results of the accuracy of the Tool's message flags
- Section 8: Conclusions about the findings and future work

- Appendices: Details about building models, test matrices, bias and uncertainty calculations, and evaluation results.

2 Overview of the Representative No Touch Audit Tool

The representative no touch audit tool uses inverse modeling and energy signature analysis to gain insights into building performance (Reichmuth and Turner 2010). The Tool uses curve-fitting techniques to solve for key parameters that reproduce typical energy use patterns in relation to outside air temperature, providing a monthly, temperature-correlated view of energy consumption by fuel type. It has four major functions: provide a usage base case model, disaggregate energy end uses, estimate potential energy savings, and identify potential causes of high energy use. This type of tool differs from others because it uses billing data for all fuel types to devise a simple whole building energy model. The parameters of this model can be benchmarked. For scenarios where the Tool is looking at building energy use over time, those parameters allow for adjustment to the model as the building undergoes both routine and non-routine changes.

2.1 Literature Review on Inverse Modeling and Energy Signature Analysis Tools

Inverse modeling and energy signature analysis are documented methods that can help identify opportunities for energy efficiency in buildings by analyzing utility bills and climatic data together. A number of tools currently on the market incorporate inverse modeling and energy signature analysis into their capabilities—somewhat similar to the Tool, others have different goals. Lee et al. (2015) reviewed the features and capabilities of 18 energy retrofit toolkits and provided a comprehensive summary of their energy conservation measures and calculation engines. Many of these tools use empirical data-driven methods or benchmarking to identify buildings with potential for energy savings. The toolkits that use more sophisticated energy-modeling software have the ability to provide a more robust and detailed analysis of energy savings opportunities, but are often complex and require significant time and expertise. Lee et al. also concluded that there is a fine line between having a detailed modeling method that requires time and expertise and a too-simplistic model that reduces fidelity of the results. This balance of robustness is important, especially for tools like the one assessed in this report whose target audience (i.e., small businesses and small commercial building owners) needs quick, streamlined, and trustworthy information without requiring much technical expertise. The efforts explained in this report consider this balance.

ASHRAE Guideline 14 was developed to fill a need for a standardized set of energy (and demand) savings calculation procedures (ASHRAE 2014). The procedures calculate energy, demand, and water savings using measured pre- and post-retrofit utility billing data. They also encompass all types of facilities and apply to all forms of energy (ASHRAE 2015). Kissock and Mulqueen (2008) also describe a four-step method to identify building energy efficiency opportunities using advanced billing analysis. The method uses a four-step process to analyze 12 months of energy data by fuel type and normalizes it to weather data. Regression models use three parameters: (1) weather independent energy use, (2) heating or cooling slopes, and (3) the balance point temperature. The weather-normalized output can then be used to benchmark a building's energy use to multiple buildings, thus enabling the average, best, and worst energy performers to be identified. It also can show how building performance changes over time.

Additionally, Abels et al. (2011) describes an inverse modeling method using regression models of utility billing data against weather and industrial production data. The models described here are incorporated into the ASHRAE Inverse Modeling Toolkit (Kissock et al. 2001) and are recommended by the International Performance Measure and Verification Protocol Committee (IPMVPC) (NREL 2002).

In the field of inverse modeling, regression fits are typically applied to only one fuel at a time, and usually a fit with statistical parameters, not engineering parameters (i.e., a correlation, not an engineering model). The Tool is different than other tools described in the literature because it fits a model of a building considering all fuel types to the billing data for all fuel types of a building.

The Tool's approach uses calibrated and aggregated engineering parameters to describe a building model that considers all fuel types, and is fitted simultaneously to the billing data for all fuel types of a building. The presumed advantage of expressing the building model in terms of engineering parameters is that these parameters can be benchmarked and lend themselves to estimating the energy use for altered states if the building undergoes change.

2.2 End Use Disaggregation

The Tool disaggregates 12-month utility bill data for different fuel types. The energy signatures are derived from curve-fitting techniques that solve for key engineering parameters, reproducing observed energy-use patterns in relation to average monthly weather data (Reichmuth and Turner 2010). The Tool uses the reproduced observed energy-use patterns to create an "equivalent analog building model." The engineering parameters from this model are compared to typical engineering parameters derived from a large sample of analyzed buildings to benchmark building energy use, to determine probable causes of high energy use, and to estimate energy savings. The primary independent variable in these models is the average monthly temperature, with seasonal changes being the primary driver.

The Tool uses a proprietary set of algorithms and equations to identify key parameters of the reference model. The key parameters were chosen by considering the energy balance of a building, engineering expertise, and experimentation with several alternative variable sets, and helps solve for the building energy end uses (Reichmuth and Turner 2010). The key parameters are listed in Table 1.

Table 1. Key Parameters Used by the Tool To Determine Building Energy End Uses

Parameter	Method to Estimate the Values
Internal electricity use by lighting and equipment per floor area	Solved by proprietary algorithms
External electricity use per floor area	Fixed ratio of internal electricity use
Aggregate heat transfer conductance	Solved by proprietary algorithms
Heating equipment efficiency	Assumed to be 75%
Cooling equipment coefficient of performance (COP)	Solved by proprietary algorithms
Rate of service hot water use	Solved by proprietary algorithms
Maximum ambient temperature with heating operation	Solved by proprietary algorithms
Minimum ambient temperature with cooling operation	Solved by proprietary algorithms
Gas heat to electric heat ratio	Solved by proprietary algorithms

Upon solving for the key parameters in Table 1 the Tool uses the parameters in the analog building model to estimate annual building energy end uses for the reference model, including:

- Electric heating, cooling, and baseload energy consumption
- Gas heating and baseload energy consumption.

To aid the analysis, the Tool considers additional descriptive information that is simple enough for a nontechnical building owner to provide. These additional inputs include:

- Building location
- Year of construction
- Overall floor area
- Space types and associated floor areas
- Number of occupants and percentage of occupied floor area
- Occupied hours per week
- Amount of space without temperature control
- Amount of space subjected to heating and cooling equipment control
- Space heating equipment type (gas furnace, heat pump, etc.)
- Type of energy for heating water (electricity, gas, etc.).

2.3 Estimation of Potential Energy Savings

The Tool uses the engineering parameters of the reference model to estimate energy savings. This process is different than processes used by automated M&V tools, as it compares parameters derived from 12 months of utility data to assumed typical values to estimate potential electricity and gas savings. In contrast, automated M&V tools use historical interval data (at much higher resolutions such as one minute time steps) to predict how much energy a building would have consumed if it had not undergone a retrofit. This predicted energy consumption is then compared to the actual energy consumption of the building that had undergone a retrofit, to calculate real energy savings. Automated M&V tools provide a more cost-effective and timely way to calculate energy savings after a building retrofit (Granderson et al. 2015).

The Tool can also estimate potential energy savings by benchmarking building energy use to a dataset of typical engineering parameters for similar building types. Benchmarking, however, requires a database of building parameters for comparison, which can be difficult to assemble in certain situations. Generally speaking, the savings estimate quantitatively outlines the potential energy and cost savings benefit that can be achieved through a building retrofit.

2.4 Identification of Probable Causes of High Energy Use

The Tool incorporates a number of automated messages (message flags) that are triggered when indicators within the benchmarking process suggest probable causes of high energy use. The message flags, denoted by a letter of the alphabet, are described in Table 2.

Table 2. Description of the Tool's Message Flags

Flag	Description	Definition
A	Low electric baseload energy	The electricity baseload per floor area is estimated in the reference model to be below 9.15 W/m^2 (0.85 W/ft^2)
B	Slightly high electric baseload energy	The electricity baseload per floor area is estimated in the reference model to be between 9.15 W/m^2 (0.85 W/ft^2) and 14.53 W/m^2 (1.35 W/ft^2)
C	Very high electric baseload energy	The electricity baseload per floor area is estimated in the reference model to be between 14.53 W/m^2 (1.35 W/ft^2) and 29.06 W/m^2 (2.7 W/ft^2)
O	Ultra-high with high internal electricity baseload	The electricity baseload per floor area is estimated in the reference model to be higher than 29.06 W/m^2 (2.7 W/ft^2) and associated with high internal gains
P	Ultra-high with high external electricity baseload	The electricity baseload per floor area is estimated in the reference model to be higher than 29.06 W/m^2 (2.7 W/ft^2) and associated with external gains
D	Inefficient shell and ventilation	Air infiltration or heat transfer across the exterior envelope is higher than expected for the building
E	Inefficient cooling equipment	Cooling efficiency seems to be lower than expected
M	Inefficient heating equipment	Heating efficiency seems to be lower than expected
H	High relative heat use	Space conditioning is running more than necessary to meet heating requirements
I	High relative cooling use	Space conditioning is running more than necessary to meet cooling requirements
K	High summer gas use	Gas uses is much higher than anticipated during summer months
L	Inappropriate operation of electric heat	Unnecessary use of reheat during the cooling season
Q	Low relative heat use	Energy used to heat a building is lower than expected
N	Possible erratic operation	Potential errors in utility bill data or data input

3 Assessment Methodology for No Touch Audit Tools

3.1 Overview

A methodology was developed to assess no touch audit tools, using the representative no touch audit tool as an example. To evaluate tool performance, the tool's estimates of energy end uses were compared to a corresponding set of simulated values to quantify the accuracy. Simulated results were created to represent realistic monthly utility and building end use values for comparison with the Tool estimates. The procedure is illustrated in the flow chart in Figure 1, and was managed using programming scripts written in R, which is a language and environment for statistical computing and graphics.

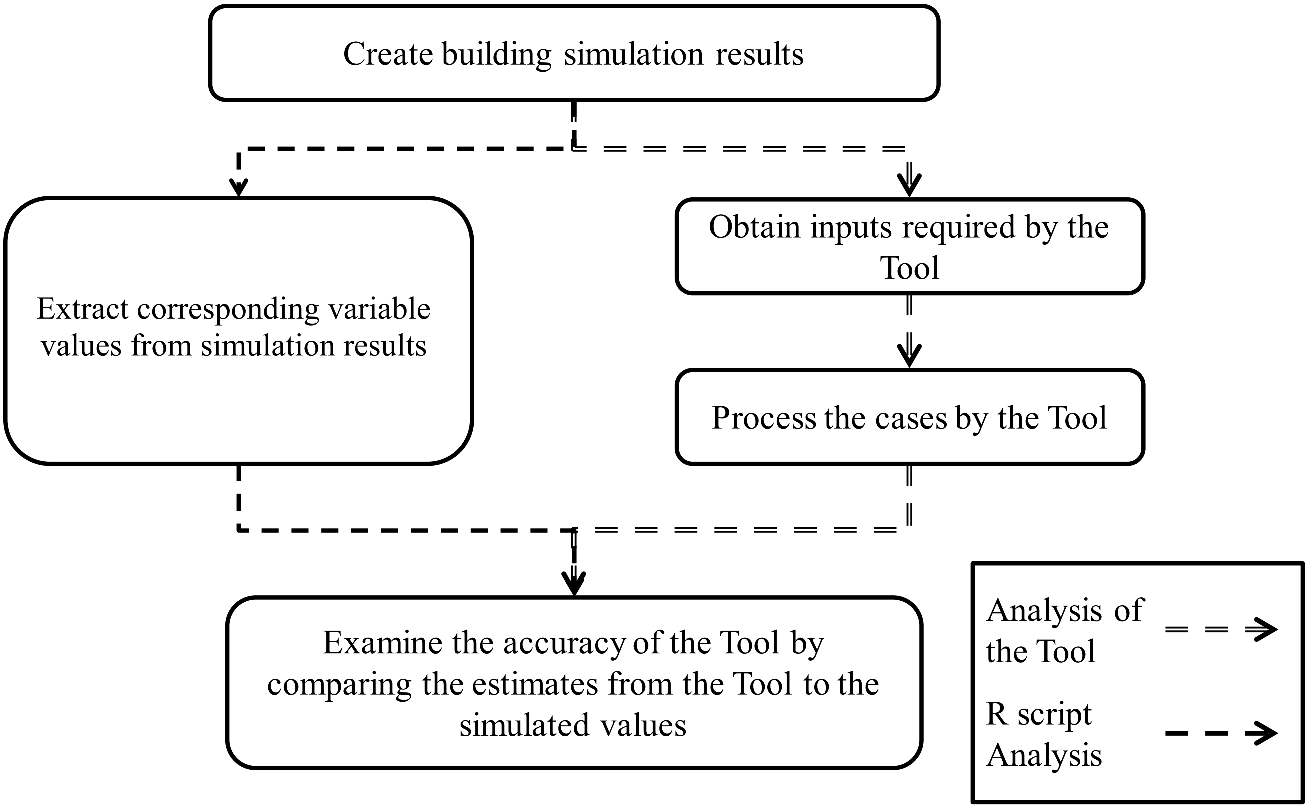


Figure 1. Flow chart providing an overview of the NREL evaluation procedure

To evaluate the performance of the Tool fairly and reliably, the building simulation results in Figure 1 satisfied criteria listed below. The simulation results:

- Considered the major types of buildings that are applicable to both the Tool and the America Saves project
- Used building models that represent typical buildings in the United States
- Considered both old and new building construction practices and heating, ventilation, and air conditioning (HVAC) technologies
- Considered all building operation issues that can be diagnosed by the Tool.

In order to satisfy these criteria, building models representing various types of typical buildings found across the United States, with different vintages and multiple HVAC technologies, were created as baseline models. The baseline models were then perturbed to create a database of simulation results with various operating conditions and issues. The database was then used to evaluate the accuracy of the Tool's end use disaggregation and potential savings estimates. A flow chart illustrating the procedure to create the simulation results is shown in Figure 2.

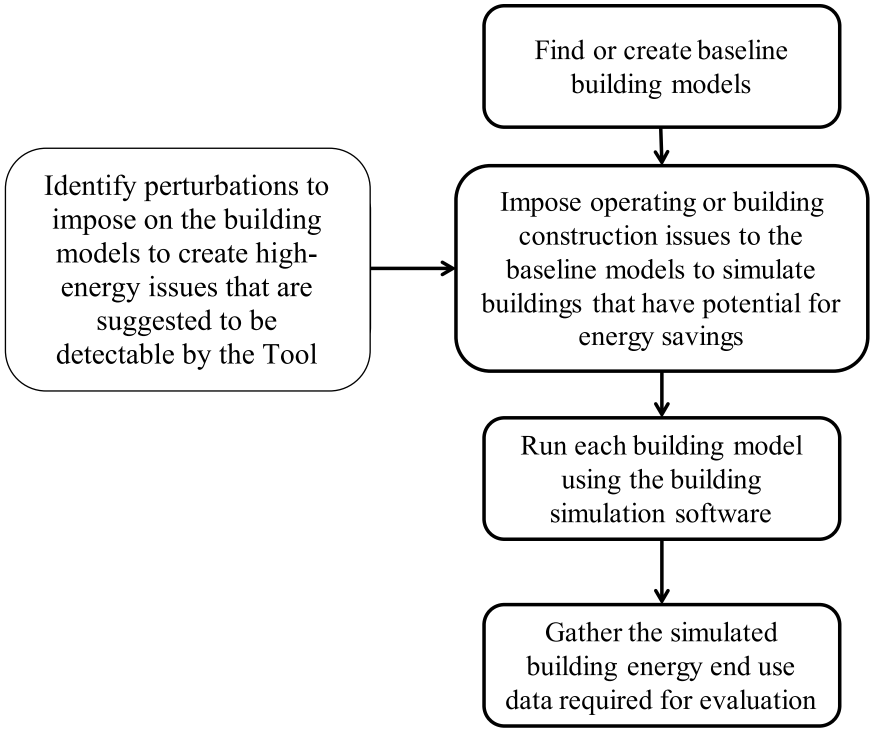


Figure 2. Flow chart illustrating the procedure to create the building simulation results

The building model simulations were executed using EnergyPlus, and the workflow and perturbations were managed with Ruby programming scripts in OpenStudio. The baseline models were mainly selected from the Commercial Reference Building Models and the Commercial Prototype Building Models. The baseline model for the main-street building type was unavailable from these two building model sets, but was derived by combining building attributes from the retail and small commercial building models. Annual simulations were carried out for each case, and monthly gas and electric totals were extracted to generate the inputs required by the Tool.

For each simulation scenario, energy end uses, potential energy savings, and building operating issues were identified from the energy simulation results. These values were considered to be the correct values that were compared to the estimated results from the Tool.

3.2 Software Tools

EnergyPlus

EnergyPlus (DOE 2015a) is an open-source building simulation program that estimates building energy use based on the building envelope, building occupancy, HVAC equipment, water use, weather, etc. It also estimates building performance metrics such as water consumption, ventilation, occupant thermal comfort, and life cycle costs, to facilitate overall building design and performance optimization.

OpenStudio

OpenStudio (DOE 2015b) is a platform that enhances the functionality of EnergyPlus by using object-oriented programming to modularize the component models in EnergyPlus. The OpenStudio graphical user interface helps users create EnergyPlus building simulation files more easily than older methods using the EnergyPlus input file editor IDFEditor. The modularization also allows users to develop their own Ruby programming scripts (OpenStudio and EnergyPlus Measure scripts) to change building models programmatically and automate parametric studies of building performance with EnergyPlus. The platform can also be connected to a cloud service (OpenStudio Server), which allows users to conduct large-scale parametric studies or building optimization without having their own supercomputer infrastructure.

Ruby

Ruby (Ruby 2015) is one of the high-level object-oriented programming languages that can be used to interact with EnergyPlus component models. Its high-level nature allows nonprogrammers to learn to use it easily, and its object-oriented nature allows users to programmatically change building models without manually and repeatedly looking up component models within several thousand lines of text in the EnergyPlus input files. In this project, it was used to write the EnergyPlus Measure scripts that were used to perturb the baseline building models with various building operating issues from the baseline models.

R

R (The R Foundation 2015) is a programming language that is designed for statistical analysis, data manipulation, and more. Because of its effectiveness in managing large data sets and its capability to run optimization and experimental design algorithms, it was used as the programming language to manage the OpenStudio Server. This project also requires similar capabilities for its data post-processing, so R was used to create the post-processing scripts to analyze the building simulation results from the OpenStudio Server.

Commercial Reference Building Models

The Commercial Reference Building Models (Deru et al. 2011) are a set of EnergyPlus building models that were developed by the national laboratories to model the energy consumption and thermal comfort of typical U.S. commercial buildings that were constructed before 2004. They include models for various building types, such as small offices and schools, in various climate zones and building vintages. Research on a multitude of building properties was conducted to ensure that the model parameters represented the typical buildings found in the United States. Since the Commercial Reference Building Model development, the reference building library has been widely used for a number of applications such as assessing the effectiveness of fault detection and diagnostics tools (Henze et al. 2015), optimal building design methods (Zhang et al. 2012), and sensitivity studies for building energy consumption (Hygh et al. 2012). These models were used in this project to represent typical commercial buildings built in the United States before 2004.

Commercial Prototype Building Models

The Commercial Prototype Building Models (Thornton et al. 2011; Goel et al. 2004) are a set of EnergyPlus building models developed based on the Commercial Reference Building Models in order to simulate the energy consumption and thermal comfort in typical commercial buildings that satisfy the energy codes published after 2004. Similar to the Commercial Reference Building Models, these models are used to study building energy consumption under different scenarios such as changes of building energy consumption in recent decades (Hong et al. 2013), the effect of urban heat island effects on buildings (Sun et al. 2014), and the effect of different insulation materials on building performance (Shrestha et al. 2014).

4 Creation of Building Simulation Results

4.1 Baseline Building Models

According to Figure 1 and Figure 2, the creation of the simulated building energy data started with the development of the baseline building models. In order to evaluate the Tool fairly, the baseline models covered a range of building types, vintages, and climate zones that are applicable to both the Tool and the America Saves project. The Commercial Reference Building Models and the Commercial Prototype Building Models were used for the baseline models, except for the main-street building models. The main-street building model was created by OpenStudio, combining building attributes from the retail and small commercial building models. In total, 2,952 building simulations were created for this study. This does not consider every combination; it was necessary to limit the number of cases to avoid infeasible time for evaluation. The process to reduce the number of cases is explained in the following sections of this report.

Building Types

The building types considered for this study are representative of small commercial buildings that are common in main-street communities across the United States. An overview of the building types, with the exception of main-street buildings, is listed in Table 3. It should be noted that the primary school models exceed the definition of "small;" however, this building type is still prevalent in the small building sector and the Tool developers had the most experience with this building type.

Table 3. Building Types Selected from the Commercial Reference Building Models and the Commercial Prototype Building Models

Building Type	Floor Area	Number of Floors
Small office	511 m^2 (5,500 ft^2)	1
Medium office	4,982 m^2 (53,628 ft^2)	3
Standalone retail	2,294 m^2 (24,962 ft^2)	1
Primary school	6,871 m^2 (73,960 ft^2)	1

The building models for each building type are briefly described in Appendix A, but readers should read Deru et al. (2011), Thornton et al. (2011), and Goel et al. (2014) for a more comprehensive description of specific input parameters used in the models.

Because the Commercial Reference Building Models and the Commercial Prototype Building Models don't include models for main-street buildings typically found in historic downtown areas (a major focus of the America Saves project), a main-street building model was created by combining baseline models for the retail and small office buildings. The building simulates a two-story historic downtown building with a total floor area of 1,191 m^2 (12,820 ft^2). It includes a retail store on the ground floor and office space on the second floor. Additional details are provided in Appendix B.

Climate Zones

In order to evaluate the Tool against a range of U.S. climate zones, four climate zones were chosen to represent the major climatic characteristics that are found across the United States. These climate zones are listed in Table 4.

Table 4. Climate Zones, Their Characteristics, and Their Representative Cities

Climate Zones	Characteristics	Representative Cities
2A	Hot and humid	Houston, Texas
4B	Mixed and dry	Albuquerque, New Mexico
4C	Mixed and marine	Seattle, Washington
5A	Cool and humid	Chicago, Illinois

The climate zone characteristics follow the classification system in ASHRAE Standard 169 (ASHRAE 2013a), and the representative cities are mapped to the representative cities found in the Commercial Reference Building Models. The representative cities determine the typical meteorological year (TMY) data used in the baseline models. The TMY3 weather data set was used (Wilcox and Marion 2008). The modeled building materials and HVAC equipment vary appropriately per climate zone. Details of these differences are described in Appendices A and B.

Vintages

Building vintage also affects the simulated building materials, HVAC equipment, electric and gas load density, etc., in each model. Similar to climate zones, the Commercial Reference Building Models and Commercial Prototype Building Models contain models for seven different vintages and it was infeasible to use all of them. To represent the buildings constructed in different years, the building models with vintages pre-1980 and post-1980 were selected from the Commercial Reference Building Models and the building models with vintages 2004 and 2013 were selected from the Commercial Prototype Building Models. The spectrum of the four vintages from the pre-1980 era to 2013 should represent the building construction practices found in those different years respectively.

For the main-street building models that are not part of the Commercial Reference Building Models and the Commercial Prototype Building Models, it was assumed that they were built before 1980 and their building materials were modeled in reference to the pre-1980 Commercial Reference Building Models. However, to create a baseline model representing a retrofitted main-street building, the HVAC equipment and electric load density were modeled in reference to both the pre-1980 Commercial Reference Building Models and the 2013 Commercial Prototype Building Models. Hence the pre-1980 vintage for main-street buildings represents an original main-street building, and the 2013 vintage represents a “retrofitted” main-street building. Additional details about the baseline main-street building models are described in Appendix B.

Baseline Model Summary

In summary, 64 baseline models, including four vintages and four climate zones, were used for small office, medium office, standalone retail, and primary schools. These baseline models were pulled from the Commercial Reference Building Models and the Commercial Prototype Building Models. Eight main-street building models, including two vintages and four climate zones, were also developed, resulting in a total of 72 baseline models.

4.2 Baseline Model Perturbations

To evaluate the accuracy of the Tool fairly, the building simulation results incorporated numerous operating issues, simulating high energy use that aligned with the Tool's message flags. The issues were created by changing the baseline model in accordance with the description of the message flags in the Tool documentation. To ensure that the model perturbation had a significant change to the simulated energy use and energy cost, the building simulation results were examined to quantify the significance of the change. The sample size of the number of cases with issues was also examined to ensure that the dataset was sufficient for a statistically significant evaluation.

Types of Building Model Perturbations Relevant to the Tool

In order to identify parameters within the model that, if perturbed, would trigger the message flags in Section 2, the HVAC equipment, shell efficiency, and internal electric loads were examined in each baseline model. The selected changes are listed in Table 5.

Table 5. Perturbations Applied to the Baseline Models

Type of Perturbation	Expected Effect on the Building Model	Potential Flags To Be Triggered	Applicable Baseline Models
Change in electric equipment power density	Change in the internal electric load of the building	Flags related to different electricity baseload levels (A, B, C, and O)	Small and medium office
Offset in the thermostat heating set point	Increase or reduction of the heating load of the building	Flags related to high heating use, low heating use, high summer gas use, or use of reheat (H, K, L, and Q)	All
Offset in the thermostat cooling set point	Increase in the building cooling load	Flags related to high cooling use (I)	All
Increase in external lighting	Increase in the external electric load of the building (multiplier applied to total electric load)	Flags related to high external electricity consumption (P)	All

Type of Perturbation	Expected Effect on the Building Model	Potential Flags To Be Triggered	Applicable Baseline Models
Increase in hot water consumption	Increase in the use of energy by the natural gas or electric hot water heater	Flags related to high summer gas use for buildings with a natural gas water heater or high electricity baseload for buildings with an electric water heater (K, B, C, and O)	All except standalone retail in pre-1980 and post-1980 vintages
Increase in infiltration airflow	Increase in infiltration and heat transfer between the building and the surroundings	Flags related to an inefficient shell and ventilation (D)	All
Reduction in gas furnace efficiency	Low efficiency gas furnace	Flags related to low heating efficiency for buildings with gas furnaces (M)	All
Increase in HVAC equipment size	Increased cycling operation of air conditioners and heat pumps	Flags related to inefficient cooling coefficient of performance or low heating efficiency for buildings with heat pumps (E and M)	All
Change in internal lighting power density	Change in the internal electric load of the building	Flags related to different electricity baseload levels in buildings with significant electricity plug-load power densities (A, B, C, and O)	Standalone retail, primary school, and main-street building
Reduction in cooling equipment rated coefficient of performance	Low efficiency cooling equipment	Flags related to inefficient cooling coefficient of performance for buildings with low operation of HVAC equipment cycling (E)	Primary school

Table 5 shows that all message flags besides flag N (denoting possible erratic operation or inconsistency in utility bill reporting) can be triggered by changing the baseline model. The process to trigger the flags is discussed at the end of this section. The A, B, C, and O flags (related to internal electric baseload levels) are triggered differently for each building type because the significance of the perturbations to building performance differs between building types. The hot water consumption in the pre-1980 and post-1980 standalone retail models is negligible, and thus was not increased. Flag E (inefficient cooling equipment), in the primary school models was triggered differently than other building models, because the primary school models have variable-air-volume (VAV) and two-speed cooling systems. With VAV and two-speed cooling systems, it was too difficult to cycle the cooling equipment on and off in the model, enough to significantly reduce the COP. Thus, the rated COP value was directly reduced to trigger Flag E in the primary school models. The quantitative values of the perturbations in Table 5 are tabulated in Table 6, Table 7, Table 8, and Table 9.

Table 6. Perturbation Values Applied to the Small and Medium Office Building Models

Changes to electric equipment plug-load power densities	-50%, +100%, +200%, +400%, +600%
Thermostat heating set point offset	-3K, -2K, +1K, +2K, +3K (-5.4°R, -3.6°R, +1.8°R, +3.6°R, +5.4°R)
Thermostat cooling set point offset	-2K, -1K (-3.6°R, -1.8°R)
Increase in external lighting	-50%, +200%, +400%, +700%, +900%
Increase in hot water consumption	+200%, +400%
Increase in infiltration airflow	+200%
Reduction in gas furnace efficiency	-50%, -25%
Increase in HVAC equipment size	+50%, +100%

Table 7. Perturbation Values Applied to the Standalone Retail Building Models in Pre-1980 and Post-1980 Vintages

Changes to internal lighting	-80%, -60%, +100%, +200%, +400%
Thermostat heating set point offset	-3K, -2K, +1K, +2K, +3K (-5.4°R, -3.6°R, +1.8°R, +3.6°R, +5.4°R)
Thermostat cooling set point offset	-2K, -1K (-3.6°R, -1.8°R)
Increase in external lighting	-50%, +200%, +400%, +700%, +900%
Increase in infiltration airflow	+200%
Reduction in gas furnace efficiency	-50%, -25%
Increase in HVAC equipment size	+50%, +100%

Table 8. Perturbation Values Applied to the Standalone Retail Building Models (2004 and 2013 Vintages) and the Main-Street Building Models (All Vintages)

Change in internal lighting	-80%, -60%, +100%, +200%, +400%
Thermostat heating set point offset	-3K, -2K, +1K, +2K, +3K (-5.4°R, -3.6°R, +1.8°R, +3.6°R, +5.4°R)
Thermostat cooling set point offset	-2K, -1K (-3.6°R, -1.8°R)
Increase in external lighting	-50%, +200%, +400%, +700%, +900%
Increase in hot water consumption	+200%, +400%
Increase in infiltration airflow	+200%
Reduction in gas furnace efficiency	-50%, -25%
Increase in HVAC equipment size	+50%, +100%

Table 9. Values of Perturbations to the Primary School Models

Changes to internal lighting	-80%, -60%, +100%, +200%, +400%
Thermostat heating set point offset	-3K, -2K, +1K, +2K, +3K (-5.4°R, -3.6°R, +1.8°R, +3.6°R, +5.4°R)
Thermostat cooling set point offset	-2K, -1K (-3.6°R, -1.8°R)
Increase in external lighting	-50%, +200%, +400%, +700%, +900%
Increase in hot water consumption	+200%, +400%
Increase in infiltration airflow	+200%
Reduction in gas furnace efficiency	-50%, -25%
Increase in HVAC equipment size	+50%, +100%
Reduction in cooling equipment COP	-20%, -10%

Design of Test Matrices

Test matrices of perturbation values were created for different combinations of perturbations that were applied to the baseline models. This helped to create additional cases that contain combinations of flags that are not represented in Table 5, such as D and H, and to evaluate whether the Tool is accurate in identifying high energy issues in these scenarios. For a comprehensive evaluation, two types of test matrices were constructed for each building type—a test matrix based on individual perturbations, and a test matrix with combinations of perturbations created using the statistical Latin hypercube sampling (LHS).

LHS is inspired by the Latin square experimental design, which aims to eliminate confounding effects of multiple experimental factors without increasing the number of experimental cases (Cheng and Druzdzel 2000). LHS uses an even sampling method to ensure that each value (or range of values) is represented evenly within the samples, no matter which value might turn out to be more important (Cheng and Druzdzel 2000). The resulting set of random inputs yields a very smooth distribution that minimizes the number of runs to obtain an accurate distribution of output variables (Langner et al. 2014). In this situation, LHS calculates the combination of perturbations and the values for each perturbation. However, because the programming scripts that offset the heating and cooling thermostat set points cannot be implemented simultaneously, two test matrices using LHS were created to offset the set points independently. For example, individual perturbations were applied to the primary school baseline model (according to Table 9), and the LHS algorithm created two test matrices of eight cases representing the perturbation combinations.

Summary of Test Matrices

After creating the test matrices for each baseline model, the resultant number of cases for each building type is shown in Table 10.

Table 10. Number of Simulation Results for Each Building Type

Building Type	Number of Cases
Small office	624
Medium office	624
Standalone retail	624
Primary school	688
Main-street building	312

The total number of cases in Table 10 is 2,872, which is close to the maximum number of simulation results that the Tool developers and NREL agreed to evaluate. The simulation scenarios were submitted to the OpenStudio Server for simulation, and the results were post-processed using R software. The details of the resultant test matrices for each baseline model are described in Appendix C.

Criteria for Evaluating Message Flags

Although the test matrices are intended to trigger the Tool's message, the perturbations do not guarantee that the message flags will be triggered. For example, if the heating and cooling loads of a building are not affected by infiltration, an increase in infiltration may not be significant enough to alter the energy consumption and trigger a flag. Conversely, if an increase in infiltration significantly increases the heating and cooling load and raises the energy cost for a building owner, a flag should be triggered. To ensure that the triggered flags are significant to a building owner, the simulation cases were evaluated with the following criteria:

- Is there is a significant operating issue related to the perturbation?
- Is the annual building energy cost significantly changed by the perturbation?

For example, a model with increased infiltration would be examined to see if the heat transfer rates and energy costs were increased significantly by the perturbation. If the perturbation does not change the energy cost significantly, the infiltration problem should not be communicated to the building owners.

The simulation results are checked by the two criteria qualitatively, and the mathematical criteria are listed in Appendix D. However, message flags (A, B, C, O, and P) are classifications of electric baseload energy, are defined quantitatively, and it is not necessary to use the above criteria to check if the Tool should raise the message flags. To determine whether a simulation result should be associated with flag A, B, C, O, or P, the electricity baseload per floor area is calculated and the result is classified according to the flag definitions in Section 2. Further details of this classification process are described in Appendix D.

Simulation Results for Message Flag N

Unlike the other message flags, flag N refers to scenarios where utility data look erratic and incorrect when compared to the building activity and average monthly temperature. This could be caused by a building owner accidentally switching the input of an electric or gas bill for a winter month with a summer month. The Tool identifies this inconsistency by raising flag N and recommends that the users correct their data. To create perturbed scenarios to trigger flag N, simulation results were created by switching two monthly electricity and gas consumption values randomly for 16 simulation results per building type. The switch was conducted between monthly utility data that were at least three months apart, so that the random switch was not conducted with months that are in the same season.

5 Evaluation of Estimated Energy End Use Disaggregation

The Tool uses monthly utility data to create a simplified building model and estimates the amount of energy allotted to the five different energy end uses (electric heating, cooling, and baseload energy consumption; and gas heating and baseload energy consumption). Because the estimated energy end uses are reported to building owners, it is necessary to check their accuracy.

The Tool uses regression algorithms to calibrate the simplified building model to provided monthly gas and electric utility data. As with any regression process, the sum of the calibrated monthly energy uses will not be identical to the annual energy consumption that is provided in the utility data, unless they are adjusted slightly to match the true values. For the analysis presented in this report, the Tool developers chose to use the calibrated, unadjusted values derived by the regression algorithms alone. Although the deviation between the actual and derived annual energy uses can propagate to the estimated energy end uses, the differences between the totals are generally between 1-2%. Appendix F.1 shows the comparison of annual electricity and gas uses determined by the tool to that of the simulation. Appendix F.2 shows a similar comparison, but with annual energy costs. However, to account for this, the Tool was evaluated in two ways: by evaluating the accuracy of the estimated annual energy uses in a standard fashion, and by evaluating the accuracy of the energy end use ratios, which are calculated by normalizing each end use to the annual energy use by fuel type. The second method enables an evaluation of the energy end uses independently of the deviation between the estimated and actual annual energy end uses.

5.1 Evaluation of Estimated Annual Energy Use

To evaluate the accuracy of the Tool's estimate of annual energy use, the percentage bias and uncertainty of the estimated annual energy uses were calculated for each simulation result listed in Table 10 and according to the method presented in Appendix E. The bias indicates how accurate the estimate is (is it over- or under-predicting?) and the uncertainty indicates the extent of repeatability of the bias. A small bias and a small uncertainty show that the estimation has a good fit, because the estimation can repeatedly give accurate estimations. If an estimate has a small bias and large uncertainty, the estimation is only occasionally accurate. Likewise, if an estimate has a large bias and a small uncertainty, the estimation is always inaccurate. The results for the different building types are plotted in Figure 3.

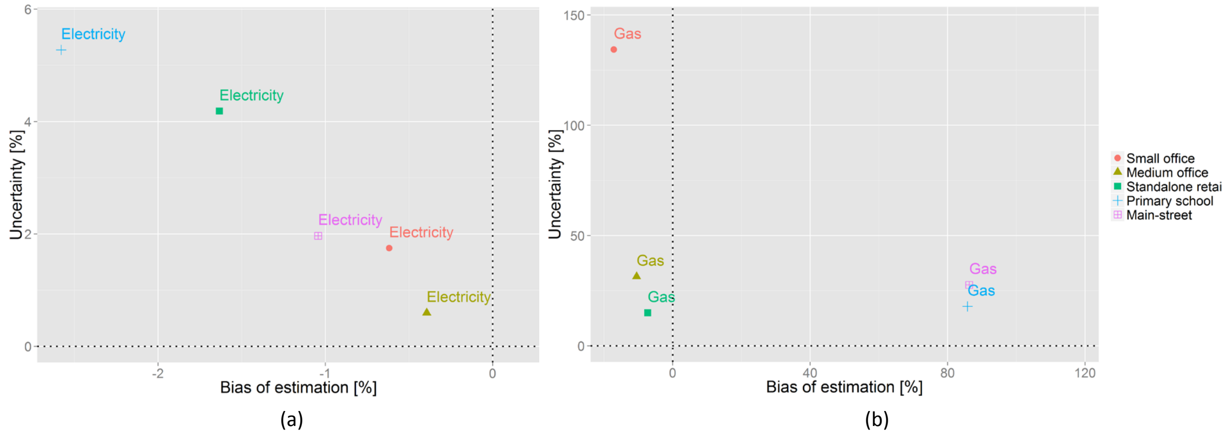


Figure 3. Percentage bias and uncertainty of annual energy uses of (a) electricity and (b) gas for different building types

The results in Figure 3 show that the largest deviation for annual electricity estimation occurs with primary schools with a bias of -2.6% and an uncertainty of 5.3%. This means that the Tool underestimates the annual electricity consumption by 2.6%, with an uncertainty of 5.3%. The gas estimates are different. The biases for primary schools and main-street buildings are higher than 75%. The complexity of the gas heating equipment can make it difficult for the tool to estimate annual gas use correctly. In small offices, medium offices, and standalone retail buildings, heat pumps or gas furnaces provide the primary heating. However, both heat pumps and gas furnaces are used as primary heating equipment in retrofitted main-street buildings, and water boilers and gas furnaces are used in primary schools. This causes large biases for primary schools and main-street buildings in Figure 3(b). The annual gas use estimation for small offices also has a bias of -17.2% and an uncertainty of 132.5%. Some small offices have negligible gas consumption because their interior is mainly heated by electricity. In these scenarios, even if the Tool estimates the gas consumption slightly different from the simulation values, it will cause a large percentage difference. This results in large bias and uncertainty for small offices in Figure 3(b).

To examine the importance of the deviations to the overall building performance, the electricity and gas consumption were evaluated by annual energy cost, rather than energy use. Combining electricity and gas use by simple summation of electricity and gas use is inappropriate because the maximum available energy for one unit of electricity is much higher than that of gas—meaning, people can do much more work with one unit of electricity than with one unit of gas. Although the appropriate thermodynamic property to consider this effect is maximum available energy (exergy) (Bejan 2006), it is not a common technical term. Source energy is used in zero energy building definitions (DOE 2015c), but this system has a different purpose of supporting the designation of zero energy. For this effort, a process is needed that is simple, directly related, and immediately connected to business financials. Hence energy cost was chosen to combine electricity and gas in this project. The annual energy cost was calculated by using the average 2013 gas and electricity cost for different U.S. states (EIA 2015a, b). The average percentage bias and uncertainty of the estimated annual energy cost for each building type is plotted in Figure 4.

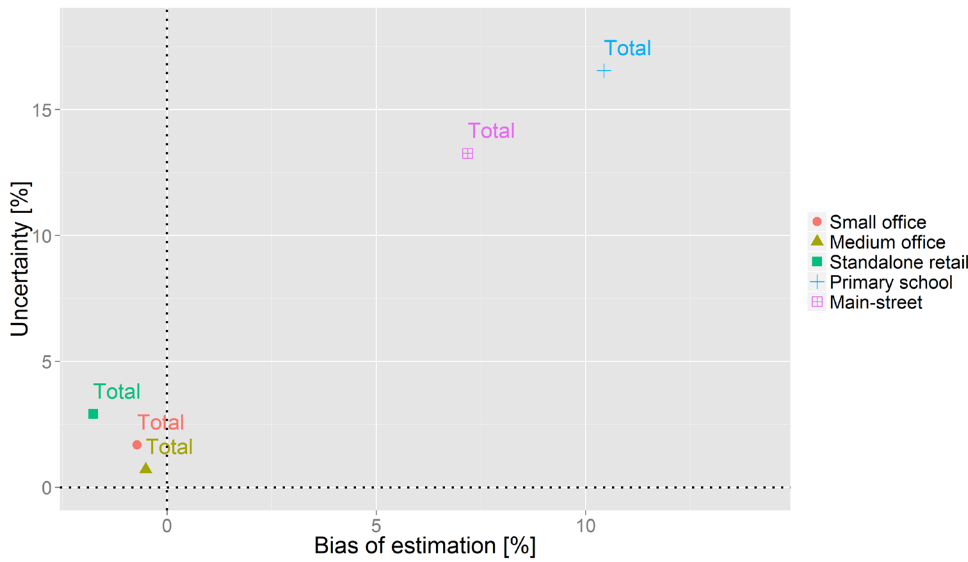


Figure 4. Percentage bias and uncertainty of annual energy cost for each building type

The results in Figure 4 show that the large gas consumption deviation for small offices (as seen in Figure 3) would not affect users significantly because the bias is only -0.7% and the uncertainty is 1.5%. However, the bias and uncertainty for primary schools and main-street buildings, as shown in Figure 4, are higher than 7% and 10% respectively, and the annual energy costs are overestimated. The cause may lie in the type of heating equipment found within the buildings.

Additional plots of bias and uncertainties for the estimates of annual energy use and annual energy costs are provided in Appendix F for reference.

5.2 Evaluation of Estimated Energy End Use Ratios

To evaluate the accuracy of how the tool estimates energy end uses independently of the deviation between the estimated and correct (from the building simulation) annual energy use, the assessment considers energy end use ratios that are calculated by normalizing the energy end uses with the annual energy use per fuel type. For the simulated energy end uses, the energy end use ratios were calculated by normalizing them with the annual energy use per fuel type. The end use ratios were calculated using equations (1) and (2).

$$\text{Correct electricity end use ratio} = \frac{\text{Electricity end use}}{\text{Total annual electricity use from building simulation}} \quad (1)$$

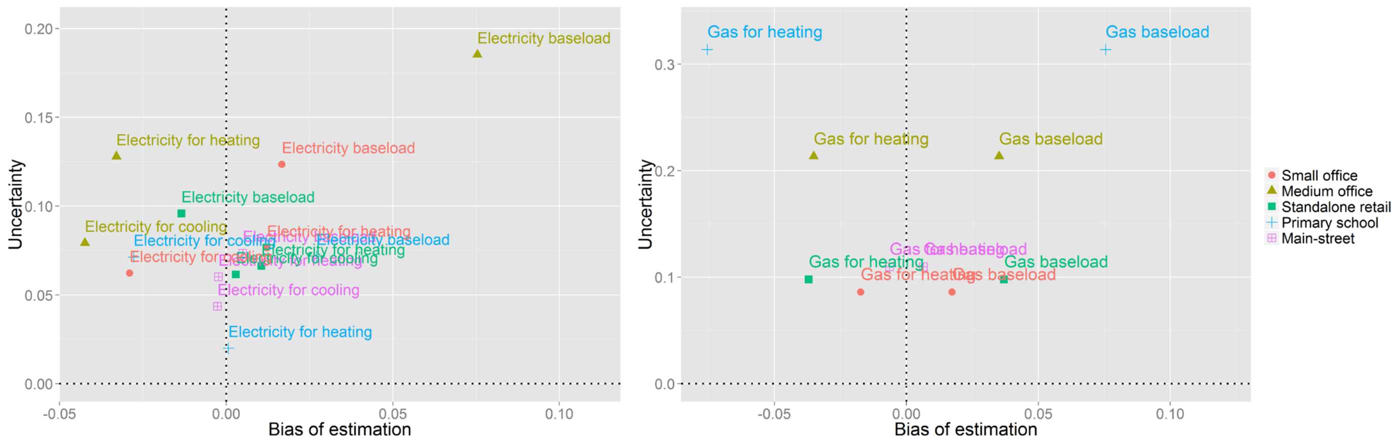
$$\text{Correct gas end use ratio} = \frac{\text{Gas end use}}{\text{Total annual gas use from building simulation}} \quad (2)$$

The energy end use ratios estimated by the Tool were calculated by equations (3) and (4).

$$\text{Estimated electricity end use ratio} = \frac{\text{Estimated electricity end use}}{\text{Total annual electricity use estimated by the Tool}} \tag{3}$$

$$\text{Estimated gas end use ratio} = \frac{\text{Estimated gas end use}}{\text{Total annual gas use estimated by the Tool}} \tag{4}$$

Because the energy end use ratios are normalized by the annual energy use, their magnitudes are independent of the difference between the estimated and correct annual energy use. The bias and uncertainty of the estimated energy end use ratios are calculated based on the differences between the correct and estimated energy end use ratios. A detailed explanation of this process is noted in Appendix E and results are plotted in Figure 5.



(a) (b)

Figure 5. Bias and uncertainty of (a) electricity and (b) gas energy end use ratios for different building types

Figure 5 shows that most electricity end uses are estimated within a bias of ±0.05, except for the electricity baseload found in medium offices. Most gas end uses are estimated within a bias of ±0.05 as well, except for the gas end uses found in primary schools. However, the repeatability of the estimation is not very high because most uncertainty values in Figure 5 are higher than 0.05.

To understand the deviation relative to the overall building performance, the energy end use cost ratios were calculated. Similar to the calculation of the energy end use ratios, the cost ratios were calculated by normalizing the energy end uses by the annual energy costs that are estimated by the average 2013 energy costs (EIA 2015a, b). Equations (5) and (6) were used to calculate the cost ratios.

$$\text{Correct cost ratios of energy end uses} = \frac{\text{Cost of energy end use}}{\text{Annual energy cost from building simulation}} \tag{5}$$

$$\text{Estimated cost ratios of energy end uses} = \frac{\text{Estimated cost of energy end use}}{\text{Annual energy cost estimated by the Tool}} \qquad (6)$$

The bias and uncertainty of the estimated cost ratios are plotted in Figure 6.

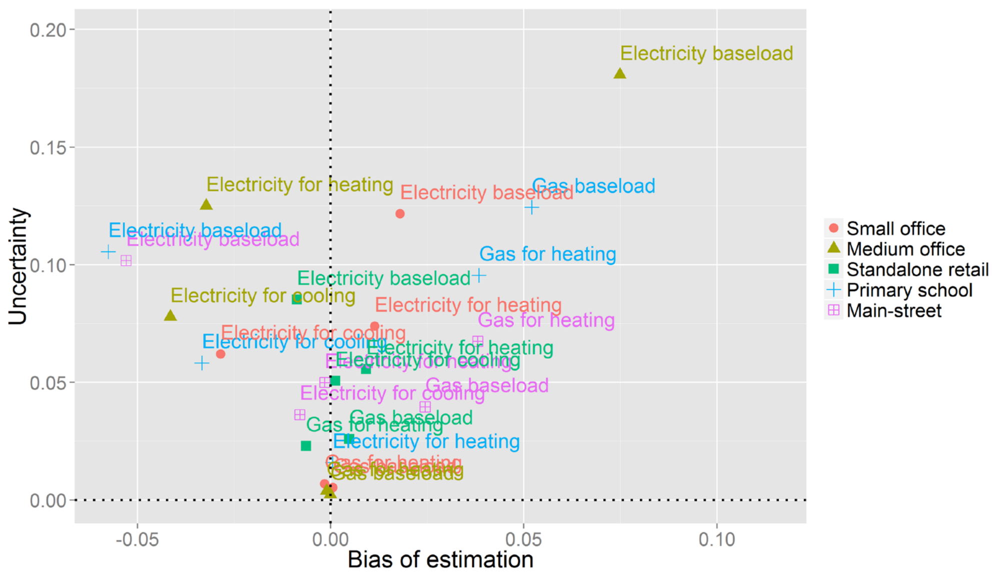


Figure 6. Bias and uncertainty of estimated cost ratios for energy end uses in different building types

The electricity costs shown in Figure 6 scatter similarly to their counterparts in Figure 5, but the gas cost ratios shown in Figure 6 cluster around the origin much more closely than that of Figure 5. This is a result of the smaller significance of cost for gas than that of electricity in the total energy cost of a building. Figure 6 also shows that the Tool is more accurate when the buildings are smaller and simpler. This is illustrated by the smaller bias and uncertainty for small offices, standalone retail buildings, and main-street buildings. The medium offices and primary schools behave differently. Further investigation was conducted to check the results of primary schools and medium offices, and the bias and uncertainty of the energy end use ratios for primary schools of different vintages are plotted in Figure 7.

(a) (b)

Figure 7. Bias and uncertainty of energy end uses of primary schools modeled in different vintages for (a) electricity and (b) gas

Figure 7 shows that the major reason for the Tool's overestimation and high uncertainty of the gas baseload cost ratio (shown in Figure 6) is associated with the 2004 and 2013 primary schools. In the pre-1980 and post-1980 primary school models, the heating equipment is modeled with water boilers and the estimates are more accurate. However, in the 2004 and 2013 models, the heating is modeled with both water boilers and gas furnaces. The complexity of heating equipment reduced the accuracy of the Tool's estimate of gas heating, thus affecting the gas baseload as well. This causes the larger bias and uncertainty of energy end uses for primary schools in Figure 6.

The larger bias and uncertainty associated with the medium office electricity baseload (shown in Figure 6) was further investigated by plotting the bias and uncertainty for electricity energy end use ratios for different vintages. The results are shown in Figure 8.

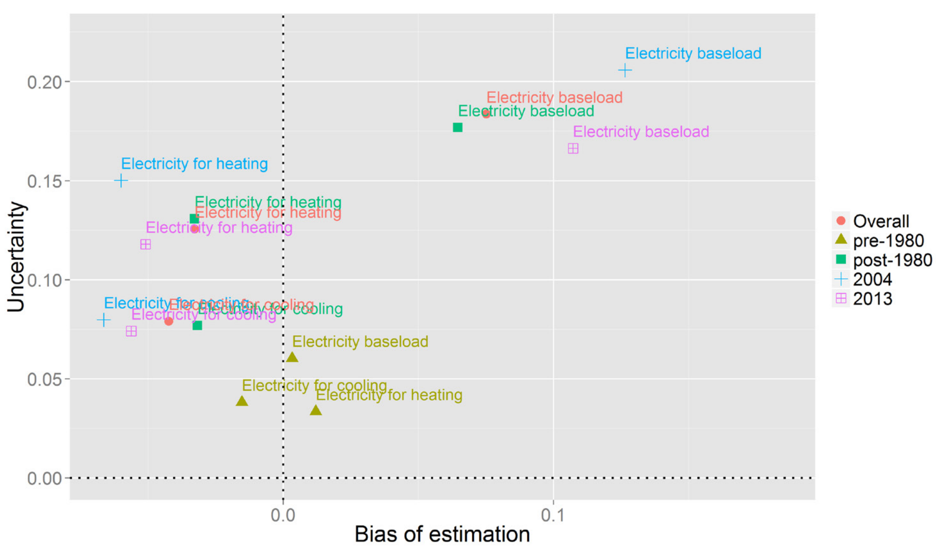


Figure 8. Bias and uncertainty of electricity energy end use ratios for medium offices of different vintages

Figure 8 shows that the electricity baseload and heating for medium offices is estimated with larger biases and uncertainties for post-1980 buildings than the pre-1980 buildings. A comparison of HVAC equipment between these vintages finds that the pre-1980 medium offices use constant speed fans with no reheat, while the other medium offices use variable-speed fans with electric reheat. The addition of the reheat and variable-speed fans to the newer buildings increases the difficulty for the Tool to estimate the electricity baseload and heating energy accurately. This results in a larger bias and uncertainty for medium office buildings as shown in Figure 6.

5.3 Summary

To summarize, the Tool shows a larger bias and uncertainty in annual energy use estimations for primary schools and main-street buildings. If the Tool developers eliminated the slight deviation between the calibrated, unadjusted energy values derived by their regression algorithms alone, and the actual annual energy calculated by the utility bill data, ,the energy end use estimation would most likely be unbiased for small office, standalone retail buildings, and main-street buildings with acceptable uncertainty. Regardless of the deviation, the tool was less accurate in estimating the gas or electricity baseload in the newer primary schools and medium offices due to the complexity of the HVAC equipment, or its ability to disaggregate larger internal loads.

Detailed plots of bias and uncertainty of energy end uses and cost ratios are listed in Appendix F for reference.

6 Evaluation Estimated Potential Energy Savings

The Tool can help encourage building owners to retrofit their buildings by identifying potential energy and cost savings. To examine the accuracy of the tool's energy savings estimates, electricity and gas savings were calculated from the building simulation results and compared to the energy savings estimated by the Tool.

6.1 Cost Savings Calculation

Although the Tool estimates electricity and gas savings separately, the evaluation considers the fuel types together because energy cost savings are dependent on the interaction of the different systems within a building. For example, if a building is diagnosed as consuming excessive electric use for plug loads, the tool should recommend that the owners reduce plug load energy and should estimate positive electricity savings. However, if the plug load electricity is reduced to the point where gas is increased to make up for a loss of heat from electric equipment, the tool should account for negative savings in gas (even if the electric savings are greater). For a proper evaluation, electric and gas savings are combined using average 2013 energy costs (EIA 2015a, b). The total cost savings are calculated using equation (7).

$$\begin{aligned}\text{Estimated potential cost savings} &= (\text{Estimated electricity savings}) \times (\text{Electricity cost}) \\ &\quad + (\text{Estimated gas savings}) \times (\text{Gas cost})\end{aligned} \tag{7}$$

Since estimated energy savings can be negative, it is possible for equation (7) to yield negative cost savings. If the value from equation (7) is negative, it is assumed that the owners will either do nothing or investigate a potential operations or maintenance fault, and that the estimated potential cost savings is zero.

6.2 Energy Cost Savings Calculation from Building Simulation Results

Because building energy savings is the difference between current and post-retrofit energy consumption, it is necessary to define post-retrofit energy consumption. For each building simulation result fromTable 10, energy cost savings was calculated for two cases: (1) energy cost savings calculated by comparing a perturbed model to the baseline building model, and (2) savings calculated by comparing a model to the standard efficient case, which is defined as the most current model (i.e., 2013 vintage) for each building type.

Baseline Case

The baseline building model, without any perturbations as described in Section 4.2, was used as the baseline case for calculating potential energy cost savings. Under this scenario, it is assumed that the building owners fix the operating or structural issues that caused the simulation results to deviate from the baseline model.

To calculate the potential cost savings, the electric and gas savings were calculated separately using equations (8) and (9).

$$\text{Electricity savings from baseline case} = \text{Annual electricity use of the current case} - \text{Annual electricity use in the baseline case} \quad (8)$$

$$\text{Gas savings from baseline case} = \text{Annual gas use of the current case} - \text{Annual gas use in the baseline case} \quad (9)$$

The electric and gas savings from equations (8) and (9) can be used to calculate the correct cost savings from the baseline case by equation (10).

$$\text{Cost savings from baseline case} = (\text{Electricity savings from baseline case})(\text{Electricity cost}) + (\text{Gas savings from baseline case})(\text{Gas cost}) \quad (10)$$

Standard Efficient Case

The newest vintage baseline model for each building type was used as the standard efficient case for calculating energy cost savings compared to an older building or perturbed building models. An example is given in Table 11.

Table 11. Example of the Baseline and Standard Efficient Case Used To Calculate Potential Energy Savings

	Building Simulation Result	Corresponding Baseline Case	Corresponding Standard Efficient Case
Climate zone	5A	5A	5A
Building type	Primary school	Primary school	Primary school
Vintage	pre-1980	pre-1980	2013
Perturbation	Increasing internal lighting intensity by 100%	None	None

As represented in Table 11, the example shows a perturbed building model that was created by increasing the internal lighting intensity by 100% from a pre-1980 primary school model in climate zone 5A. Because the newest vintage considered in the project is 2013, the corresponding standard efficient case comes from the 2013 primary school model in climate zone 5A without an applied perturbation.

Under this scenario, it is assumed that the building owners could achieve energy savings by retrofitting their building with standard 2013 building practices applied to the building envelope, HVAC equipment, and other building systems. This scenario usually results in more energy savings than retrofitting to the baseline case, but would require much more capital to implement the retrofit. The electricity and gas savings were calculated using equations (11) and (12).

$$\text{Electricity savings from standard efficient case} = \text{Annual electricity use of the current case} - \text{Annual electricity use in the standard efficient case} \tag{11}$$

$$\text{Gas savings from standard efficient case} = \text{Annual gas use of the current case} - \text{Annual gas use in the standard efficient case} \tag{12}$$

The energy cost savings from the standard efficient case were calculated using equation (13).

$$\begin{aligned} &\text{Cost savings for the standard efficient case} \\ &= (\text{Electricity savings from standard efficient case}) \times (\text{Electricity cost}) \\ &\quad + (\text{Gas savings from the standard efficient case}) \times (\text{Gas cost}) \end{aligned} \tag{13}$$

6.3 Evaluation Results and Discussion

To examine the accuracy of the Tool's savings estimates, the normalized deviation, calculated by equation (14), was used to quantify the bias and uncertainties according to Appendix E.

$$\frac{\text{Estimated potential cost savings - Correct cost savings}}{\text{Annual energy costs from the building simulation result}} \tag{14}$$

Because building owners might check the significance of cost savings relative to their annual energy costs, the deviation of the estimated cost savings from the correct cost savings was normalized by the annual energy costs of the building before a retrofit. The deviations were used in calculating bias and uncertainties that are plotted in Figure 9.

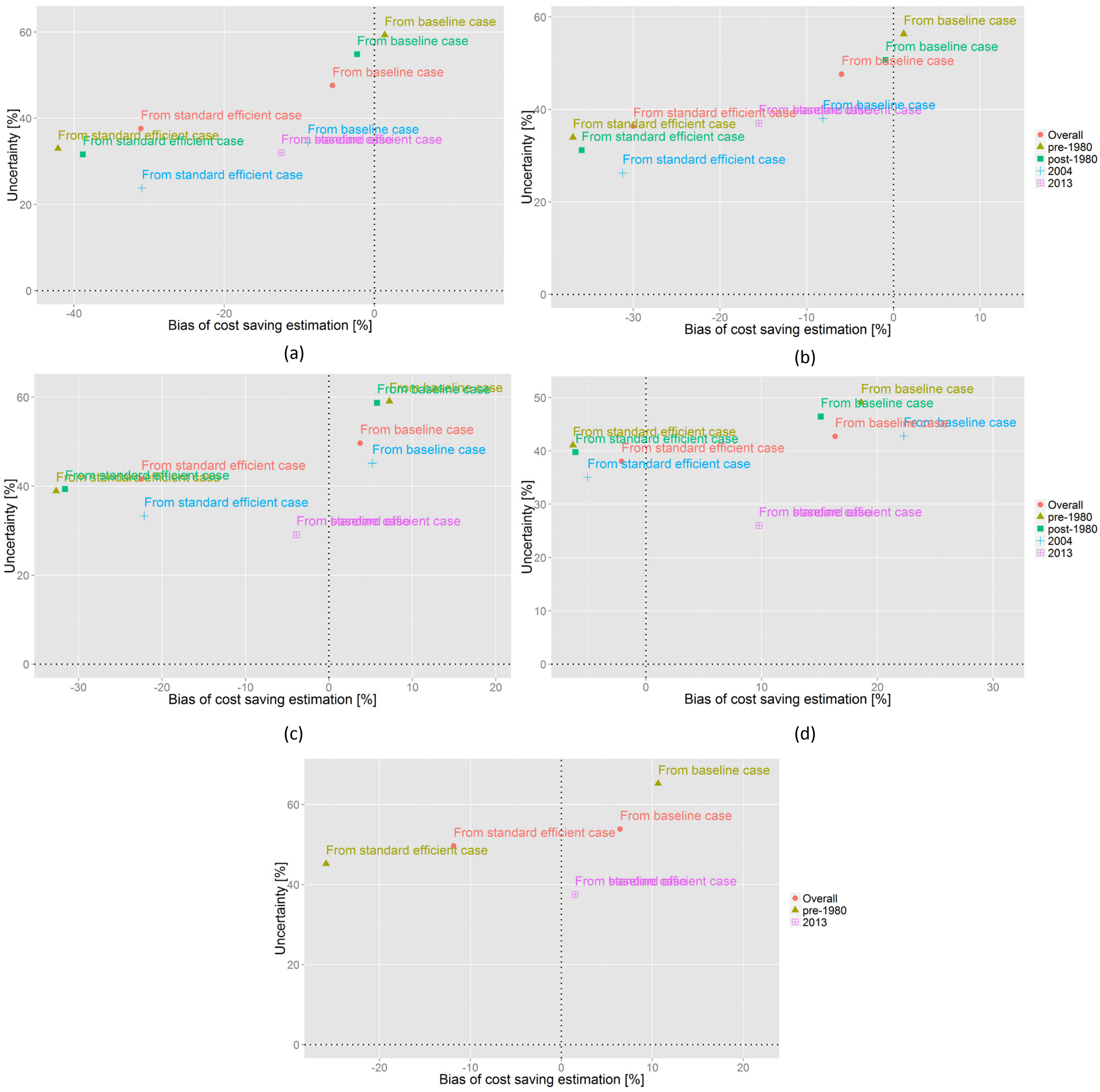


Figure 9. Bias and uncertainty of the Tool's energy cost savings estimates for (a) small offices, (b) medium offices, (c) standalone retail buildings, (d) primary schools, and (e) main-street buildings

Figure 9 shows that the cost savings from the baseline case have smaller biases than the cost savings from the standard efficient case for small and medium offices. This means that the Tool is effectively using the baseline case for calculating cost savings for the small and medium offices. For standalone retail and main-street buildings, the biases for the cost savings from the baseline case are positive and the biases for the cost savings from the standard efficiency case are negative. This shows that the Tool's baseline for calculating energy savings lies somewhere between the baseline model and the standard efficient case. For primary schools, Figure 9(d) shows that the Tool's cost savings estimates for the standard efficient cases have a small bias and the Tool cost savings estimates for pre-1980 to 2004 vintages are relative to the standard efficient case.

It is clear from Figure 9 that the Tool significantly overestimates the cost savings potential for 2013 primary schools. Although there are significant biases for other types of 2013 buildings, the biases are negative. This means that the Tool will not give false hopes of cost savings for building owners interested in retrofitting their buildings. However, the overestimation of cost savings for the 2013 primary school could encourage building owners to investigate a retrofit even if it is not warranted. Hence the intent of this tool in providing a high-level first look at potential savings must be emphasized to the users, and owners should understand that a more in-depth assessment will be needed.

Figure 9 also shows that the tool's cost savings estimates are unreliable with uncertainties greater than 20%. Because the older buildings usually show larger uncertainty than newer buildings, this issue is explored further in Figure 10 using histogram plots of normalized cost savings deviations from the baseline case for small office buildings.

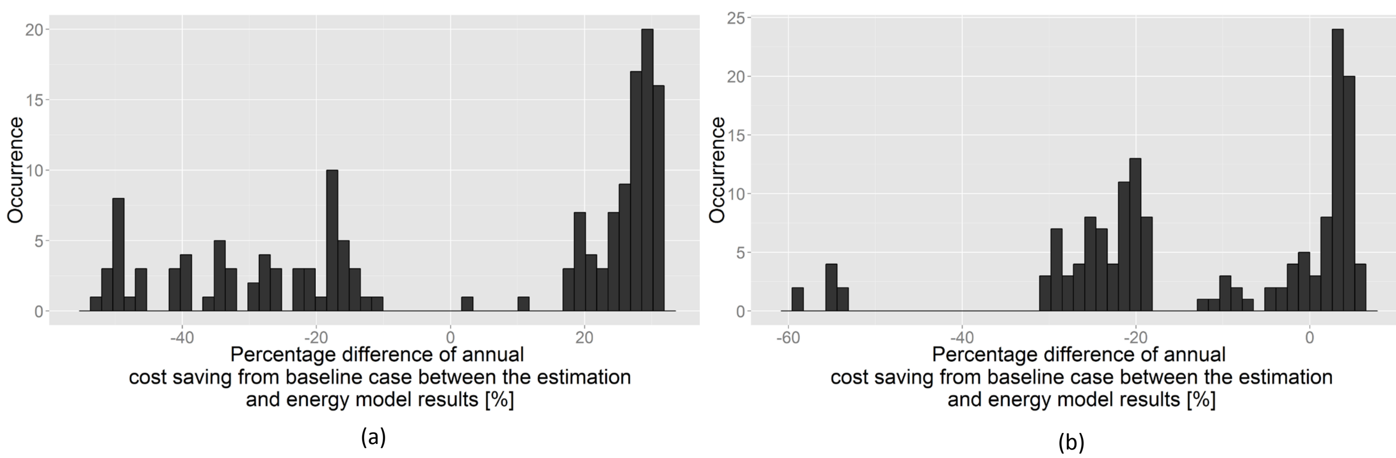


Figure 10. Histograms of normalized cost savings deviations for (a) pre-1980 main-street buildings, and (b) 2013 small office buildings

The histogram in Figure 10(a) has a wider spread than the histogram in Figure 10(b), which shows that the deviation of the estimated cost savings compared to the true cost savings is likely to be larger for pre-1980 small office buildings than for 2013 small office buildings. This explains the larger uncertainty in cost savings estimates for older vintages (Figure 9) than for newer buildings. In summary, the Tool cost savings estimates are less reliable for older buildings than for newer ones. It should also be noted that the spread in cost savings, shown in both histograms in Figure 10, are large with multiple peaks. This suggests that the tool does not have a consistent retrofit target for evaluating potential energy savings.

6.4 Summary

To summarize, the accuracy of the cost savings estimates predicted by the Tool was evaluated. Because the post-retrofit performance target that the Tool uses to estimate savings is proprietary, two reference cases were considered—the baseline case and the standard efficient case. By comparing the biases and uncertainties of the cost savings estimates calculated in reference to these two cases, it was found that the cost savings estimates of small and medium offices, on average, reference the baseline case while the cost savings estimates for primary schools are more in line with the standard efficient case. It can also be concluded that the standalone retail and main-street buildings lie between the baseline and standard efficient case, and the reference cases change pending building type and vintage. It is important to note that the Tool overestimates potential cost savings for the 2013 primary school buildings. It is also important to note that the the Tool cost savings estimates have large uncertainty, which implies that the Tool's cost savings estimates could be unreliable and the tool is inconsistent in identifying retrofit performance targets for different building types.

7 Evaluation of Message Flag Accuracy

In addition to estimating energy end uses and potential energy cost savings for retrofits, the Tool also alerts building owners to operational and construction-related issues in their buildings using a series of message flags. This section first discusses the design of the message flags in terms of their effectiveness and clarity to communicate building issues to the owners. Then an assessment of message flag accuracy is presented and discussed.

7.1 Message Flag Definitions

Categorizing Electricity Baseload Energy Use (Flags A, B, C, O, and P)

The Tool provides four main categories of message flags for electricity baseload use: (A) low, (B) slightly high, (C) very high, and (O or P) ultra-high. The categories are designed to be mutually exclusive and only one of them would be raised per analysis. However, because the categorization does not include a class for "normal" electric baseload energy use, building owners will always be alerted with at least one electric baseload issue.

The definitions for these flags are fixed quantitatively regardless of the building type. However, different types of buildings may have different standards for high and low electricity baseload based on industry codes and standards. For example, according to ASHRAE Standard 90.1-2013 (ASHRAE 2013b), the standard lighting power density of a retail space is 13.6 W/m^2 (1.26 W/ft^2) and is 54.5% higher than that of an office space, which is 8.8 W/m^2 (0.82 W/ft^2). When the Tool uses the same classification scheme for all building types, it may mistakenly classify retail stores as "slightly high electricity baseload" when the baseload is normal. Hence the tool developers should consider using different electricity baseload classification schemes for different types of buildings.

Message Flags Unrelated to Energy Savings: Flag Q

While the Tool aims to alert building owners of operation issues that affect energy use, there is a flag, "low relative heat use," that is meant to indicate that a building has poor thermal comfort or inadequate occupancy scheduling. However, the meaning of this flag is not well communicated in the documentation of the Tool, and the developers should consider adding a clarified explanation or removing the flag.

7.2 Evaluation Methodology for Message Flags Independent of One Another (Flags N, E, I, K, and L)

The N, E, I, K, and L message flags are not related to other flags, and their accuracy can be assessed by using 2x2 confusion matrices, which are frequently used to assess the accuracy of classification algorithms (Fawcett 2006). The matrices for these flags are developed by considering the building simulation results as the correct results and the Tool analysis results as the estimated results. With these results, the data points are allocated to four classes as shown in Table 12.

Table 12. Example of a 2x2 Confusion Matrix Used To Assess the Accuracy of the Tool in Triggering a Flag Related to a Particular Building Operating Issue

	Simulation Results with the Issue	**Simulation Results without the Issue**
The Tool raises the flag related to the issue	True diagnosis	False alarm
The Tool does not raise the flag	False negative	True negative

Table 12 categorizes the simulation results in a 2x2 confusion matrix based on the Tool analysis results. In a case where an issue exists and the Tool correctly flags the issue, the result is categorized under "True diagnosis." If an issue exists but the Tool does not flag it, then it is categorized under "False negative." A "False alarm" is associated with a case where the Tool flags an issue that does not exist. Finally, when a fault does not exist and the Tool does not issue a flag, then this situation is categorized as "True negative." After categorizing all cases and counting the number of cases in each class, the true diagnosis and false alarm rates are calculated by equations (15) and (16).

$$\text{True diagnosis rate} = \frac{\text{Number of true diagnoses}}{\text{Number of simulation cases with the flag}} \tag{15}$$

$$\text{False alarm rate} = \frac{\text{Number of false alarms}}{\text{Number of simulation cases without the flag}} \tag{16}$$

A high true diagnosis rate would demonstrate that the tool can correctly identify issues that exist. Even more important than a high true diagnosis rate, the false alarm rate should be low so as to not alert users to issues that don't exist.

Flag N

Flag N refers to possible erratic operation or occupancy. For example, 80 cases were generated with the issue in addition to the cases listed in Table 10. To assess the accuracy of the Tool in triggering flag N, confusion matrices were drawn considering the 80 results with the issue and the additional cases in Table 10 without the issue. The true diagnosis and false alarm rates are tabulated in Table 13.

Table 13. True Diagnosis and False Alarm Rates for Flag N

Building Type	True Diagnosis Rate	False Alarm Rate
Small office	12.5%	0%
Medium office	6.3%	0%
Standalone retail	18.8%	0%
Primary school	37.5%	0%
Main-street building	0.0%	0%

Table 13 shows that the false alarm rates were zero for all types of buildings. The true diagnosis rates are relatively low but greater than zero except for the main-street buildings. For the main-street buildings, the tool is not useful in diagnosing possible erratic operation and occupancy issues.

One possible reason for the low true diagnosis rates is that switching monthly data may not have caused a change significant enough to affect the energy usage pattern in ways that would trigger Flag N.

Flag E

Flag E is related to inefficient cooling COP of building cooling equipment. This was simulated by either increasing the size of the HVAC equipment to increase the on-off cycling of equipment (which is less efficient), or by reducing the COP of the equipment from the rated condition. The true diagnosis and false alarm rates for this flag are tabulated in Table 14.

Table 14. True Diagnosis and False Alarm Rates for Flag E

Building Type	True Diagnosis Rate	False Alarm Rate	Cases with Flag E from the Simulation Results
Small office	0.0%	0.9%	82
Medium office	0.0%	3.4%	61
Standalone retail	13.3%	2.8%	45
Primary school	1.5%	0.8%	195
Main-street building	0.0%	0.0%	24

These results show that the tool cannot identify any cases with low cooling COP in small offices, medium offices, and main-street buildings. For standalone retail buildings and primary schools, the results are somewhat better with true diagnosis rates that are higher than the false alarm rates.

Flag I

Flag I is related to the overuse of cooling in a building. The true diagnosis and false alarm rates are tabulated in Table 15.

Table 15. True Diagnosis and False Alarm Rates for Flag I

Building Type	True Diagnosis Rate	False Alarm Rate	Cases with Flag I from the Simulation Results
Small office	16.3%	8.0%	86
Medium office	20.8%	14.0%	87
Standalone retail	31.2%	18.6%	93
Primary school	54.2%	17.5%	83
Main-street building	8.6%	3.6%	35

The results show that the true diagnosis rates are always higher than the false alarm rates indicating that the tool is more likely to get the right answers than wrong ones. However, only the true diagnosis rate for primary schools is greater than 50%.

Flag K

Flag K relates to unusually high summer gas use. Its true diagnosis and false alarm are tabulated in Table 16.

Table 16. True Diagnosis and False Alarm Rates for Flag K

Building Type	True Diagnosis Rate	False Alarm Rate	Cases with Flag K from the Simulation Results
Small office	25.8%	0.2%	31
Medium office	0.0%	0.0%	93
Standalone retail	73.6%	12.6%	110
Primary school	100.0%	100.0%	114
Main-street building	100.0%	72.0%	44

The results show that the tool performs best in standalone retail buildings, with a true diagnosis rate of 73.6% and low false alarm rate of 12.6%. The performance for the small office is also reasonable with a true diagnosis rate of 25.8% and false alarm rate of 0.2%. However, the tool does not perform well for main-street buildings, with a false alarm rate of 72.0%. Furthermore, the tool performance is low for the medium office (no correct diagnoses), primary schools (always diagnoses the flag regardless of the case), and main-street buildings (a high false alarm rate that is not much lower than the true diagnosis rate).

Flag L

Flag L relates to the use of reheat during warm and summer months. The Tool performance results for this flag are tabulated in Table 17.

Table 17. True Diagnosis and False Alarm Rates for Flag L

Building Type	True Diagnosis Rate	False Alarm Rate	Cases with Flag L from the Simulation Results
Small office	Not available	5.9%	0
Medium office	50.0%	7.1%	20
Standalone retail	Not available	2.7%	0
Primary school	9.1%	7.4%	22
Main-street building	Not available	0.6%	0

Table 17 shows that the flag is only applicable to two building types because not all of the building types use reheat devices. Thus, the true diagnosis rate of flag L is undefined for those buildings without reheat. However, the tool flags this issue despite the absence of reheat devices in the buildings and hence, the results show false alarm rates for small office, standalone retail, and main-street buildings. Among the buildings with reheat devices, the tool performs much better for the medium office with a 50% true diagnosis rate and a low false alarm rate.

7.3 Evaluation Methodology and Results for Pairs of Mutually Exclusive Flags (Flags H and Q)

When operating issues associated with two flags that are mutually exclusive exist, the accuracy of the tool in triggering those flags can be simultaneously assessed by using a 3x3 confusion matrix. The matrix creates nine different classes according to the diagnosis results and the existence of the operating issues. An example is illustrated in Table 18 where flag 1 and flag 2 are mutually exclusive.

Table 18. Example of a 3x3 Confusion Matrix Used To Assess the Accuracy of the Tool in Triggering a Pair of Flags

	Simulation Results with Issues Related to Flag 1	Simulation Results without Any Issues Related to Flag 1 or 2	Simulation Results with Issues Related to Flag 2
Diagnosed with flag 1	True diagnosis for flag 1	False alarm for flag 1	False diagnosis for flag 1
No diagnosis for flag 1 or 2	False negative	True negative	False negative
Diagnosed with flag 2	False diagnosis for flag 2	False alarm for flag 2	True diagnosis for flag 2

After classifying each simulation result into the nine classes shown in Table 18, the number of cases under each class is counted and used to calculate the performance indicators for flags 1 and 2: true diagnosis, false diagnosis, and false alarm rates. For flag 1, these indicators are calculated using equations (17), (18), and (19).

$$\text{True diagnosis rate for flag 1} = \frac{\text{Number of true diagnoses for flag 1}}{\text{Number of simulation cases with flag 1 issue}} \quad (17)$$

$$\text{False alarm rate for flag 1} = \frac{\text{Number of false alarms for flag 1}}{\text{Number of simulation cases without flag 1 and 2 issues}} \quad (18)$$

$$\text{False diagnosis rate for flag 1} = \frac{\text{Number of false diagnoses for flag 1}}{\text{Number of simulation cases with flag 2 issues}} \quad (19)$$

True diagnosis rate for flag 1 in equation (17) indicates how often the Tool can correctly diagnose the cases with the flag 1 issue. False alarm rates for flag 1 in equation (18) indicates how often the Tool diagnoses flag 1 without the presence of flag 1 and 2 issues. False diagnosis rate for flag 1 in equation (19) indicates how often the cases with a flag 2 issue are diagnosed with flag 1. A high false diagnosis rate implies that the tool is likely to incorrectly raise flag 1 when flag 2 should be raised. The same rates are calculated for flag 2 to evaluate the performance of the tool in diagnosing flag 2 issues.

There is only one pair of mutually exclusive flags in the Tool: flag H (high relative heating use) and flag Q (low relative heating use). The performance of the tool for these message flags was evaluated using a 3x3 confusion matrix. The true diagnosis and false alarm rates for flags H and Q across the different building types were calculated according to equations (17), (18), and (19) and tabulated in Table 19.

Table 19. True Diagnosis, False Diagnosis Rates, and False Alarm Rates for Flags H and Q for Different Building Types

Building Type	Flag H			Flag Q		
	True Diagnosis Rate	False Diagnosis Rate for Flag H	False Alarm Rate for Flag H	True Diagnosis Rate	False Diagnosis Rate for Flag Q	False Alarm Rate for Flag Q
Small office	28.5%	0%	10.0%	14.3%	10.0%	7.8%
Medium office	34.9%	0%	14.7%	8.7%	0.0%	6.0%
Standalone retail	13.7%	0%	3.6%	33.3%	27.4%	22.6%
Primary school	33.0%	0%	17.7%	23.1%	33.0%	18.9%
Main-street building	4.5%	0%	0.7%	50.0%	22.7%	24.6%

The results show that the true diagnosis rate of flag H is higher than the false alarm rates for all building types. However, none of the true diagnosis rates for flag H are higher than 50%. Similar performance can be seen for the diagnosis of flag Q, except in the cases for primary schools. In these cases, the true diagnosis rate is lower than the false alarm rate and the tool is more likely to claim a case with high usage of heat when it actually has a low heat use.

7.4 Evaluation Methodology and Results for Multiple Flags with Similar Operating Issues (Flags D and M)

The Tool documentation indicates that flag D (inefficient shell and ventilation) and flag M (possible low heating efficiency) may both be raised when there is an excessive infiltration problem. Although their main functions would alert building owners to operating issues that are physically different, the documentation indicates that some inefficient shell and ventilation issues may be alerted by flag M instead of flag D because of the difficulty in separating the issues completely in its simplified mathematical model. To accommodate this supplementary function, the two flags were evaluated together. To determine appropriate performance indicators for the two flags, confusion matrices were constructed to classify the cases, first without considering the supplementary function, and secondly to classify them according to the rules in Section 7.2. The resultant confusion matrices are shown in Table 20 and Table 21.

Table 20. Confusion Matrix for Flag M with and without Considering the Supplementary Function of Flag M To Diagnose Cases with an Inefficient Shell and Ventilation

	With Inefficient Shell and Ventilation (D)		**Without Inefficient Shell or Ventilation (D)**	
	With Low Heating Efficiency (M)	**Without Low Heating Efficiency (M)**	**With Low Heating Efficiency (M)**	**Without Low Heating Efficiency (M)**
Diagnosed with flag D and M	True diagnosis	False alarm	True diagnosis	False alarm
Diagnosed with flag D but not M	False negative	True negative	False negative	True negative
Diagnosed with flag M but not D	True diagnosis	False alarm	True diagnosis	False alarm
No diagnosis of D nor M	False negative	True negative	False negative	True negative

Table 21. Confusion Matrix for Flag D without Considering the Supplementary Function of Flag M To Diagnose Cases with Inefficient Shell and Ventilation

<table>
<tr><th rowspan="2"></th><th colspan="2">With Inefficient Shell and Ventilation (D)</th><th colspan="2">Without Inefficient Shell or Ventilation (D)</th></tr>
<tr><th>With Low Heating Efficiency (M)</th><th>Without Low Heating Efficiency (M)</th><th>With Low Heating Efficiency (M)</th><th>Without Low Heating Efficiency (M)</th></tr>
<tr><td>Diagnosed with flag D and M</td><td colspan="2" rowspan="2">True diagnosis</td><td colspan="2" rowspan="2">False alarm</td></tr>
<tr><td>Diagnosed with flag D but not M</td></tr>
<tr><td>Diagnosed with flag M but not D</td><td colspan="2" rowspan="2">False negative</td><td colspan="2" rowspan="2">True negative</td></tr>
<tr><td>No diagnosis of D nor M</td></tr>
</table>

When the supplementary function of flag M is not considered, the classes for flag D and M (shown in Table 20 and Table 21) are not affected by the conditions or diagnoses related to the other flag. The classes of true diagnosis, false alarm, false negative, and true negative in Table 20 only depend on whether flag M is triggered by the Tool and if the building equipment has low heating efficiency. Similarly, the classes in Table 21 only depend on if flag D is triggered by the Tool and if the building has an inefficient shell and ventilation issue. The true and false alarm rates calculated from Table 20 and Table 21 are tabulated in Table 22.

Table 22. True and False Alarm Rates for Flags D and M without Considering the Supplementary Function of Flag M

Building Type	Flag D		Flag M	
	True Diagnosis Rate	False Alarm Rate	True Diagnosis Rate	False Alarm Rate
Small office	62.9%	45.1%	15.5%	9.4%
Medium office	60.7%	39.1%	21.3%	14.9%
Standalone retail	92.7%	76.3%	0.7%	3.9%
Primary school	93.9%	84.3%	5.9%	6.2%
Main-street building	100.0%	82.8%	7.7%	13.3%

Table 22 shows that all true diagnosis rates are higher than the false alarm rates for flag D, and the tool is often able to distinguish the inefficient shell and ventilation issues from cases without those issues. However, the results also show that the false alarm rate for flag D ranges from 39.1% to 84.3%. For flag M, the false alarm rates are higher than the true diagnosis rates for standalone retail buildings, primary schools, and main-street buildings. This shows that the tool does a poor job of distinguishing low heating efficiency from cases with normal heating efficiency. The true diagnosis rates for flag M for small and medium size offices are also low with a maximum rate of 21.3%.

To examine whether the low true diagnosis rates for flag M are related to the supplementary function, the confusion matrix in Table 20 was modified to accommodate the supplementary function in Table 23.

Table 23. Confusion Matrix for Flag M Considering the Supplementary Function of Flag M To Diagnose Cases with an Inefficient Shell and Ventilation

	With Inefficient Shell and Ventilation (D)		**Without Inefficient Shell or Ventilation (D)**	
	With Low Heating Efficiency (M)	**Without Low Heating Efficiency (M)**	**With Low Heating Efficiency (M)**	**Without Low Heating Efficiency (M)**
Diagnosed with flag D and M	True diagnosis	False alarm	True diagnosis	False alarm
Diagnosed with flag D but not M	False negative	True negative	False negative	True negative
Diagnosed with flag M but not D	<u>Weak diagnosis</u>	<u>Weak diagnosis</u>	True diagnosis	False alarm
No diagnosis of D nor M	False negative	<u>False negative</u>	False negative	True negative

The differences between Table 20 and Table 23 are underlined in Table 23. When the building has both issues but only diagnoses flag M, it is considered incomplete and termed a weak diagnosis. When the building has an inefficient shell and ventilation but normal (not low) heating efficiency, flag M only serves as the supplementary function and is considered a weak diagnosis. It is not considered a true diagnosis because it could be interpreted as low heating efficiency when in reality, the building does not have that issue. If the building has the inefficient shell and ventilation but no diagnosis of flags D or M, flag M fails to serve its supplementary function and is considered a false negative. In comparing Table 20 with Table 23, it's shown that the supplementary function reduces the number of false alarms for flag M, at the sacrifice of true diagnosis and true negative rates. Because the supplementary function introduced a weak diagnosis, a new performance indicator called "weak diagnosis rate" was introduced. This is calculated by equation (20).

$$\text{Weak diagnosis rate} = \frac{\text{Number of weak diagnoses}}{\text{Number of simulation cases with inefficient shell and ventilation}} \quad (20)$$

Because the number of true negatives for flag M was changed by the supplementary function, the true negative rate was calculated by equation (21).

$$\text{True negative rate} = \frac{\text{Number of true negatives}}{\text{Number of simulation cases without low heating efficiency issues}} \quad (21)$$

Without the supplementary function, the true negative rate can be calculated by subtracting 1 from the false alarm rate. Because the supplementary function only involves the function of flag M, it does not change the functions of flag D in the confusion matrix of Table 21. Hence the performance of the tool in triggering flag D remains the same with the supplementary function. The performance indicators for flag M with and without the supplementary function are tabulated in Table 24.

Table 24. Performance Indicators for Flag M with and without the Supplementary Function

	With the Supplementary Function				Without the Supplementary Function		
	True Diagnosis Rate	False Alarm Rate	Weak Diagnosis Rate	True Negative Rate	True Diagnosis Rate	False Alarm Rate	True Negative Rate
Small office	13.8%	7.6%	10.5%	86.2%	15.5%	9.4%	90.6%
Medium office	13.1%	13.7%	11.2%	80.8%	21.3%	14.9%	85.1%
Standalone retail	0.7%	3.7%	0.9%	95.4%	0.7%	3.9%	96.1%
Primary school	5.9%	6.2%	0.0%	93.8%	5.9%	6.2%	93.8%
Main-street building	7.7%	13.3%	0.0%	86.7%	7.7%	13.3%	86.7%

The results show that the supplementary function has negligible effect on the true diagnosis, false alarm, and true negative rates for standalone retail buildings, primary schools, and main-street buildings. The supplementary function improves the diagnosis of flag M for small offices as shown by its large weak diagnosis rate. However, the supplementary function negatively affects the diagnosis performance in medium offices because it reduces the true diagnosis rate from 21.3% to 13.1%, and the tool can no longer distinguish between the cases with or without low heating efficiency issues.

To evaluate whether the accuracy of the Tool can be improved by combining flags D and M into one flag, true diagnosis and false alarm rates for the different building types were calculated by considering the diagnoses related to flags D and M to be the same diagnosis. This is shown in the confusion matrix tabulated in Table 25.

Table 25. Confusion Matrix Combining the Diagnoses of Flags D and M into a Single Diagnosis

	Buildings with Inefficient Shell and Ventilation or Equipment with Low Heating Efficiency (D or M)	**Buildings without Inefficient Shell and Ventilation and Equipment with Low Heating Efficiency (D or M)**
Diagnoses with flags D or M	True diagnosis	False alarm
No diagnoses with flags D and M	False negative	True negative

The true diagnosis and false alarm rates are tabulated in Table 26.

Table 26. True Diagnosis and False Alarm Rates after Combining the Diagnoses of Flags D and M into a Single Diagnosis

Building Type	**True Diagnosis Rate**	**False Alarm Rate**
Small office	76.8%	44.4%
Medium office	65.1%	47.3%
Standalone retail	94.7%	74.2%
Primary school	93.6%	86.1%
Main-street building	100.0%	82.2%

In comparing Table 26 and Table 22, the results show that combining flags D and M increases the true diagnosis rates but does not reduce the false alarm rates significantly. This shows that diagnoses with flags D and M are being triggered by building issues other than an inefficient shell and ventilation and inefficient heating equipment. These two flags should be studied more carefully to identify the causes of the high false alarm rates.

7.5 Evaluation Methodology and Results For Multiple Flags That Are Mutually Exclusive (Flags A, B, C, O, and P)

Although a confusion matrix to evaluate multiple flags that are mutually exclusive can be built similarly to Table 18, the evaluation would contain too many performance indicators if the false diagnosis rates were calculated in a similar manner. A new evaluation scheme was determined based on the classification methods that categorize flags A, B, C, O, and P. The classification for these flags considers two methods:

- Electricity baseload per floor area
- Internal electric load relative to the electricity baseload.

If the electricity baseload per floor area is estimated to be lower than 29.1 W/m^2 (2.7 W/ft^2), it will be classified with flags A, B, or C. For buildings with an estimated electricity baseload per floor area higher than 29.1 W/m^2 (2.7 W/ft^2), they will be classified as O or P, according to the dominance of internal electricity load to the total electricity baseload. If the internal electricity load is the dominant factor, the building will be diagnosed with a flag O. Otherwise, flag P will be triggered, indicating the presence of a high exterior electric load.

To assess the accuracy of the Tool in classifying building electricity baseload correctly, the confusion matrix in Table 27 was drawn.

Table 27. Confusion Matrix To Assess the Accuracy of the Tool in Classifying Buildings according to Electricity Baseload

Electricity Baseload per Floor Area	Lower than 9.1 W/m^2 (0.85 W/ft^2)	Between 9.1 W/m^2 (0.85 W/ft^2) and 14.5 W/m^2 (1.35 W/ft^2)	Between 14.5 W/m^2 (1.35 W/ft^2) and 29.1 W/m^2 (2.7 W/ft^2)	Higher than 29.1 W/m^2 (2.7 W/ft^2)
Diagnosed with flag A	Correct diagnosis	False diagnosis	False diagnosis	False diagnosis
Diagnosed with flag B	False diagnosis	Correct diagnosis	False diagnosis	False diagnosis
Diagnosed with flag C	False diagnosis	False diagnosis	Correct diagnosis	False diagnosis
Diagnosed with flag O or P	False diagnosis	False diagnosis	False diagnosis	Correct diagnosis

Since the confusion matrix does not contain any false alarms that are important enough to be evaluated separately, only the accuracy of the tool in categorizing buildings into the correct class was evaluated. The accuracy of the Tool was calculated by equation (22).

$$\begin{matrix}\text{Accuracy to classify building} \\ \text{electricity baseload}\end{matrix} = \frac{\text{Number of correct diagnoses in Table 27}}{\text{Total number of simulation cases in Table 27}} \tag{22}$$

To assess the accuracy of the tool in differentiating between flags O and P, a confusion matrix was developed, as seen in Table 28.

Table 28. Confusion Matrix To Assess the Accuracy of the Tool in Identifying whether Internal Electricity Load Dominates the Total Electricity Baseload

	Internal Electricity Load Dominates the Total Electricity Baseload	External Electricity Load Dominates the Internal Electricity Baseload
Diagnosed with flag O	Correct diagnosis	False diagnosis
Diagnosed with flag P	False diagnosis	Correct diagnosis

Similar to the confusion matrix in Table 27, there are no false alarms, and only the accuracy of correct classification is needed. The accuracy of the Tool in identifying whether the internal electricity load dominates the total electricity baseload was calculated by equation (23).

$$\begin{array}{c}\text{Accuracy to identify whether the}\\ \text{internal electricity load dominates}\\ \text{the total electricity baseload}\end{array} = \frac{\text{Number of correct diagnoses in Table 28}}{\text{Total number of simulation cases in Table 28}} \tag{23}$$

The accuracies from equations (22) and (23) for different building types are tabulated in Table 29.

Table 29. Accuracies to Evaluate How the Tool Diagnoses Flags A, B, C, O, and P

	Accuracy in Classifying Buildings Based on Electricity Baseload	Accuracy in Identifying whether the Internal Electricity Load Dominates the Total Electricity Baseload
Small office	92.0%	65.1%
Medium office	84.9%	72.1%
Standalone retail	85.7%	33.0%
Primary school	78.9%	0.0%
Main-street building	86.9%	52.4%

Table 29 shows that the Tool classified the buildings correctly between 78.9% and 92.0% of the cases. In an effort to understand why the accuracy is higher for small offices than for primary schools, the confusion matrices for the two building types are presented in Table 30 and Table 31.

Table 30. Confusion Matrix To Assess the Accuracy of the Tool in Classifying Buildings Correctly according to the Electricity Baseload for Small Offices

Electricity Baseload per Floor Area	Lower than 9.1 W/m^2 (0.85 W/ft^2)	Between 9.1 W/m^2 (0.85 W/ft^2) and 14.5 W/m^2 (1.35 W/ft^2)	Between 14.5 W/m^2 (1.35 W/ft^2) and 29.1 W/m^2 (2.7 W/ft^2)	Higher than 29.1 W/m^2 (2.7 W/ft^2)
Diagnosed with flag A	65	2	0	0
Diagnosed with flag B	3	109	36	0
Diagnosed with flag C	0	0	251	3
Diagnosed with flag O or P	0	0	6	149

Table 31. Confusion Matrix To Assess the Accuracy of the Tool in Classifying Buildings Correctly according to the Electricity Baseload for Primary Schools

Electricity Baseload per Floor Area	Lower than 9.1 W/m^2 (0.85 W/ft^2)	Between 9.1 W/m^2 (0.85 W/ft^2) and 14.5 W/m^2 (1.35 W/ft^2)	Between 14.5 W/m^2 (1.35 W/ft^2) and 29.1 W/m^2 (2.7 W/ft^2)	Higher than 29.1 W/m^2 (2.7 W/ft^2)
Diagnosed with flag A	14	24	0	0
Diagnosed with flag B	1	145	113	0
Diagnosed with flag C	0	0	300	0
Diagnosed with flag O or P	0	0	7	84

A comparison of Table 30 with Table 31 shows that there are more cases in the upper triangular portion of the confusion matrix in Table 31 than that of Table 30. This is due to an underestimation of electricity baseload in primary schools.

Table 29 also shows that the Tool is less capable in identifying the source of high electricity baseload in standalone retail buildings and primary schools than for the other building types. A possible reason is the uneven internal electricity baseload in standalone retail buildings and primary schools. In both building types, there are particular areas where the internal electricity per area is higher than the rest of the building, including special sales areas in standalone retail buildings and kitchen areas in the primary schools. If the Tool does not estimate the electricity load in these areas correctly, it may consider the electricity as part of an external load and incorrectly classify the building.

7.6 Summary

The Tool's performance in triggering message flags that diagnose building operating issues was analyzed qualitatively and quantitatively. Results from the qualitative analysis concluded that the tool should be modified to output a category of "normal" electricity baseload, and to utilize a classification scheme that changes with different building types. It should also focus its development on issues that have the biggest effect on energy cost savings.

Methods to quantify the performance of the Tool in correctly triggering message flags depend on the individual flags or combination of flags. For each circumstance, different confusion matrices were developed according to the nature of the message flags. The Tool performed best in the following areas:

- Diagnosis of high summer gas use in standalone retail buildings
- Diagnosis of unnecessary reheat in medium offices

- Classification of electricity baseload
- Identification of high electricity baseload in office and main-street buildings.

The accuracy for the other message flags could use improvement. The performance indicators for all of the message flags are summarized in Appendix G for reference.

8 Conclusions

A comprehensive test suite for evaluating the performance of no touch audit tools was developed by NREL and Purdue. The testing approach compares building simulation results to no-touch audit tool estimates for a wide variety of building types, building vintages, climate zones, envelope performance, and building operating issues. Three functions found in no-touch audit tools were evaluated: estimations of building energy end uses, estimations of potential energy savings, and tool alerts that signal envelope performance and building operation issues. The energy end use estimation accuracy was examined by comparing normalized utility costs of electric and gas end uses. The accuracy of potential energy savings estimates was examined by comparing pre- and post-retrofit cases. The tool alerts were evaluated by determining true diagnosis and false alarm rates for each type of alert. The test set developed in this work could be used for additional no touch audit tools that perform in the same or similar manner.

In order to develop the comprehensive test suite and methodology for assessing no-touch audit tools, a representative tool was used as a test case in this study. Results from the assessment show that the tool had fairly good accuracy in estimating energy end uses for all buildings except medium office buildings. Its performance in triggering alerts that signal performance and operation issues varied —the most accurate alerts (also referenced as “message flags”) are summarized in Section 7.6. False alarm rates were generally too high. The magnitude of the tool’s bias and uncertainty for estimating energy end use cost ratios were less than 5% and 10%, respectively, for small offices, standalone buildings, and main-street buildings. The tool underestimates predicted energy savings for most buildings, except for primary schools, where the tool overestimates predicted energy savings. With respect to the various performance and operational alerts, the tool could reliably diagnose issues related to high summer gas use in standalone retail buildings, possible reheat in medium office buildings, and high electricity baseloads in office and main-street buildings. The tool’s true diagnosis and false alarm rates for the other operating issues require improvement before it should be considered reliable. There were also small deviations (less than 1-2%) between the tool estimates and simulation results for total annual energy use in primary schools and main-street buildings, due to the tool’s regression algorithms used to calibrate the model. The testing results may improve if the calibrated, simplified model values are adjusted to match the simulated monthly utility bill data.

In a parallel companion report, NREL is supplementing the simulated evaluation with actual building energy performance data (Cai et al. 2016). NREL plans to compare the energy end use breakdowns and identify possible causes of high energy use from the Tool with submetered field data.

References

Abels, B., Sever, F., Kissock, K., and Ayele, D. 2011. *Understanding industrial energy use through lean energy analysis*. (No. 2011-01-0326). SAE Technical Paper.

America Saves. 2015. "Saving Energy + Money Together." Accessed September 2015: http://www.americasavesenergy.org/.

ANSI BSR/RESNET 1201-2016, *Standard Method of Test for the Evaluation of Building Energy Analysis Model Calibration Methods*. American National Standards Institute/Residential Energy Services Network.

ANSI/ASHRAE Standard 140-2014*, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA.

ASHRAE. 1999. *Ventilation for Acceptable Indoor Air Quality, ANSI/ASHRAE Standard 62.1-1999*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA.

ASHRAE. 2010. *Engineering Analysis of Experimental Data, ASHRAE Guideline 2-2010*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA.

ASHRAE. 2013a. *Climatic Data for Building Design Standards, ANSI/ASHRAE Standard 169-2013*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA.

ASHRAE. 2013b. *Energy Standard for Buildings Except Low-Rise Residential Buildings, ANSI/ASHRAE Standard 90.1-2013*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA.

ASHRAE. 2014. *Measurement of Energy, Demand, and Water Savings, ASHRAE Guideline 14-2014*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA.

ASHRAE. 2015. "Standards & Guidelines – Titles, Purposes and Scopes." Accessed September 2015: https://www.ashrae.org/standards-research--technology/standards--guidelines/titles-purposes-and-scopes.

Cai, J., Braun, J., and Langner, R. 2016. *Methodology to Assess No-Touch Audit Software Using Field Data*. NREL Technical Report. Available from www.nrel.gov/publications.

Cheng, J. and Druzdzel, M. 2000. *Latin Hypercube Sampling in Bayesian Networks*. American Association for Artificial Intelligence. Pittsburgh, PA: University of Pittsburgh. Accessed October 2013 from www.pitt.edu/~druzdzel/psfiles/flairs00.pdf.

Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., Liu, B., Halverson, M., Winiarski, D., Yazdazian, M., Huang, J., and Crawley, D. 2011. *U.S. Department of Energy*

Commercial Reference Building Models of the National Building Stock. NREL/TP-5500-46861, National Renewable Energy Laboratory, Golden, Colorado.

DOE. 2015a. *EnergyPlus*, Department of Energy, Accessed September 2015: https://energyplus.net/.

DOE. 2015b. *OpenStudio*, Department of Energy, Accessed September 2015: https://www.openstudio.net/.

DOE. 2015c. *A Common Definition for Zero Energy Buildings*, Accessed November 2015: http://energy.gov/sites/prod/files/2015/09/f26/bto_common_definition_zero_energy_buildings_093015.pdf.

Encyclopædia Britannica. 2015. Accessed September 2015: http://www.britannica.com/science/summer-season.

EIA. 2015a. *2013 Total Electric Industry - Average Retail Price*. U.S. Energy Information Administration, Accessed March 2015: http://www.eia.gov/electricity/sales_revenue_price/pdf/table4.pdf.

EIA. 2015b. *U.S. Natural Gas Prices*. U.S. Energy Information Administration, Accessed February 2014: http://www.eia.gov/dnav/ng/ng_pri_sum_dcu_nus_a.htm.

Fawcett, T. 2006. "An introduction to ROC analysis." *Pattern Recognition Letters* (27); pp. 861–874.

Granderson et al. 2015. *Assessment of Automated Measurement and Verification (M&V) Methods*. Lawrence Berkeley National Laboratory. July 2015. Accessed September 2015: http://eetd.lbl.gov/sites/all/files/lbnl-187225_0.pdf.

Goel, S., Rosenberg, M., Athalye, R., Xie, Y., Wang, W., Hart, R., Zhang, J., and Mendon, V. 2014, *Enhancements to ASHRAE Standard 90.1 Prototype Building Models*. PNNL-23269, Pacific Northwest National Laboratory. Richland, Washington.

Henze, G. P., Pavlak, G. S., Florita, A. R, Dodier, R. H., and Hirsch, A. I. 2015. "An Energy Signal Tool for Decision Support in Building Energy Systems." *Applied Energy* (138); pp. 51–70.

Hong, T.. Chang, W.. and Lin, H. 2013. "A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data." *Applied Energy* (111); pp. 333–350.

Hygh, J. S., DeCarolis, J. F, Hill, D. B., and Ranjithan, S. R. 2012. "Multivariate regression as an energy assessment tool in early building design." *Building and Environment* (57), pp. 165–175.

Judkoff, R., Wortman, D., and O'Doherty, B. 1983/2008. *A Methodology for Validating Building Energy Analysis Simulations*. NREL/TP-550-42059.

Judkoff, R., and Neymark, J. 1995. *International Energy Agency Building Energy Simulation Test (BESTEST) and Diagnostic Method*. NREL/TP-472-6231.

Judkoff, R., and Neymark, J. 2006. Model Validation and Testing: The Methodological Foundation of ASHRAE Standard 140. ASHRAE Annual Meeting, Quebec City, Canada. Preprint NREL/CP-550-40360.

Judkoff, R., Polly, B., Bianchi, M., and Neymark, J. 2010. *Building Energy Simulation Test for Existing Homes (BESTEST-EX)*. NREL/TP-550-47427.

Judkoff, R., Polly, B., Bianchi, M., Neymark, J., Kennedy,M. 2011. *Building Energy Simulation Test for Existing Homes (BESTEST-EX): Instructions for Implementing the Test Procedure, Calibration Test Reference Results, and Example Acceptance-Range Criteria.* NREL/TP-5500-52414.

Judkoff, R., Polly, B., and Neymark, J. 2016. "A Method to Test Model Calibration Techniques." *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings*, Pacific Grove California. Preprint, NREL/CP-5500-65981.

Kissock, K., Haverl, J., and Claridge, D. 2001. *Development of a Toolkit for Calculating Linear, Change-point Linear and Multiple-Linear Inverse Building Energy Analysis Models*. Final Report, ASHRAE 1050-RP. December.

Kissock, K. and Mulqueen, S. 2008. *Targeting Energy Efficiency in Commercial Buildings Using Advanced Billing Analysis*. ACEEE Summer Study on Energy Efficiency in Buildings. August 17–22, 2008.

Langner et al. 2014. *Reducing Transaction Costs for Energy Efficiency Investments and Analysis of Economic Risk Associated with Building Performance Uncertainties*. National Renewable Energy Laboratory. NREL/TP-5500-60976. Accessed September 2015: http://www.nrel.gov/docs/fy14osti/60976.pdf.

Lee, S. H., Hong, T., Piette, M. A., and Taylor-Lange, S. C. 2015. "Energy retrofit analysis toolkits for commercial buildings: A review." *Energy* (89), pp. 1087–1100.

Montgomery, D. C. 2004. *Design and analysis of experiments*. John Wiley & Sons.

Neymark, J., and Judkoff, R. 2002. *International Energy Agency Building Energy Simulation Test and Diagnostic Method for Heating, Ventilating, and Air-Conditioning Equipment Models (HVAC BESTEST); Volume 1: Cases E100-E200.* NREL/TP-550-30152.

Neymark, J., Judkoff, R., Beausoleil-Morrison, I., Ben-Nakhi, A., Crowley M., Deru, M., Henninger, R.,Ribberink, H., Thornton, J., Wijsman, A., Witte, M. 2008. *International Energy Agency Building Energy Simulation Test and Diagnostic Method (IEA BESTEST): In-Depth Diagnostic Cases for Ground Coupled Heat Transfer Related to Slab-on-Grade Construction.* NREL/TP-550-43388.

Neymark, J., Judkoff, R., Alexander, D., Felsmann, C., Strachan,P., Wijsman, A. 2008. *International Energy Agency Building Energy Simulation Test and Diagnostic Method (IEA BESTEST) Multi-Zone Non-Airflow In-Depth Diagnostic Cases: MZ320 -- MZ360.* NREL/TP-550-43827.

Neymark, J., Kennedy, M., Judkoff, R., Gall, J., Knebel,D., Henninger, R., Witte, M., Hong, T., McDowell, T., Yan, D., Zhou, X. 2016. *Airside HVAC BESTEST: Adaptation of ASHRAE RP 865 Airside HVAC Equipment Modeling Test Cases for ASHRAE Standard 140. Volume 1, Cases AE101-AE445.* NREL/TP-5500-66000.

NMSC. 2015. *Welcome to the National Main Street Center, Inc*. Accessed September 2015: http://www.preservationnation.org/main-street/?referrer=https://www.google.com/#.Vfd0HmRViko.

NREL. 2002. *International Performance Measurement & Verification Protocol Volume I: Concepts and Options for Determining Energy Savings*. National Renewable Energy Laboratory. NREL TP-710-31505.

R Foundation. 2015. "The R Project for Statistical Computing." *R Foundation*, Accessed September 2015: https://www.r-project.org/.

Reichmuth, H. and Turner, C. 2010. *A Tool for Efficient FirstView of Commercial Building Energy Performance*. 2010 ACEEE Summer Study on Energy Efficiency in Buildings, 325–338.

Ruby. 2015. "Ruby Programming Language." Accessed September 2015: https://www.ruby-lang.org/en/.

Sun, Y. and Augenbroe, G. 2014. "Urban heat island effect on energy application studies of office buildings." *Energy and Buildings* (77), pp. 171–179.

Shrestha, S. S., Bhandari, M. S., Biswas, K., and Desjarlais, A. O. 2014. *Lifetime Energy and Environmental Impacts of Insulation Materials in Commercial Building Applications–Assessment Methodology and Sample Calculations*. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Thornton, B. A., Rosenberg, M. I., Richman, E. E., Wang, W., Xie, Y., Zhang, J., Cho, H., Mendon, V. V., Athalye, R. A., and Liu, B. 2010. *Achieving the 30% Goal: Energy and Cost Savings Analysis of ASHRAE Standard 90.1-2010*. PNNL-20405, Pacific Northwest National Laboratory. Richland, Washington.

Wilcox, S. and Marion, W. 2008. *Users Manual for TMY3 Data Sets*. Technical Report, NREL/TP-581-43156, National Renewable Energy Laboratory, Golden, Colorado.

Zhang, J., Liu, G., Lutes, R. G., and Brambley, M. R. 2013. *Energy Savings for Occupancy-Based Control (OBC) of Variable-Air-Volume (VAV) Systems*. PNNL-22072, Pacific Northwest National Laboratory. Richland, Washington.

Appendix A. Small Office, Medium Office, Standalone Retail, and Primary School Building Model Descriptions

This appendix briefly describes the configuration of the Commercial Reference Building Models (Deru et al. 2011) and the Commercial Prototype Building Models (Thornton et al. 2011; Goel et al. 2014) for small office, medium office, standalone retail, and primary school buildings. For additional details regarding electricity load densities in the thermal zones and the reasoning behind the definition of the materials, see Deru et al. (2011), Thornton et al. (2011), and Goel et al. (2014).

A.1 Small Office Buildings

The small office model simulates a typical small office building found in the United States. It represents a single-story building with a floor area of 511 m^2 (5,500 ft^2). Its internal space is divided into five thermal zones as shown in Figure A-1.

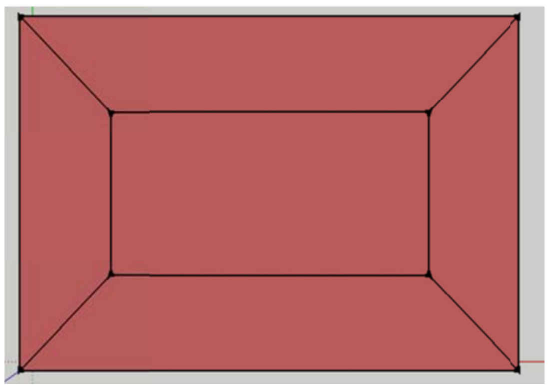


Figure A-1. Division of thermal zones for the small office model (Thornton et al. 2011)

All thermal zones are considered for office use, and the HVAC load densities, occupancy density schedules, and electricity load densities are identical. The internal equipment power densities and building envelope characteristics change according to the vintages, showing the advances of building technologies through time. The HVAC equipment for the different vintages is tabulated in Table A-1.

Table A-1. HVAC Equipment in the Small Office Models Based on Vintage

Vintages	Heating Equipment	Cooling Equipment	Water Heater	Air Distribution
Pre-1980 and post-1980	Gas furnaces	Packaged air conditioners	Natural gas water heaters	Single-zone constant air volume (CAV) system
2004 and 2013	Air-source heat pumps with supplemental gas heat	Air-source heat pumps	Electric water heaters	

A.2 Medium Office Buildings

The medium office model represents a typical medium office building in the United States. It represents a three-story office building with a total floor area of 4,982 m^2 (53,628 ft^2). It is divided into 15 thermal zones with 5 thermal zones on each floor as shown in Figure A-2.

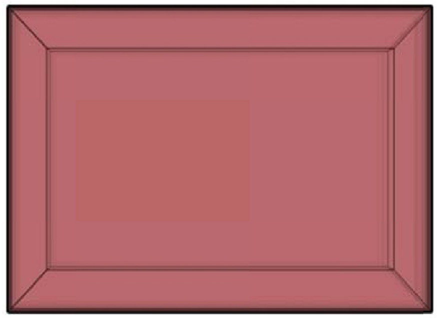


Figure A-2. Division of thermal zones for the medium office model (Thornton et al. 2011)

Similar to the small office model, all thermal zones are considered for office use, and the occupancy schedules, load densities, etc., per zone are the same. The HVAC equipment for different vintages is listed in Table A-2.

Table A-2. HVAC Equipment in the Medium Office Models Based on Different Vintages

Vintages	Heating Equipment	Cooling Equipment	Water Heater	Air Distribution
Pre-1980	Gas furnaces	Packaged air conditioners	Natural gas water heaters	Single-zone VAV system
Post-1980, 2004, and 2013				Multi-zone VAV system with electric reheat

A.3 Standalone Retail Buildings

The standalone retail models simulate the operation of small retail stores that occupy an individual building in the United States. The model simulates one floor with a floor area of 2,294 m^2 (24,962 ft^2), and its space is divided into four thermal zones as shown in Figure A-3.

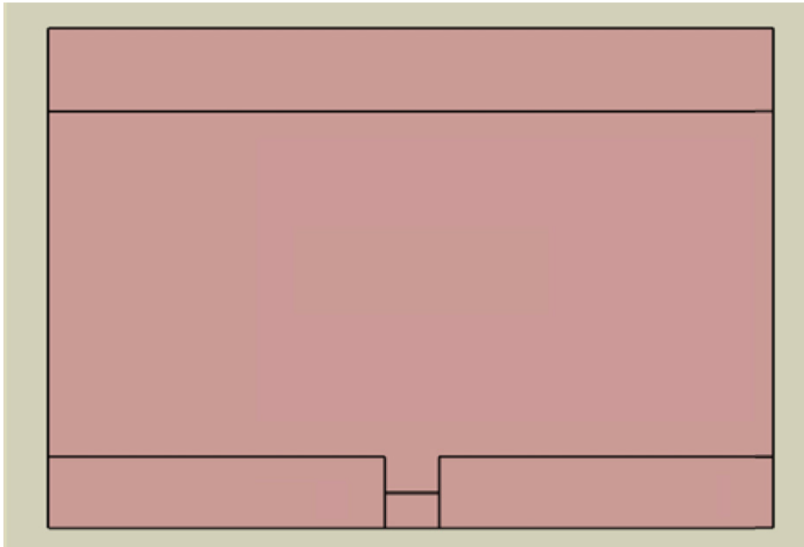


Figure A-3. Division of thermal zones for the standalone retail model (Thornton et al. 2011)

The division of thermal zones in Figure A-3 is due to the different functions of each building area. The smallest thermal zone serves as the entrance to the retail store and its HVAC equipment consists of a unit heater. The small thermal zones to the left of the entrance serve as a special sales area and have a higher electric load density than the main sales area (large middle zone). The top thermal zone represents storage or inventory space and has a lower electric load and occupancy densities than the main sales area.

The HVAC equipment simulated in the standalone retail model are generally the same despite the building vintage. Besides the front entrance thermal zone, the other zones are supported by packaged air conditioners with gas furnaces and constant air volume systems. The only difference between the old (pre-1980 and post-1980) and new buildings (2004 and 2013), is the use of electric heaters versus gas furnaces in the front entrance, respectively.

A.4 Primary School Buildings

The primary school model simulates a typical primary education building in the United States. It consists of one floor with a floor area of 6,871 m^2 (73,960 ft^2). It is divided into 25 zones as shown in Figure A-4.

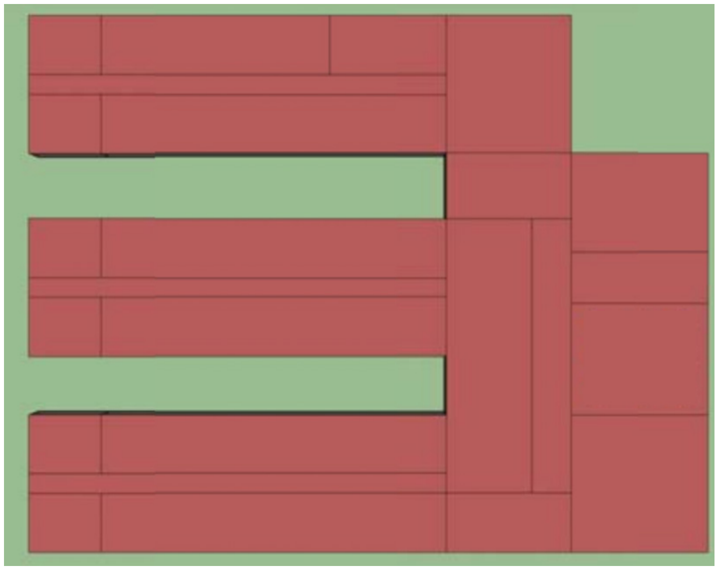


Figure A-4. Division of thermal zones for the primary school model (Thornton et al. 2011)

The different thermal zones are used for different functions as tabulated in Table A-3.

Table A-3. Functions of Thermal Zones in a Primary School Model

Function	Number of Thermal Zones
Classrooms	12
Computer room	1
Corridors	4
Administrative area	1
Gymnasium	1
Mechanical room	1
Media center	1
Lobby	1
Kitchen	1
Cafeteria	1
Bathroom	1

The HVAC equipment in the primary school models is tabulated in Table A-4.

Table A-4. HVAC Equipment in the Primary School Models Based on Different Vintages

Vintages	Heating Equipment	Cooling Equipment	Water Heater	Air Distribution
Pre-1980 and post-1980	Boiler	Packaged air conditioners	Natural gas water heaters	Mix of CAV and VAV systems with hot water reheat
2004 and 2013	Boiler and gas furnaces			

Appendix B. Description of the Main-Street Building Model

A main-street building model was developed in OpenStudio and manipulated to create multiple simulation cases to test the accuracy of the Tool on historic downtown buildings. The model represents a two-story building with a retail store on the ground floor and an inventory space at the back, and office space on the second floor. Because main-street buildings are usually older or historic buildings, its building envelope, occupancy schedule, lighting, electric load density, etc., reference to the pre-1980 load profiles for the small office and standalone retail reference models (Deru et al. 2011). A drawing of its configuration is shown in Figure B-1.

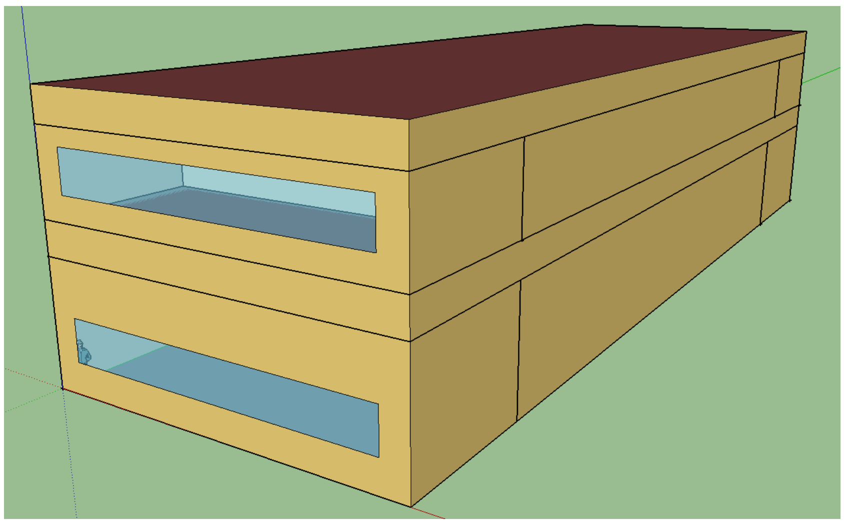


Figure B-1. 3-D drawing of the main-street building model

To illustrate the difference in building performance at different locations, TMY3 weather data files (Wilcox and Marion 2008) for four locations (listed in

Table 4) and affiliated construction materials (Deru et al. 2011) were used. To represent main-street buildings that have been retrofitted with new HVAC equipment, the models were modified to simulate the HVAC equipment in the 2013 small office and standalone retail models (Goel et al. 2014). In total, eight baseline models were created to represent main-street buildings with old and new HVAC equipment in four climate zones.

A detailed description of the pre-1980 baseline model is tabulated in Table B-1.

Table B-1. Detailed Description of the Pre-1980 Main-Street Building Model

Form	
Floor area	1,191 m^2
Number of floors	2
Overall dimension	14.09 m x 42.27 m x 10.06 m
Window area	12.09 m x 1.52 m
Window locations	All exposed surface (excluding exposed surface of the plenum floor and the back side of the ground floor)
Shading geometry	None
Boundary conditions	Adiabatic on the two sides
Thermal zoning	Three thermal zones on the top floor and three thermal zones on the ground floor
Floor to ceiling height	4.57 m on the ground floor and 3.05 m on the upper floor
Floor to floor height	5.79 m on the ground floor and 4.27 m on the upper floor
Glazing sill height	1.14 m on the ground floor and 0.9 m on the upper floor
Construction	
Exterior walls	Steel framed wall Wood siding + Steel framed wall insulation + 0.5 inch gypsum
Roof	Insulation above deck Roof membrane + insulation + metal decking
Ceiling below roof	Uninsulated stud wall
Plenum interface	Carpet and concrete mass wall
Foundation	Carpet and concrete mass wall
Internal furnishing	6-in. standard wood
Infiltration per exterior surface	0.001133 $m^3/s\text{-}m^2$
HVAC	
Heating equipment	One gas furnace per thermal zone
Cooling equipment	One packaged air conditioner per thermal zone
Ventilation	One economizer for outdoor air ventilation per thermal zone on the ground floor

Air distribution	One CAV per thermal zone
Supply fan	Pressure rise at 622 Pa with efficiency at 65%
Sizing	Autosized to design day
HVAC efficiency	Equipment efficiency on the upper floor follows pre-1980 small office building Equipment efficiency on the ground floor follows pre-1980 retail building
HVAC Control	
Thermostat set point	24.0°C cooling/ 21.0°C heating
Thermostat setback	26.7°C cooling/ 15.6°C heating
Supply air temperature	Maximum 50°C, minimum 10°C
Economizer control	Algorithm based on dry bulb temperature difference between outdoor air and return air
Ventilation	Algorithm follows outdoor air requirement in ASHRAE Standard 62.1-1999
Service water heating	
Storage tank volume	0.1764 m^3
Fuel	Natural gas
Thermal efficiency	80%
Maximum capacity	984,870 W
Water temperature set point at outlet	60°C
Water peak flow rate	0.0000631 m^3/s
Occupancy	
Office space	18.58 m^2/person
Retail space	6.19 m^2/person
Retail back space	27.87 m^2/person
Schedules	Occupancy schedule on the upper floor follows pre-1980 small office building Occupancy schedule on the ground floor follows pre-1980 retail building
Lighting	
Office space	19.48 W/m^2
Retail space	36.25 W/m^2
Retail back space	12.55 W/m^2
Schedules	Lighting control schedule on the upper floor follows

	pre-1980 small office building Lighting control schedule on the ground floor follows pre-1980 retail building
Electric plug load	
Office space	10.76 W/m^2
Retail space	3.23 W/m^2
Retail back space	8.07 W/m^2
Schedules	Electric plug load control schedule on the upper floor follows pre-1980 small office building Electric plug load control schedule on the ground floor follows pre-1980 retail building
Exterior lighting	
Peak power	5,717 W
Schedule	Lights on in the evening

To model the retrofitted 2013 main-street building, changes to the HVAC equipment, water heater and water fixture, lighting technologies, and electric plug loads were made to the pre-1980 building model. This involves the modeling of more efficient HVAC equipment, reduction of peak water flow rates, and reduction of electricity power consumption. The changes are tabulated in Table B-2.

Table B-2. Changes To Create the 2013 Main-Street Building Model

HVAC	
Heating equipment	One gas furnace per thermal zone on the ground floor and one air-source heat pump per thermal zone on the upper floor
Cooling equipment	One packaged air conditioner per thermal zone on the ground floor and one air-source heat pump per thermal zone on the upper floor
HVAC efficiency	Equipment efficiency on the upper floor follows 2013 small office building Equipment efficiency on the ground floor follows 2013 retail building
Service water heating	
Storage tank volume	0.3528 m^3
Fuel	Natural gas
Thermal efficiency	82%
Maximum capacity	16,719 W
Water temperature set point at outlet	60°C
Water peak flow rate	0.000022948 m^3/s
Lighting	
Office space	8.83 W/m^2
Retail space	15.5 W/m^2
Retail back space	7.57 W/m^2
Electric plug load	
Office space	6.78 W/m^2
Exterior lighting	
Peak power	2,325 W

Appendix C. Test Matrices To Create Simulation Test Set

This appendix lists the test matrices, using the methods discussed in Section 4.2, to create the simulation results inTable 10. For small and medium office buildings, variables in each baseline model were perturbed with multipliers according to Table C-1, creating 39 simulation results for each baseline model.

Table C-1. Baseline Model Perturbations for Small and Medium Office Buildings

Internal Electricity Load Multiplier	Heating Set Point Offset	Cooling Set Point Offset	External Electricity Load Multiplier	Hot Water Peak Flow Multiplier	Infiltration Airflow Multiplier	Gas Furnace Efficiency Multiplier	HVAC Equipment Size Multiplier
0.50	0	0	1	1	1	1	1
0.50	0	-1	1	3	1	1	1.50
0.50	1	0	1	3	1	1	1.50
1	0	0	1	1	1	0.50	1
1	0	0	1	1	1	0.75	1
1	0	0	0.50	1	1	1	1
1	0	-2	1	1	1	1	1
1	0	-1	1	1	1	1	1
1	-3	0	1	1	1	1	1
1	-2	0	1	1	1	1	1
1	1	0	1	1	1	1	1

1	2	0	1	1	1	1	1
1	3	0	1	1	1	1	1
1	0	0	3	1	1	1	1
1	0	0	5	1	1	1	1
1	0	0	8	1	1	1	1
1	0	0	10	1	1	1	1
1	0	0	1	3	1	1	1
1	0	0	1	5	1	1	1
1	0	0	1	1	3	1	1
1	0	0	10	5	1	0.75	1.50
1	2	0	10	5	1	0.75	1.50
1	0	0	1	1	1	1	1.50
1	0	0	1	1	1	1	2
1	0	-1	3	3	3	1	2
1	0	0	3	3	3	1	2
2	0	0	1	1	1	1	1
3	0	-2	0.50	5	1	0.50	1
3	-3	0	0.50	5	1	0.50	1
3	0	0	8	3	3	0.75	1
3	3	0	8	3	3	0.75	1
3	0	0	1	1	1	1	1
5	0	0	1	1	1	1	1
5	0	-1	8	1	3	1	1
5	1	0	8	1	3	1	1
7	0	0	1	1	1	1	1
7	0	-2	3	1	1	0.50	2
7	-3	0	3	1	1	0.50	2

For standalone retail buildings, because the pre-1980 and post-1980 models do not simulate any hot water consumption, the perturbations applied to the pre-1980 and post-1980 baseline models were different from those applied to the 2004 and 2013 baseline models. For pre-1980 and post-1980 baseline models, the list of changes in each of their baseline models is shown in Table C-2.

Table C-2. Perturbations Applied to the Pre-1980 and Post-1980 Standalone Retail Baseline Models

Internal Lighting Load Multiplier	Heating Set Point Offset	Cooling Set Point Offset	External Electricity Load Multiplier	Infiltration Airflow Multiplier	Gas Furnace Efficiency Multiplier	HVAC Equipment Size Multiplier
0.20	0	-2	10	1	0.75	1.50
0.20	-2	0	10	1	0.75	1.50
0.20	0	-2	8	3	0.75	1
0.20	0	0	1	1	1	1
0.20	-3	0	8	3	0.75	1
0.40	-3	0	10	3	1	1.50
0.40	0	0	1	1	1	1
0.40	0	-2	10	3	1	1.50
1	-3	0	1	1	1	1
1	0	0	1	1	0.50	1
1	2	0	1	1	1	1
1	0	0	3	1	1	1
1	0	0	1	3	1	1
1	0	0	1	1	0.75	1
1	0	-2	1	1	1	1
1	0	0	1	1	1	2
1	0	-1	1	1	1	1
1	2	0	1	1	1	2
1	0	0	1	1	1	1.50
1	0	0	0.50	1	1	1
1	3	0	1	1	1	1
1	0	0	8	1	1	1
1	1	0	1	1	1	1
1	0	0	10	1	1	1
1	-2	0	1	1	1	1
1	0	0	5	1	1	1
1	0	0	1	1	1	2
2	0	0	1	1	1	1

2	0	-1	0.50	3	0.50	2
2	0	0	1	3	0.75	1.50
2	3	0	1	3	0.75	1.50
2	1	0	0.50	3	0.50	2
3	0	0	5	1	0.50	1
3	0	-1	5	1	0.50	1
3	0	0	1	1	1	1
5	0	0	1	1	1	1
5	3	0	3	1	1	1
5	0	0	3	1	1	1

The 2004 and 2013 standalone retail building models simulate a water heater and the hot water peak flow was changed. The same type of perturbations were also be made to all of the main-street building baseline models. These perturbations are listed in Table C-3.

Table C-3. Perturbations Applied to the Baseline Models for the 2004 and 2013 Standalone Retail and Main-Street Buildings

Internal Lighting Load Multiplier	Heating Set Point Offset	Cooling Set Point Offset	External Electricity Load Multiplier	Hot Water Peak Flow Multiplier	Infiltration Airflow Multiplier	Gas Furnace Efficiency Multiplier	HVAC Equipment Size Multiplier
0.20	-3	0	10	1	1	0.75	2
0.20	0	-2	10	1	1	0.75	2
0.20	0	0	1	1	1	1	1
0.40	1	0	0.50	5	1	0.50	1.50
0.40	2	0	1	3	3	1	1.50
0.40	0	0	1	1	1	1	1
0.40	0	0	1	3	3	1	1.50
0.40	0	-1	0.50	5	1	0.50	1.50
1	0	0	1	1	1	1	1
1	0	0	1	1	1	1	2
1	0	0	1	1	1	0.50	1
1	-3	0	5	1	3	0.50	1
1	0	0	5	1	1	1	1
1	0	0	10	1	1	1	1
1	1	0	1	1	1	1	1
1	0	0	1	5	1	1	1

1	-3	0	1	1	1	1	1
1	0	0	1	1	1	0.75	1
1	0	0	3	1	1	1	1
1	0	-1	1	1	1	1	1
1	2	0	1	1	1	1	1
1	0	0	1	3	1	1	1
1	0	-2	1	1	1	1	1
1	0	0	8	1	1	1	1
1	0	0	1	1	1	1	1.50
1	0	0	0.50	1	1	1	1
1	0	0	1	1	3	1	1
1	-2	0	1	1	1	1	1
1	3	0	1	1	1	1	1
1	0	-2	5	1	3	0.50	1
2	0	0	1	1	1	1	1
3	1	0	8	5	1	1	1
3	0	-1	8	5	1	1	1
3	0	0	5	3	3	1	1
3	3	0	5	3	3	1	1
3	0	0	1	1	1	1	1
5	0	0	1	5	1	0.75	2
5	0	0	1	1	1	1	1
5	0	-1	1	5	1	0.75	2

The rated cooling equipment COPs in the baseline primary school models were perturbed to create more cases with inefficient cooling equipment COPs. These perturbations are listed in Table C-4.

Table C-4. Perturbations Applied to the Primary School Baseline Models

Internal Lighting Load Multiplier	Heating Set Point Offset	Cooling Set Point Offset	External Electricity Load Multiplier	Hot Water Peak Flow Multiplier	Infiltration Airflow Multiplier	Gas Furnace Efficiency Multiplier	HVAC Equipment Size Multiplier	Cooling COP Multiplier
0.20	0	0	1	1	1	1	1	1
0.20	0	-1	3	1	3	0.50	1.5	1
0.20	0	0	3	1	3	0.50	1.5	1
0.40	3	0	8	5	1	0.75	1	0.90
0.40	2	0	10	3	3	0.50	2	0.80
0.40	0	0	10	3	3	0.50	2	0.80
0.40	0	0	8	5	1	0.75	1	0.90
0.40	0	0	1	1	1	1	1	1
1	0	0	1	1	1	1	1	1
1	0	0	1	1	1	0.75	1	1
1	0	0	3	1	1	1	1	1
1	-2	0	1	1	1	1	1	1
1	0	-1	1	1	1	1	1	1
1	0	0	1	1	3	1	1	1
1	0	0	1	1	1	1	1	0.80
1	0	0	10	1	1	1	1	1
1	0	-2	1	1	1	1	1	1
1	0	0	1	1	1	1	1.50	1
1	0	0	0.50	1	1	1	1	1
1	0	0	5	1	1	1	1	1
1	0	0	1	1	1	1	2	1
1	2	0	1	1	1	1	1	1
1	0	0	1	5	1	1	1	1

1	0	0	1	1	1	1	1	0.90
1	3	0	1	5	1	0.75	2	0.80
1	0	0	1	5	1	0.75	2	0.80
1	3	0	1	1	1	1	1	1
1	-3	0	1	1	1	1	1	1
1	0	0	1	3	1	1	1	1
1	1	0	1	1	1	1	1	1
1	0	0	1	1	1	0.50	1	1
1	0	0	8	1	1	1	1	1
2	0	0	1	1	1	1	1	1
2	0	0	0.50	5	3	0.50	1	0.90
2	0	-1	0.50	5	3	0.50	1	0.90
3	1	0	0.50	3	1	1	2	0.80
3	0	-1	0.50	3	1	1	2	0.80
3	-2	0	5	1	1	1	1	0.90
3	0	0	1	1	1	1	1	1
3	0	-2	5	1	1	1	1	0.90
5	0	0	1	1	1	1	1	1
5	-3	0	5	3	3	0.75	1.5	1
5	0	-2	5	3	3	0.75	1.5	1

Appendix D. Criteria to Assign Flags to the Building Simulation Results

This appendix explains the mathematical criteria to determine the correct message flags associated with each building simulation result. To examine if the Tool triggers its message flags correctly, correct message flags should be assigned to the building simulation results from Table 10 according to the definitions of the message flags, and the assignment should be compared with the Tool estimates to make sure they match. The assignment, except A, B, C, O, P, and N, is conducted by examining the building simulation cases according to the following criteria:

- Is there a significant issue?
- Is the cost of energy of the building changed significantly by the issue?

To programmatically examine the criteria, the criteria are interpreted mathematically. Because each flag has its own definition, the mathematical expressions of the criteria for each flag differ. Flags of different natures (flags A, B, C, O, P, and N) use a different set of mathematical criteria to determine if a building simulation result is associated with them.

Flags A, B, and C

The assignment of flag A, B, and C depends on the electricity baseload per floor area as shown in Table D-1 according to the documentation of the Tool.

Table D-1. Definition of Flags A, B, and C according to the Documentation of the Tool

Flag	Electricity Baseload per Floor Area
A	Lower than 9.15 W/m^2 (0.85 W/ft^2)
B	Between 9.15 W/m^2 (0.85 W/ft^2) and 14.53 W/m^2 (1.35 W/ft^2)
C	Between 14.53 W/m^2 (1.35 W/ft^2) and 29.06 W/m^2 (2.7 W/ft^2)

The assignment of the message flags to the building simulation results is conducted according to the electricity baseload per floor area. Unlike the electricity baseload in the Tool that is estimated from the tool inputs, the electricity baseload in the building simulation results is calculated by the electricity plug load defined in the input files to EnergyPlus, and the electricity baseload per floor area is calculated by equation (D-1).

$$\begin{array}{c}\text{Electricity baseload}\\ \text{per floor area}\end{array} = \frac{\text{Annual internal electricity gain} + \text{Annual exterior lighting}}{\text{Floor area}} \qquad \text{(D-1)}$$

Flags O and P

The assignments of flags O and P to the building simulation results are similar to that of flags A, B, and C—the assignment depends on the electricity baseload per floor area. Flags O and P can only be assigned to buildings with electricity baseload per floor area from equation (D-1) greater than 29.06 W/m^2 (2.7 W/ft^2). However, flags O and P are mutually exclusive and only one of them can be assigned each time. The selection of flag O or P can be done by interpreting the difference between flags O and P mathematically. Flag O refers to buildings in which high electricity baseload is caused by an increase of internal electricity load, and flag P refers to buildings in which high electricity baseload is caused by an increase of external electricity load. To determine if the internal electricity load is the dominant factor of the high electricity baseload, the ratio in equation (D-2) can be calculated.

$$\frac{\text{Annual internal electricity gain - Annual internal electricity gain of the standard efficient case}}{\text{Annual electricity baseload - Annual electricity baseload of the standard efficient case}} \quad \text{(D-2)}$$

where the standard efficient case is defined in Section 6.2.

If the ratio in equation (D-2) is higher than 0.5, the internal electricity gain is the major cause of the high electricity baseload and the flag O should be assigned to the building simulation result. Otherwise, flag P should be assigned.

Flag D

Flag D refers to buildings with inefficient shell and ventilation. To examine if flag D should be assigned to a building simulation result, the infiltration airflow of a simulation must first be increased according to Table 5. Its annual infiltration heat transfer gain and annual infiltration heat transfer loss are also examined to see if their magnitudes differ from that of the baseline case by more than 10%, as defined in Section 6.2. This tests if the related issue, inefficient shell and ventilation, is significantly triggered by the perturbation—the increase of infiltration airflow. The annual energy cost of the result is also examined to see if the issue significantly increases the energy cost of the building. This is done by checking if the cost increase from the annual energy cost of the baseline case by increased infiltration is higher than 5% of the annual energy cost of the standard efficient case. The calculation is conducted by equation (D-3).

$$\frac{\text{Annual energy cost - Annual energy cost of the baseline case}}{\text{Annual energy cost of the standard efficient case}} \quad \text{(D-3)}$$

The reason to normalize equation (D-3) by the annual energy cost of the standard efficient case is to emphasize the significance of energy cost in buildings with a costly baseline case. This can be exemplified by considering the two buildings in the same climate zones of the same type of building in Table D-2.

Table D-2. Example Comparison of Two Buildings with Infiltration Airflow Increased by 200% in the Same Climate Zone of the Same Building Type

Building	Vintage	Energy Cost [$/m^2]	Energy Cost of the Baseline Case [$/m^2]	Energy Cost of the Standard Efficient Case [$/m^2]	Percentage Increase from the Baseline Case	Ratio from Equation (D-3)
1	pre-1980	30	25	20	20%	25%
2	2013	24	20	20	20%	20%

Table D-2 shows that the significance of the infiltration airflow increase to the cost of two buildings is the same if the percentage increase from the baseline case is considered directly. However, because the same type of buildings in the same climate zone are considered, the higher energy cost of building 1 than that of building 2 should be emphasized. This emphasis is brought by normalizing the difference of energy cost by the energy cost of the standard efficient case instead of the cost of the baseline case. This results in the last column in Table D-2, which shows that the cost increase in building 1 is more significant than that of building 2.

Flag E

Flag E is triggered when there is inefficient cooling COP in the building and is only assigned to simulation cases with reduced HVAC equipment size or rated cooling equipment COP. This ensures that the equipment is degraded by perturbations that are related to cooling equipment operating efficiency. The cooling cost per cooling load of the building simulation result is examined to see if it is at least 10% higher than that of the baseline case to ensure that the cooling equipment COP is degraded significantly by the perturbations. Lastly, similar to flag D, its annual energy cost is compared with that of the baseline case by equation (D-3), and the result from equation (D-3) must be greater than 10% to ensure that the cooling COP degradation is significant to the overall building operation. If a building simulation result passes these tests, it is assigned with a flag E to indicate that its cooling equipment has a low COP.

Flag M

Flag M refers to buildings with inefficient heating equipment. The existence of the issue in a building simulation result can be confirmed by first checking if the gas furnace efficiency is reduced or if the HVAC equipment size has been increased. If the building model is changed by either of these perturbations, the total heating cost per heating load in the result is calculated. If the total heating cost per heating load has been increased by more than 10% from the baseline case, the inefficient heating operation can be confirmed and the ratio in equation (D-3) is calculated. If the ratio is greater than 5%, the significant increase of energy cost by the low heating efficiency can be confirmed and the building simulation result should be associated with a flag M.

Flag H

Flag H refers to buildings that deliver much more heating than necessary. To examine if flag H should be assigned to a building simulation result, the thermostat heating set point is checked to ensure that it is higher than that of the baseline case. If the thermostat heating set point is increased from that of the baseline model, the heating load delivered by the heating equipment in the building simulation result is compared with that of the baseline case. If the heating load is 10% higher than that of the baseline case, it is proved that the change of thermostat set point increases the heating load significantly. The ratio in equation (D-3) is then calculated. If the ratio is higher than 5%, it shows that the energy cost of the building is significantly affected by the increase of heating load, and flag H is assigned to the building simulation result.

Flag I

Flag I is triggered when the cooling use in the building is higher than necessary. The existence of the issue in the building simulation result is examined by first checking if the thermostat cooling set point is lower than that of the baseline case. If it is, the annual average cooling load per floor area in the result is compared to that of the baseline case. If the cooling load has increased by more than 10% from the baseline case, the energy cost per floor area in the result is examined. If the ratio in equation (D-3) is higher than 5%, the energy cost is considered to be significantly increased by the high cooling use, and flag I is assigned to the building.

Flag K

Flag K refers to high summer gas use in the building. A building simulation result is associated with flag K only if its perturbation is related to high summer gas use according to Table 5—an increase of hot water consumption, an increase of thermostat heating set point, or a reduction of gas furnace efficiency. If the perturbation exists, its gas cost during summertime (from June 21 to September 23, Encyclopædia Britannica 2015) is calculated. If the increase of gas cost from the baseline case is higher than 0.5% of the summer energy cost of the baseline case, the perturbation is said to have increased the summer gas cost significantly and the ratio in equation (D-3) is calculated based on summer energy cost rather than annual energy cost. Flag K is only assigned to a building simulation if the ratio is higher than 5% to prove that the increase of summer gas cost is important to the building owners.

Flag L

Flag L refers to unnecessary reheat operation when the average daily temperature is higher than 18.3°C (65°F). Only results with the increase of heating thermostat set point shown in Table 5 may be associated with flag L. Because not all types of buildings tested contain reheat devices, only buildings that have reheat devices listed in Table D-3 may be associated with flag L.

Table D-3. List of Buildings Containing Reheat Devices

Vintage\Building Type	Small Office	Medium Office	Standalone Retail	Primary School	Main-Street Building
Pre-1980	No	No	No	Yes	No
Post-1980	No	Yes	No	Yes	No
2004	No	Yes	No	Yes	No
2013	No	Yes	No	Yes	No

If a building has its thermostat heating set point increased and has a reheat device according to Table D-3, its average heating cost per day when the daily average temperature is higher than 18.3°C (65°F) is examined. If the cost has been increased from that of the baseline case for more than 0.5% of the total energy cost during the high temperature period, the reheat issue exists. However, to know if the issue is significant enough to be communicated to the building owners, the ratio in equation (D-3) is calculated with the average energy cost per day during the high temperature period in the standard efficient case as the denominator. If the ratio is higher than 5%, the reheat issue can be said to have increased the building energy cost significantly, and flag L should be assigned to the simulation result.

Flag Q

Flag Q refers to buildings in which heating load is lower than average. To test if flag Q should be assigned to a building simulation result, the criteria in Table D-4 are checked.

Table D-4. Criteria To Assign Flag Q to a Building Simulation Result

Criteria	Variable	Value
Type of perturbation	Lowering of heating thermostat set point	Exists
Significance of heating load to perturbation	Annual heating delivered by heating equipment	10% lower than that of the baseline case
Significance of energy cost to heating load	Ratio in equation (D-3)	Lower than -5%

If all criteria in Table D-4 are met, the building simulation result will be assigned with flag Q.

Flag N

Flag N refers to possible erratic operation or occupancy in buildings. The process of assigning flag N is discussed in Section 4.2.

Appendix E. Calculation of Bias and Uncertainty To Assess Accuracy of an Estimation Method

Bias and uncertainty are statistical representations of the deviation between estimates of a variable and their corresponding correct values. Bias describes the average deviation between an estimate of a variable and its correct value, and uncertainty quantifies the repeatability of the bias for each estimate. A method with high magnitude of bias but small uncertainty is a method that either overestimates or underestimates the variable consistently, and a method with small magnitude of bias but large uncertainty is a method that can only correctly estimate the variable occasionally. An accurate estimation method should give unbiased estimation consistently and its magnitudes of bias and uncertainty should be small.

To assess the accuracy of an estimation method of a variable by calculating bias and uncertainty, a good sample of variables and their corresponding estimates from the estimation method should be obtained. The deviation between a correct value of the variable and its corresponding estimates should be normalized with an appropriate reference for proper comparison. The bias and uncertainty of the estimation method can be calculated based on the normalized values of deviation in the sample to assess the accuracy of the estimation method.

E.1 Choice of a Sample

To fairly assess the accuracy of an estimation method, the bias and uncertainty should be calculated from a sample that represents the population of the variables in reality. If the sample is not appropriately chosen to represent the population, the bias and the uncertainty calculated may not fairly justify the accuracy of an estimation method. For instance, if the accuracy of an estimation method of building electricity consumption based on ambient temperature is assessed by using building data from hot climate zones only, the bias and uncertainty calculated from the assessment will not represent its performance for buildings in cold climate zones. Hence the sample of variables must be chosen to represent the whole population of variables so that the bias and uncertainty calculated will fairly assess the accuracy of an estimation method.

E.2 Quantification of the Deviation between Estimates and Correct Values of a Variable

To calculate the meaningful bias and uncertainty of an estimation method, the deviation used to calculate the bias and uncertainty should be normalized appropriately to fit the application of the estimation method. For instance, if the estimated variable is the energy savings after retrofit, the deviation will be the difference between the estimated and correct values of energy savings normalized by their building energy cost before retrofit. The energy cost before retrofit is chosen to normalize the difference because building owners justify the importance of energy savings relative to their energy cost before retrofit. If the deviations are not normalized, it will be difficult to analyze the bias and uncertainty if the sample contains a wide range of energy savings values. Hence the deviation between the estimates and correct values must be normalized appropriately in order to calculate meaningful values of bias and uncertainty.

E.3 Calculation of Bias

Upon defining the deviation and calculating the deviations for each data point in the sample, the bias of the estimation method can be calculated by equation (E-1).

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}\Delta x_i \tag{E-1}$$

Equation (E-1) shows that bias is calculated by averaging all deviations between estimates and their correct values in a sample. A positive bias means overestimation and a negative bias means underestimation. If an estimation method is unbiased, the magnitude of its bias should be negligible.

Although no normalization of variables is presented in equation (E-1), the evaluation of the Tool calculates the bias with a normalized variable x, and the normalization procedures are discussed in Sections 5 and 6.

E.4 Calculation of Uncertainty

Uncertainty of a variable can be quantified by multiple methods. It can be quantified by different statistical indicators of the spread of population such as confidence interval for experimental data (ASHRAE 2010) and coefficient of variation for M&V tools (ASHRAE 2014). Because the Tool will only be used once per building and each building owner will usually be responsible for one building only, the uncertainty of its estimates should show the possible range of values that an estimate may appear for one future observation. This range equals to the prediction interval of an estimate and can be calculated by equation (E-2) (Montgomery 2004).

$$\text{Prediction interval} = t_{N-1,1-\frac{\alpha}{2}}\sigma\sqrt{(1+\frac{1}{N})} \tag{E-2}$$

Although a similar equation in ASHRAE Guideline 14 (ASHRAE 2014) considers N as the number of months in the data point, the evaluation in this report examines the accuracy of the tool to estimate variables calculated annually, such as annual energy savings and annual energy end uses, and N in equation (E-2) refers to the number of building simulation cases involved in the evaluation rather than the number of months in the data points.

The classification error α in equation (E-2) is usually set to be 0.05 to calculate the uncertainty with a confidence level of 95% according to ASHRAE Guideline 14. The sample standard deviation σ in equation (E-2) is calculated by equation (E-3).

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\Delta x_i^2}{N-1}} \tag{E-3}$$

Appendix F. Bias and Uncertainty Plots To Evaluate Accuracy of Estimates of Annual Energy Uses, Annual Energy Costs, Energy End Use Ratios, and Cost Ratios

This appendix lists the plots of uncertainty against bias to evaluate the estimation accuracy of the Tool for annual energy uses, annual energy costs, energy end uses, and cost ratios for reference.

F.1 Annual Energy Use Plots

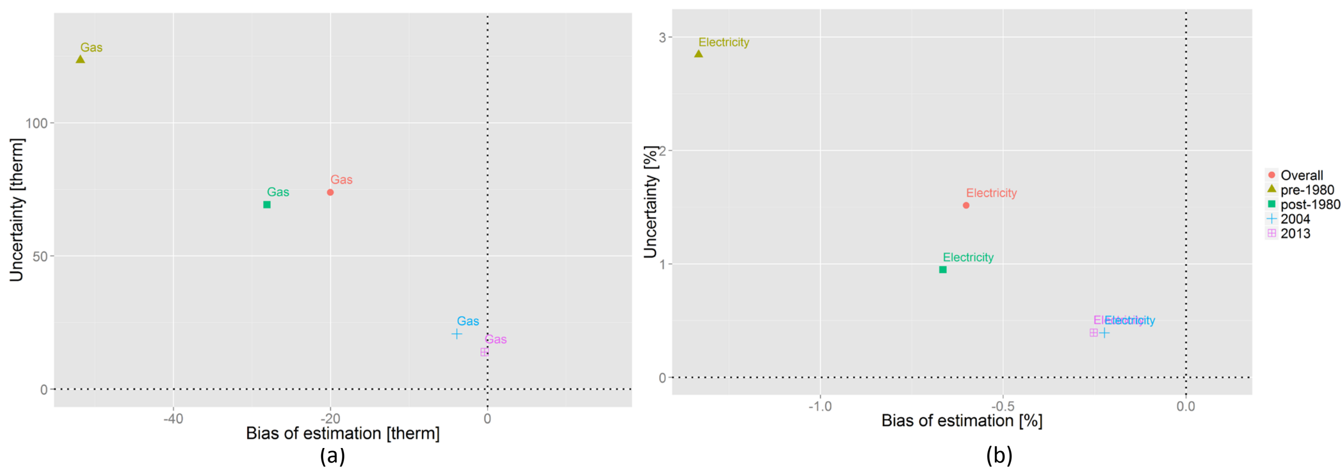


Figure F-1. Bias and uncertainty of annual energy use estimates for small office buildings under different vintages in terms of (a) gas and (b) electricity

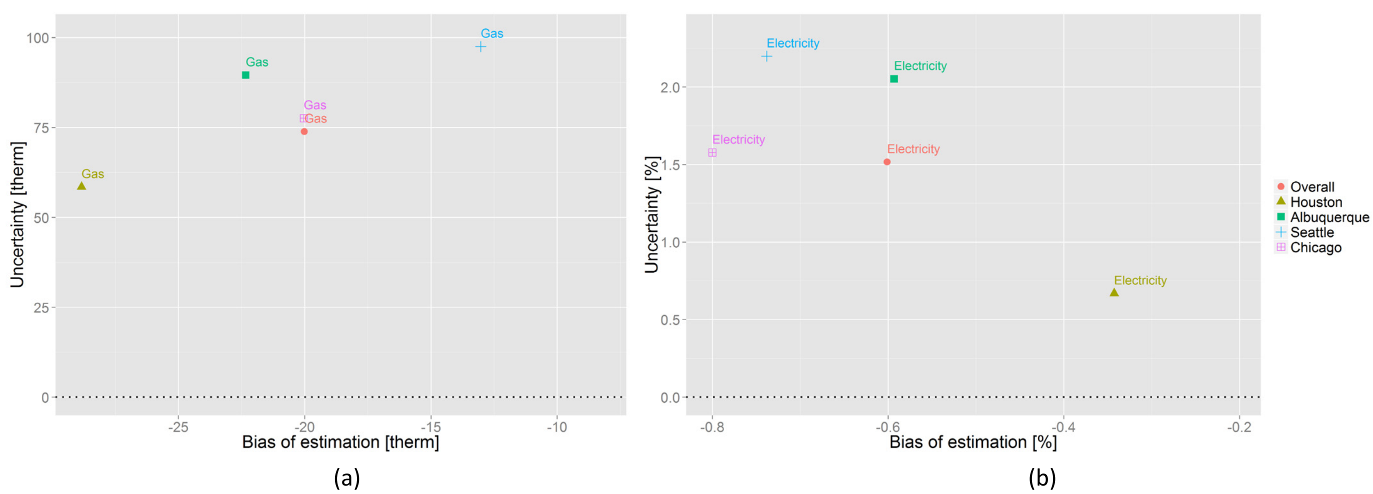
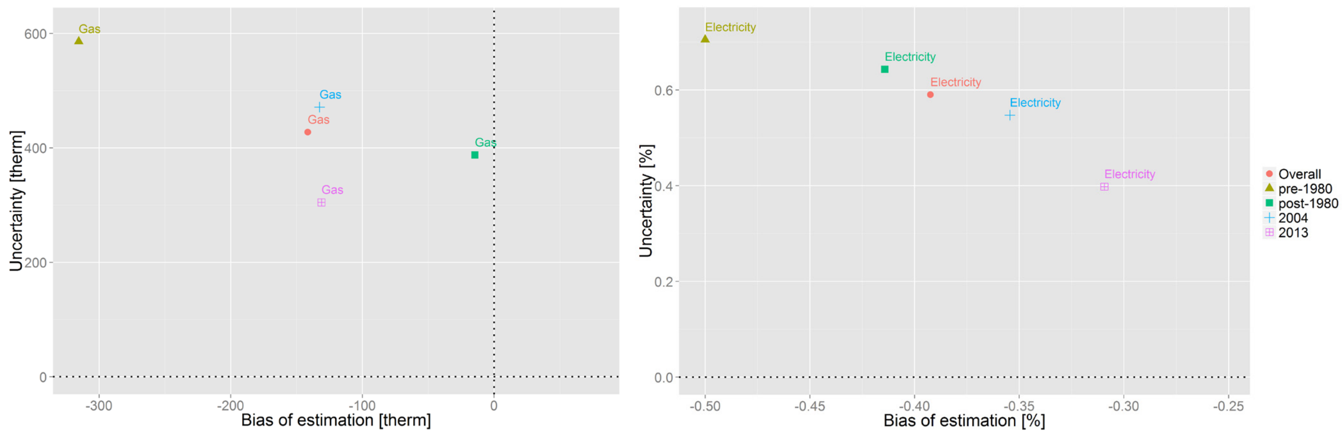
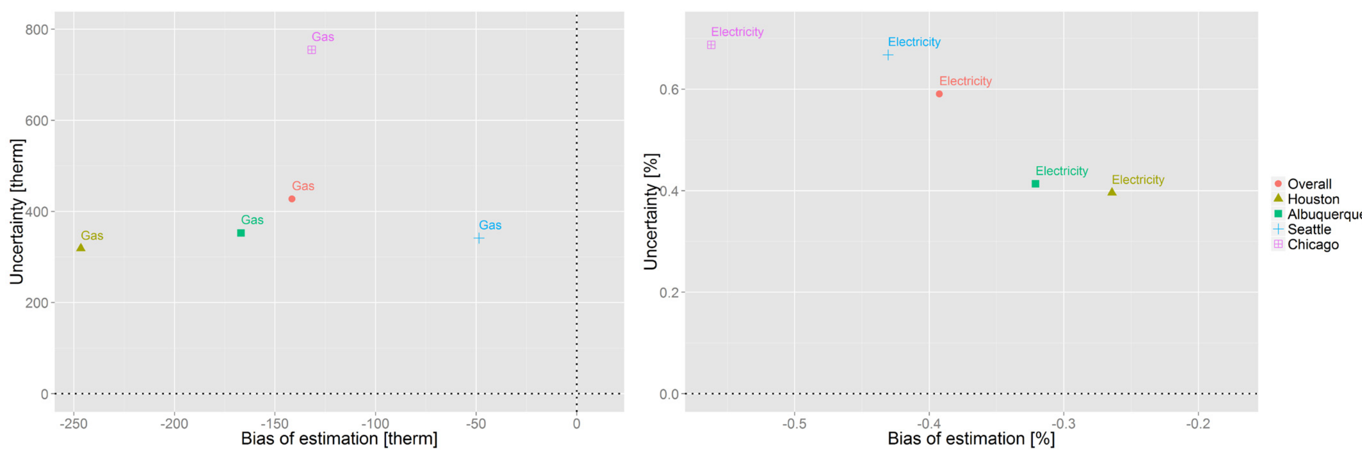


Figure F-2. Bias and uncertainty of annual energy use estimates for small office buildings under different climate zones in terms of (a) gas and (b) electricity



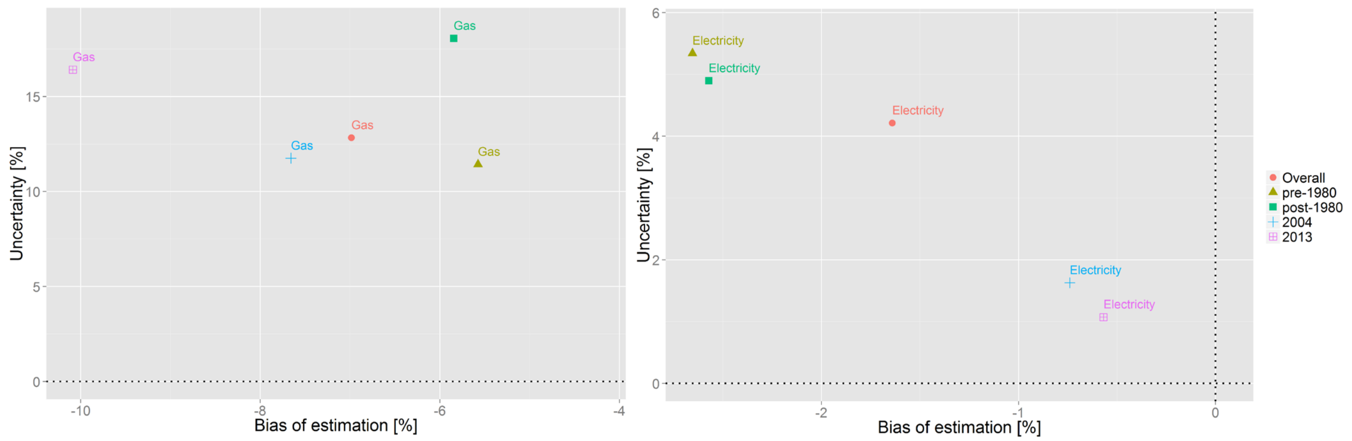
(a) (b)

Figure F-3. Bias and uncertainty of annual energy use estimates for medium office buildings under different vintages in terms of (a) gas and (b) electricity



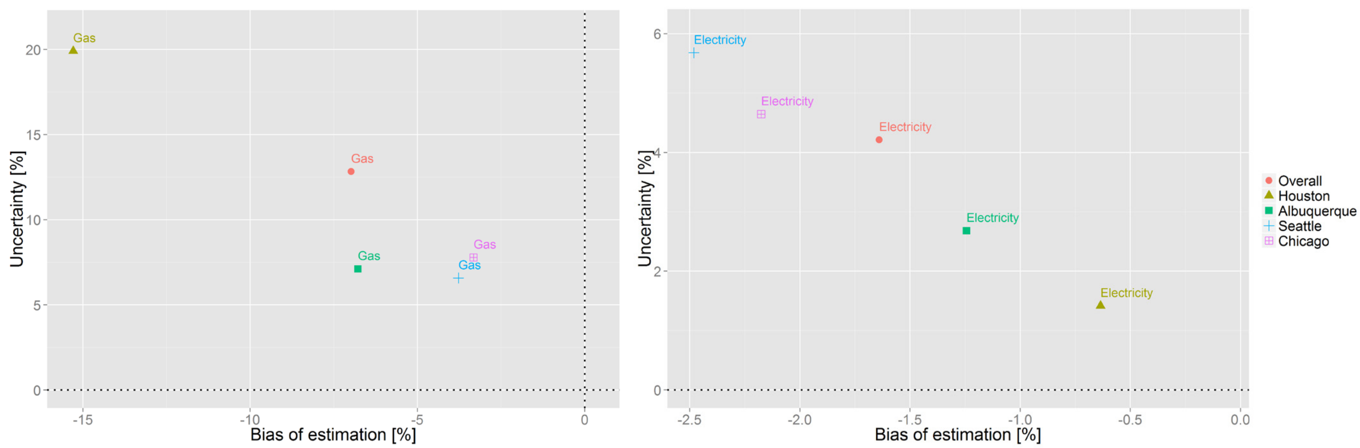
(a) (b)

Figure F-4. Bias and uncertainty of annual energy use estimates for medium office buildings under different climate zones in terms of (a) gas and (b) electricity



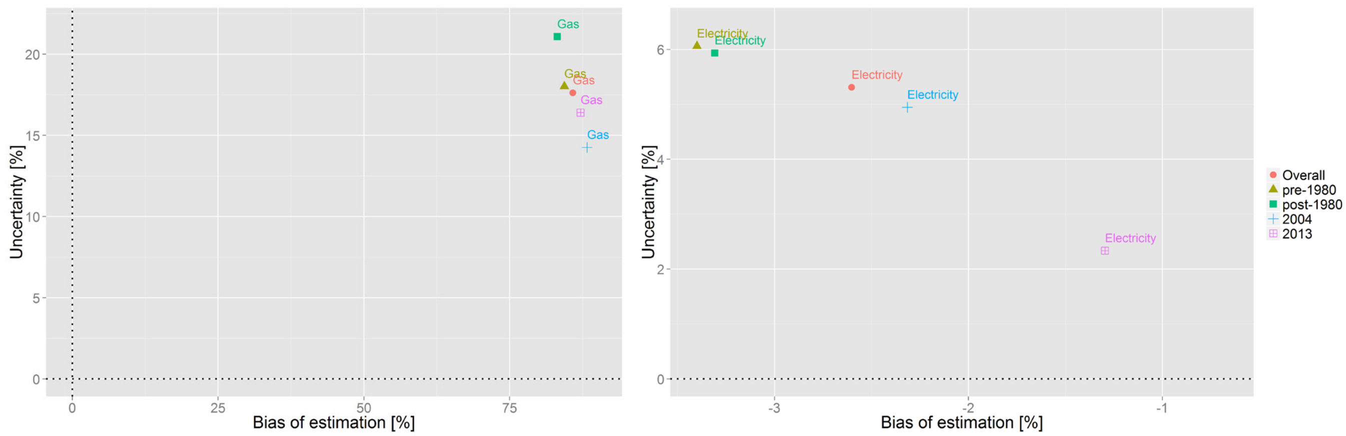
(a) (b)

Figure F-5. Bias and uncertainty of annual energy use estimates for standalone retail buildings under different vintages in terms of (a) gas and (b) electricity



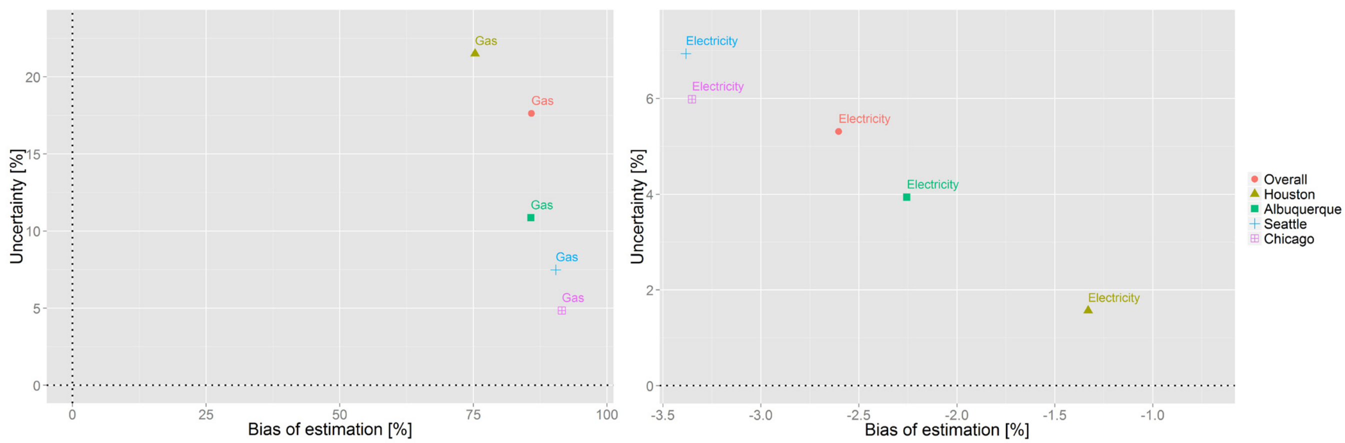
(a) (b)

Figure F-6. Bias and uncertainty of annual energy use estimates for standalone retail buildings under different climate zones in terms of (a) gas and (b) electricity



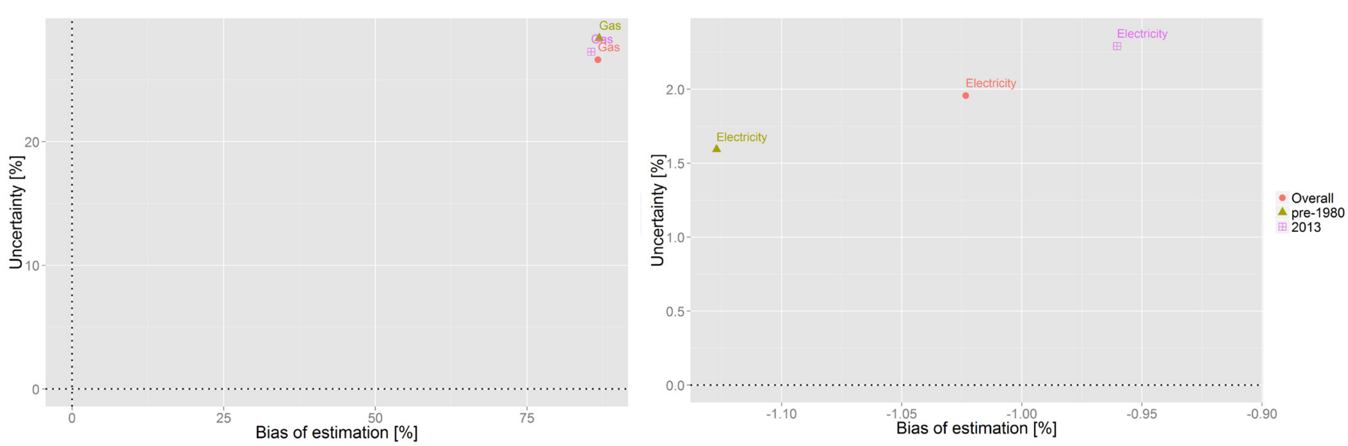
(a) (b)

Figure F-7. Bias and uncertainty of annual energy use estimates for primary school buildings under different vintages in terms of (a) gas and (b) electricity



(a) (b)

Figure F-8. Bias and uncertainty of annual energy use estimates for primary school buildings under different climate zones in terms of (a) gas and (b) electricity



(a) (b)

Figure F-9. Bias and uncertainty of annual energy use estimates for main-street buildings under different vintages in terms of (a) gas and (b) electricity

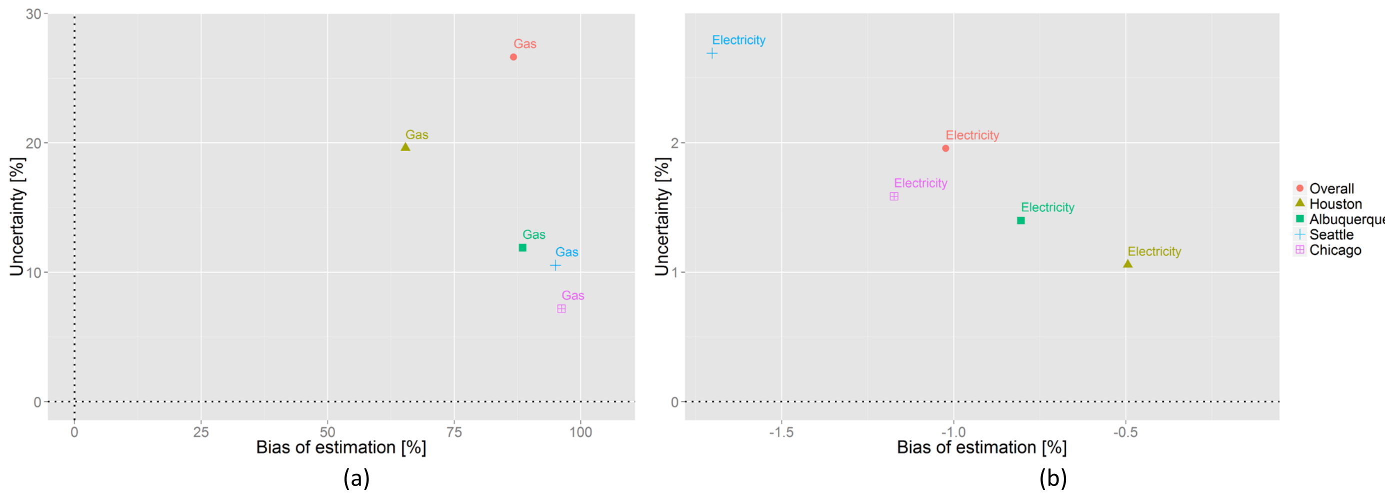


Figure F-10. Bias and uncertainty of annual energy use estimates for main-street buildings under different climate zones in terms of (a) gas and (b) electricity

F.2 Annual Energy Cost Plots

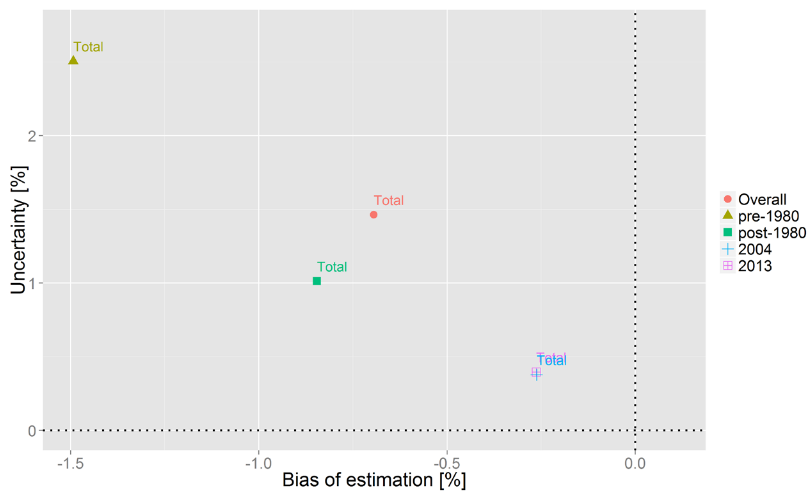


Figure F-11. Bias and uncertainty of annual energy cost estimates for small office buildings under different vintages

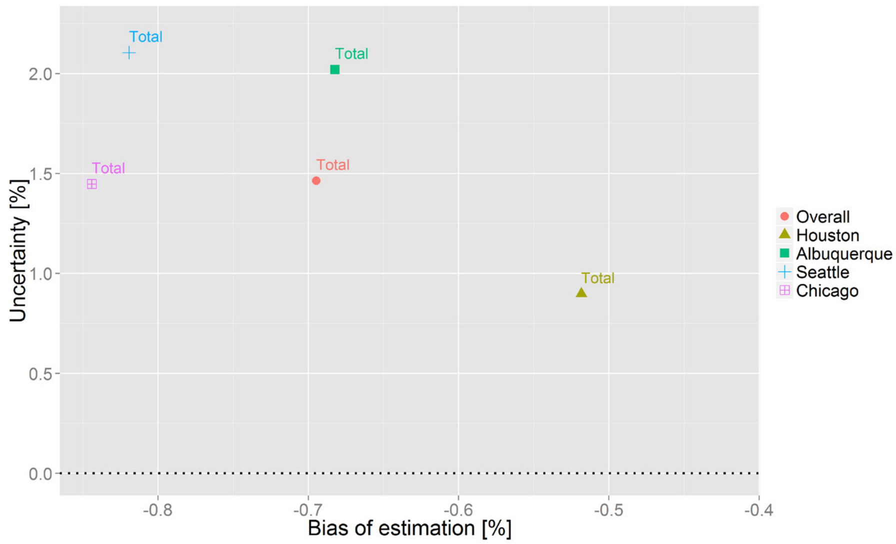


Figure F-12. Bias and uncertainty of annual energy cost estimates for small office buildings under different climate zones

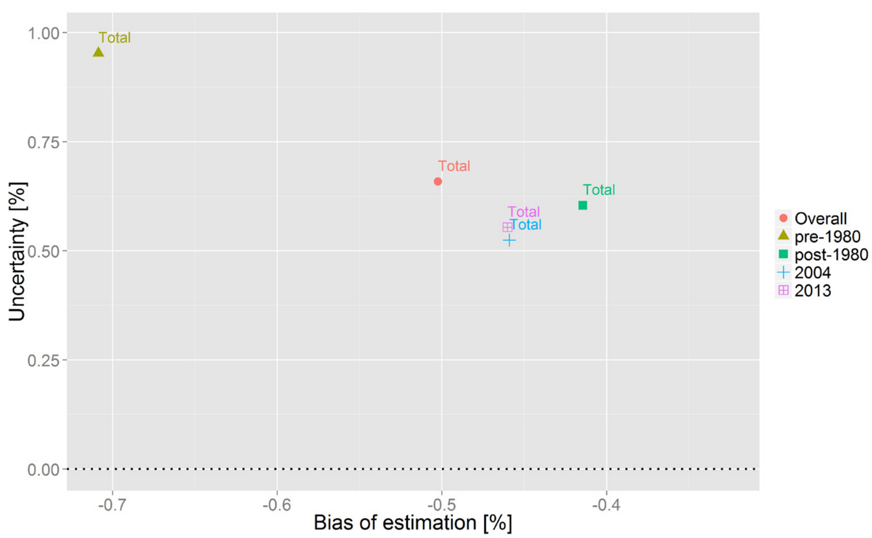


Figure F-13. Bias and uncertainty of annual energy cost estimates for medium office buildings under different vintages

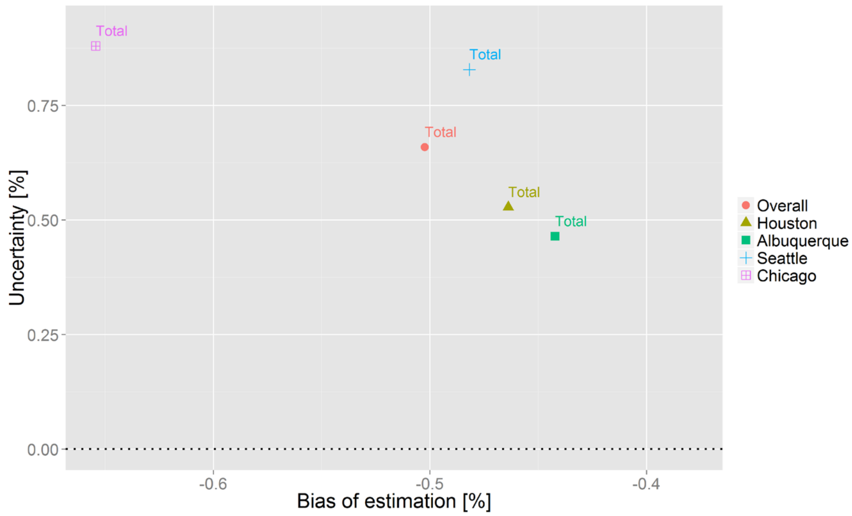


Figure F-14. Bias and uncertainty of annual energy cost estimates for medium office buildings under different climate zones

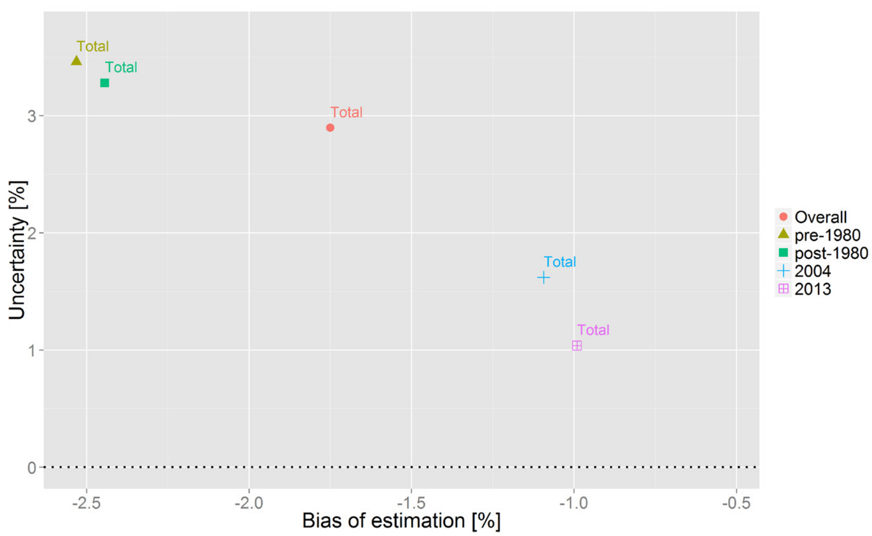


Figure F-15. Bias and uncertainty of annual energy cost estimates for standalone retail buildings under different vintages

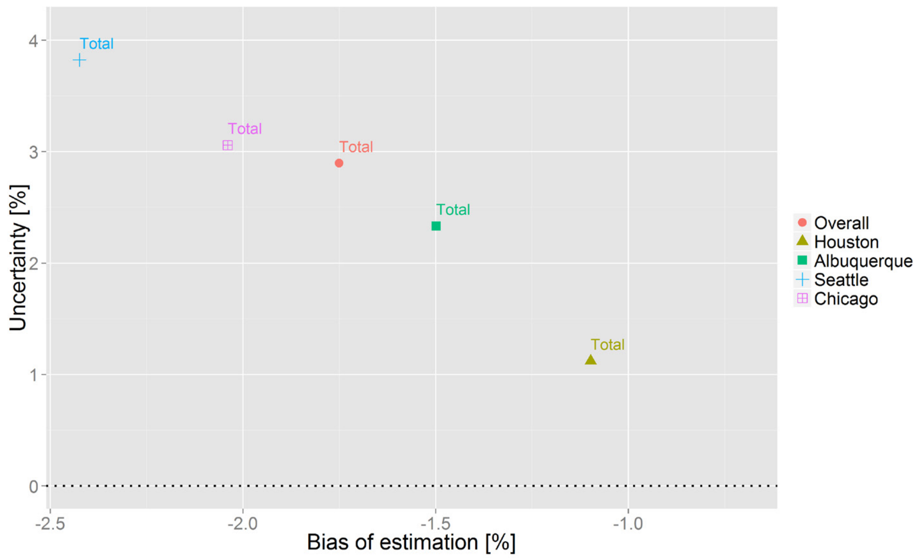


Figure F-16. Bias and uncertainty of annual energy cost estimates for standalone retail buildings under different climate zones

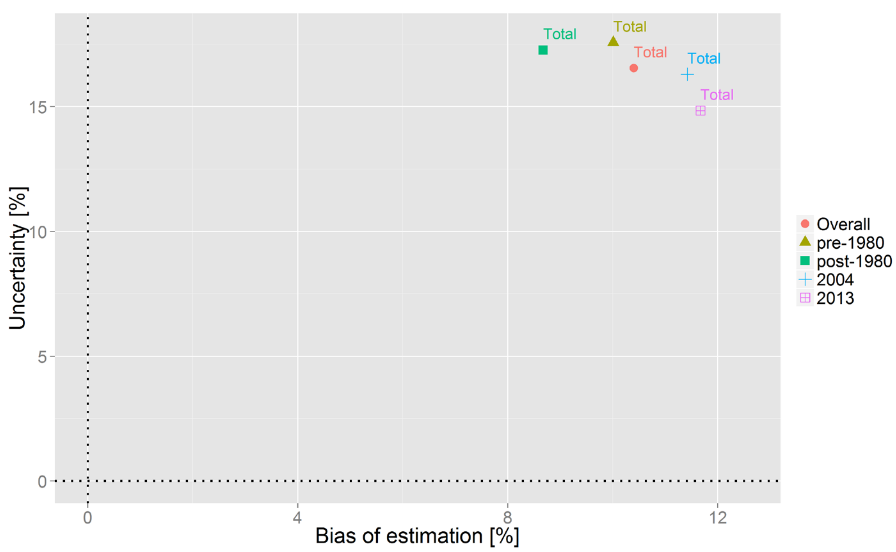


Figure F-17. Bias and uncertainty of annual energy cost estimates for primary school buildings under different vintages

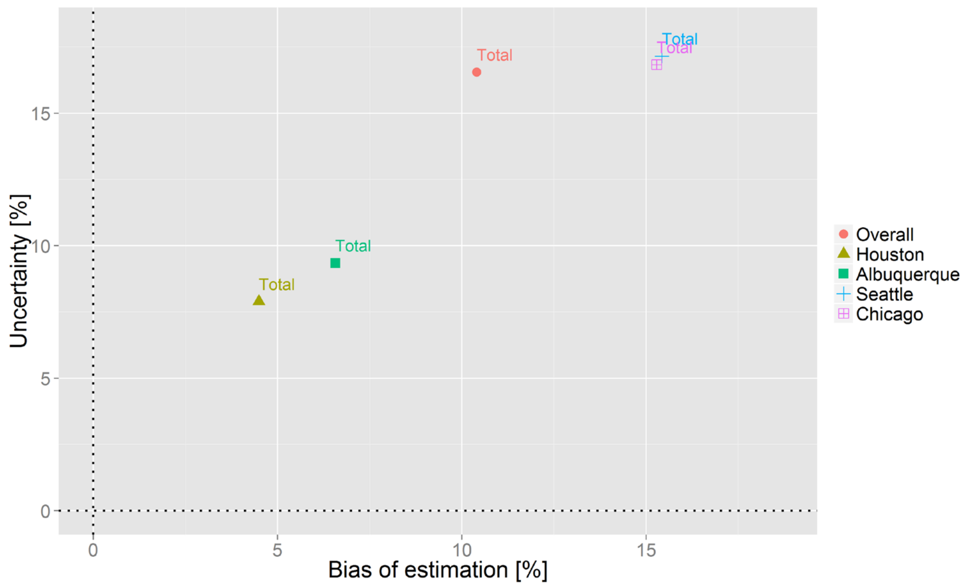


Figure F-18. Bias and uncertainty of annual energy cost estimates for primary school buildings under different climate zones

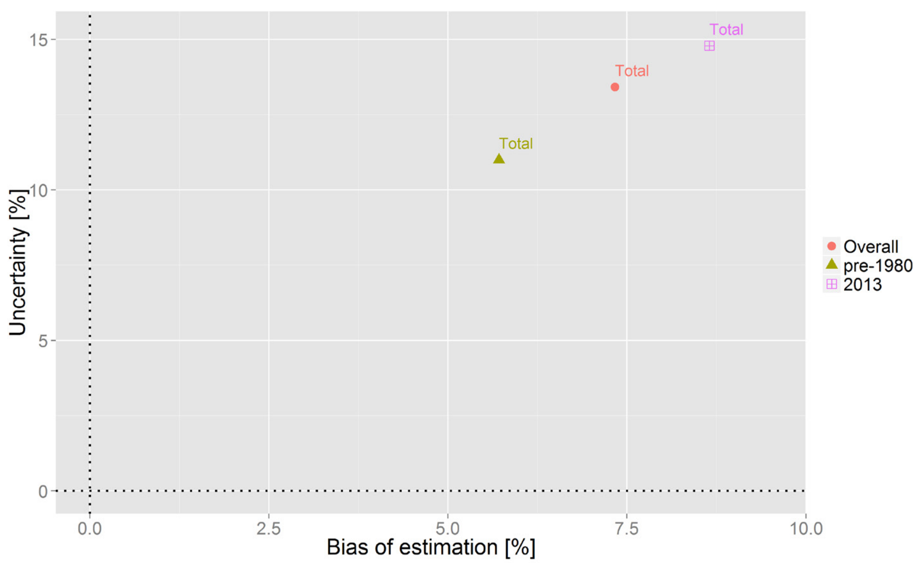


Figure F-19. Bias and uncertainty of annual energy cost estimates for main-street buildings under different vintages

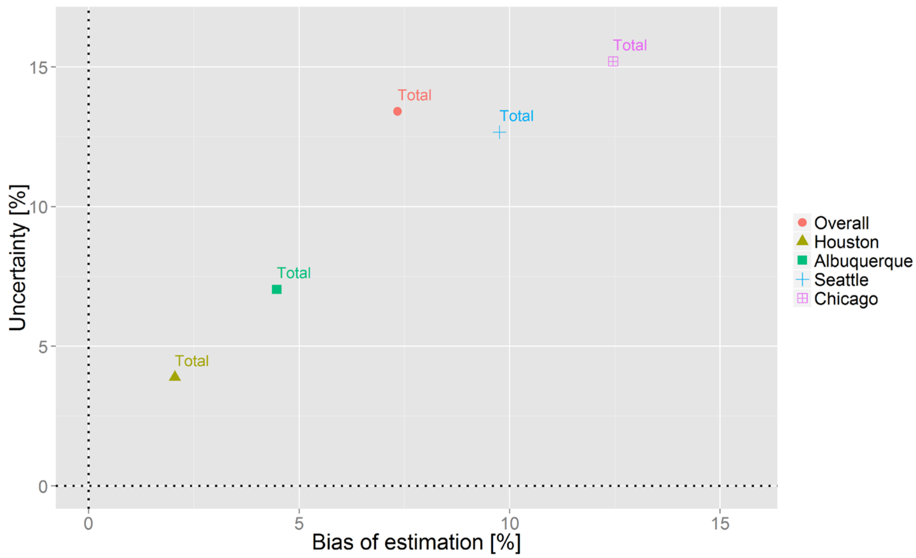


Figure F-20. Bias and uncertainty of annual energy cost estimates for main-street buildings under different climate zones

F.3 Energy End Use Ratio Plots

Figure F-21. Bias and uncertainty of energy end use ratio estimates for small office buildings under different vintages

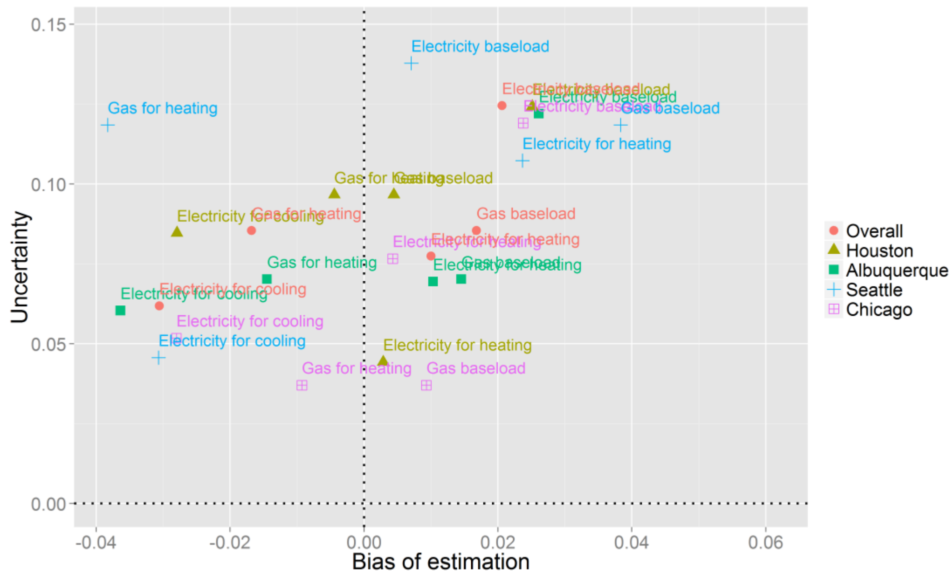


Figure F-22. Bias and uncertainty of energy end use ratio estimates for small office buildings under different climate zones

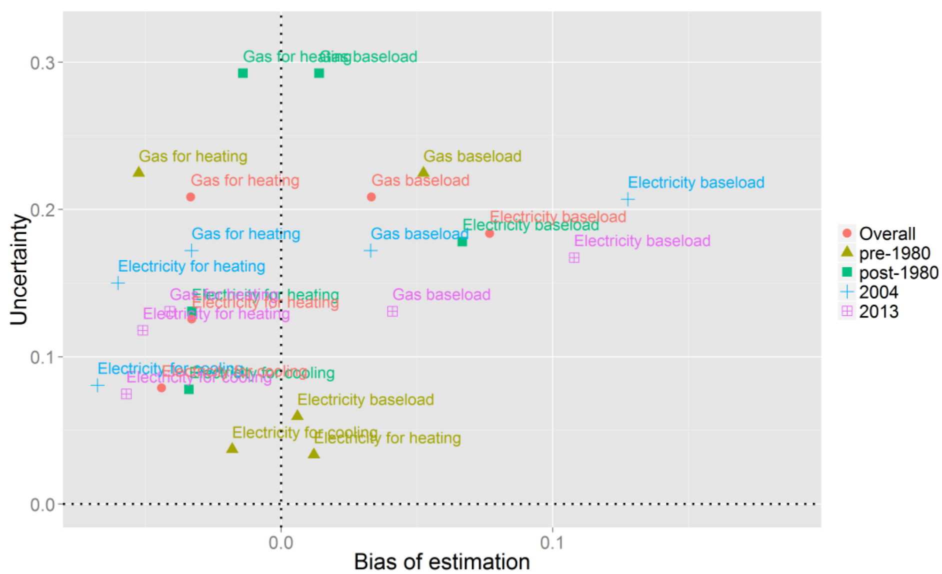


Figure F-23. Bias and uncertainty of energy end use ratio estimates for medium office buildings under different vintages

Figure F-24. Bias and uncertainty of energy end use ratio estimates for medium office buildings under different climate zones

Figure F-25. Bias and uncertainty of energy end use ratio estimates for standalone retail buildings under different vintages

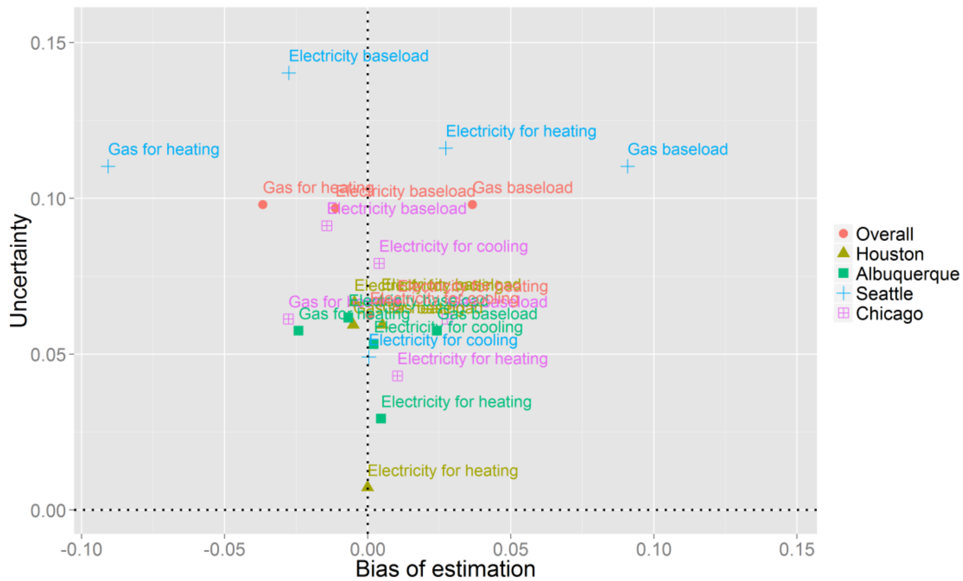


Figure F-26. Bias and uncertainty of energy end use ratio estimates for standalone retail buildings under different climate zones

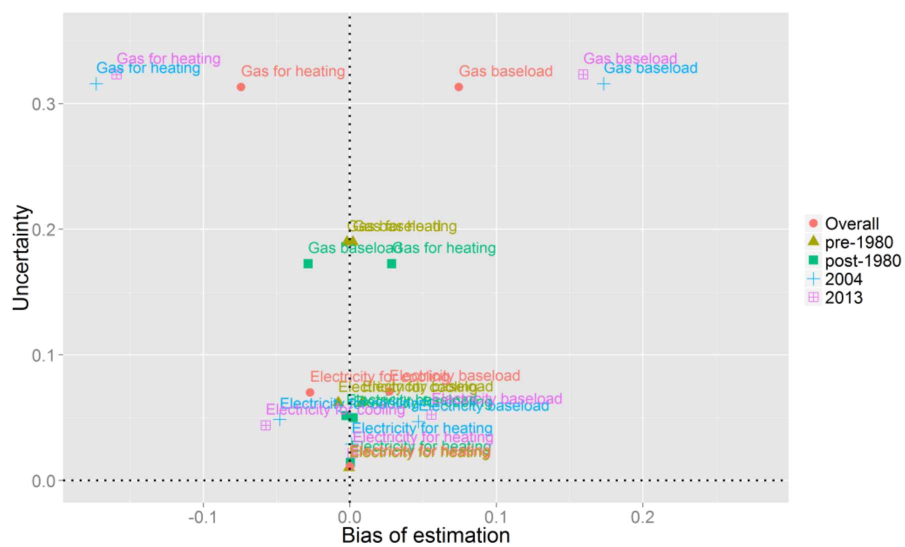


Figure F-27. Bias and uncertainty of energy end use ratio estimates for primary school buildings under different vintages

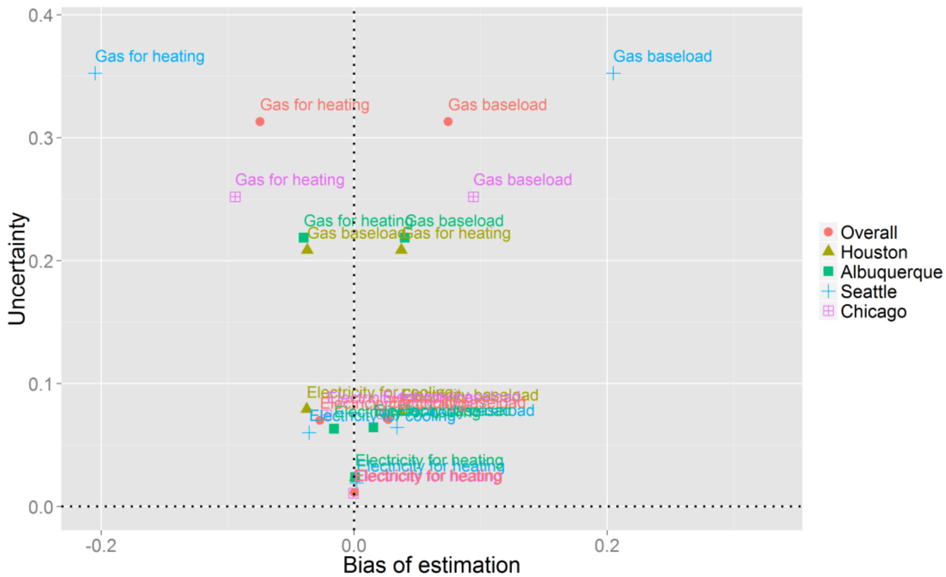


Figure F-28. Bias and uncertainty of energy end use ratio estimates for primary school buildings under different climate zones

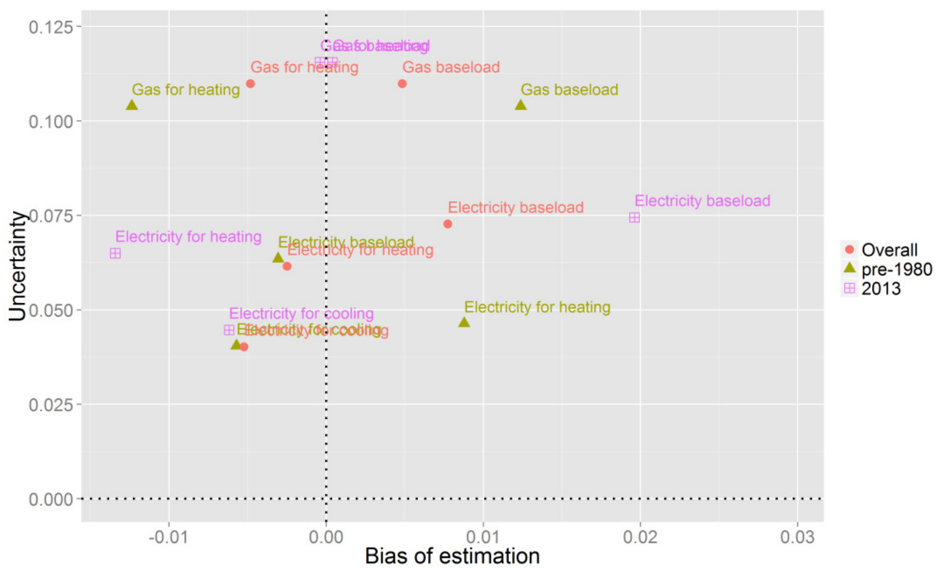


Figure F-29. Bias and uncertainty of energy end use ratio estimates for main-street buildings under different vintages

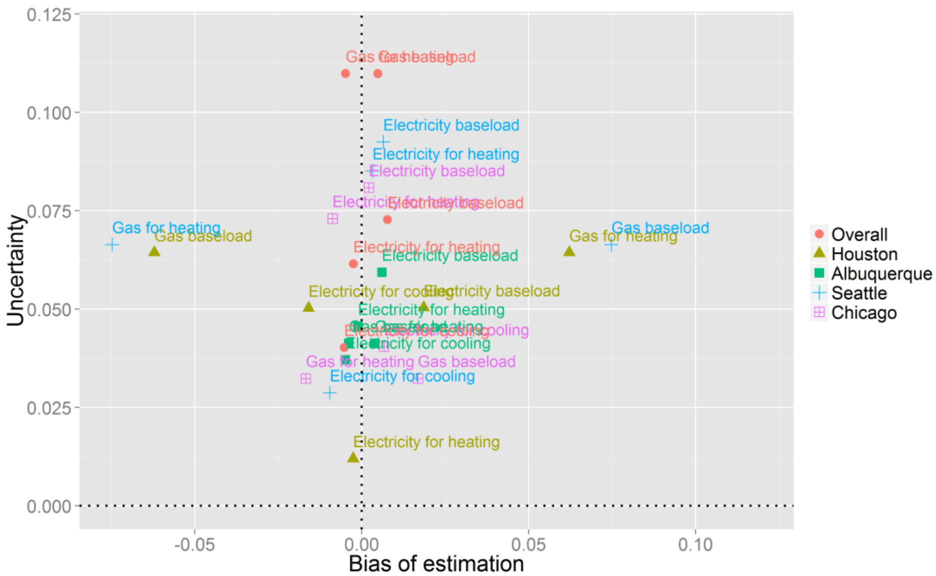


Figure F-30. Bias and uncertainty of energy end use ratio estimates for main-street buildings under different climate zones

F.4 Cost Ratio Plots

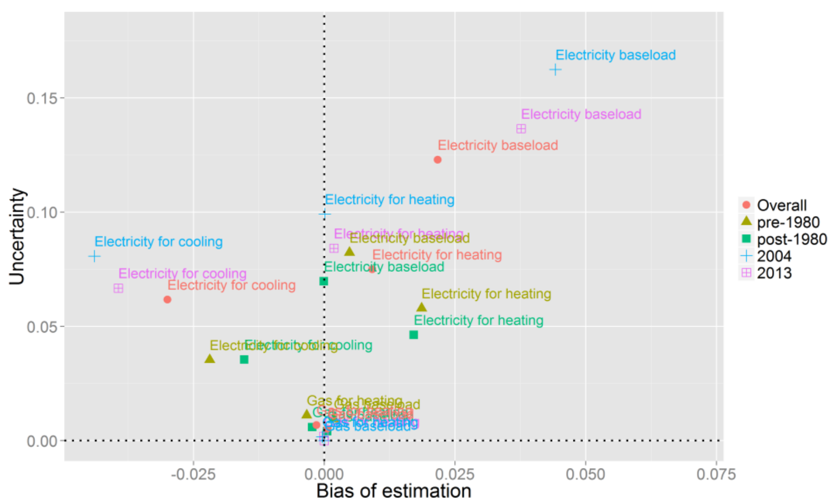


Figure F-31. Bias and uncertainty of cost ratio estimates for small office buildings under different vintages

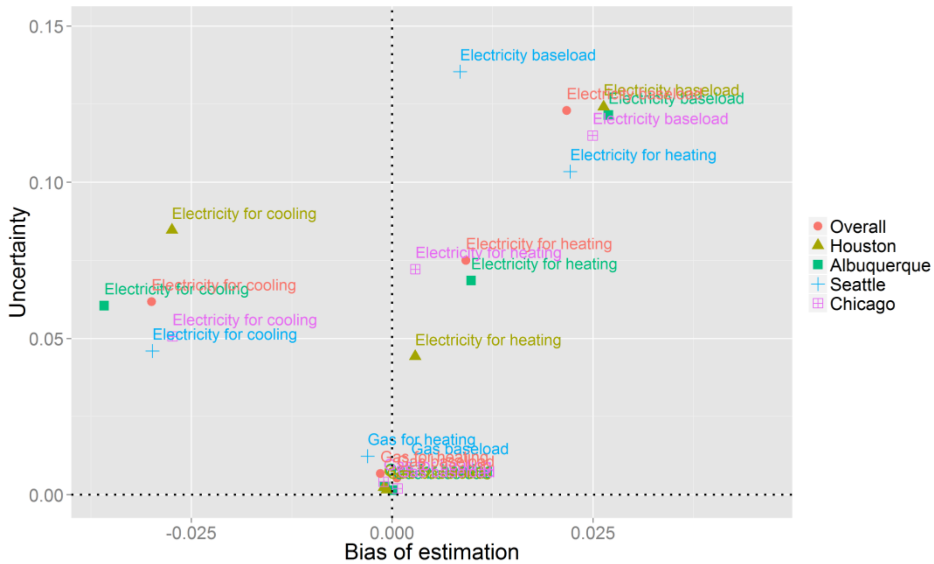


Figure F-32. Bias and uncertainty of cost ratio estimates for small office buildings under different climate zones

Figure F-33. Bias and uncertainty of cost ratio estimates for medium office buildings under different vintages

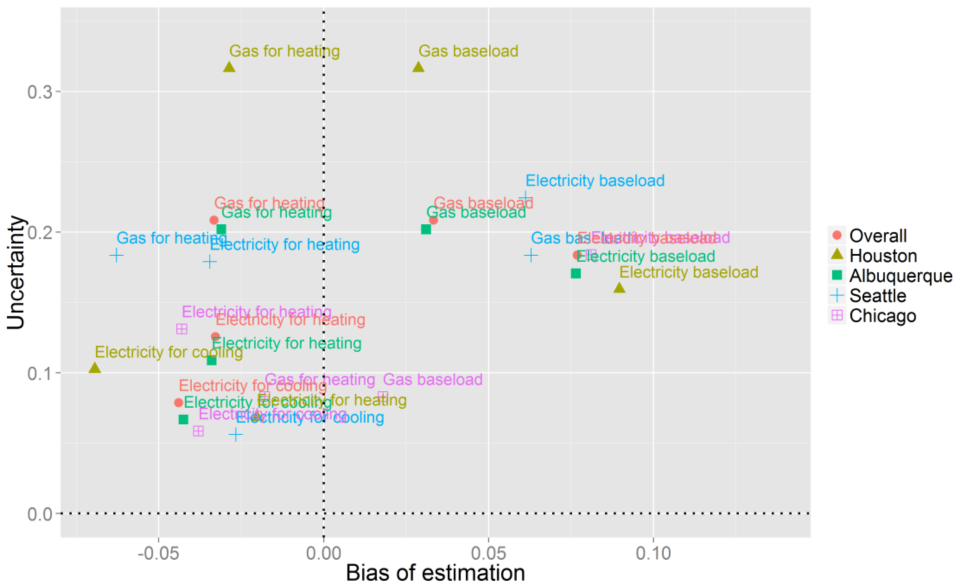


Figure F-34. Bias and uncertainty of cost ratio estimates for medium office buildings under different climate zones

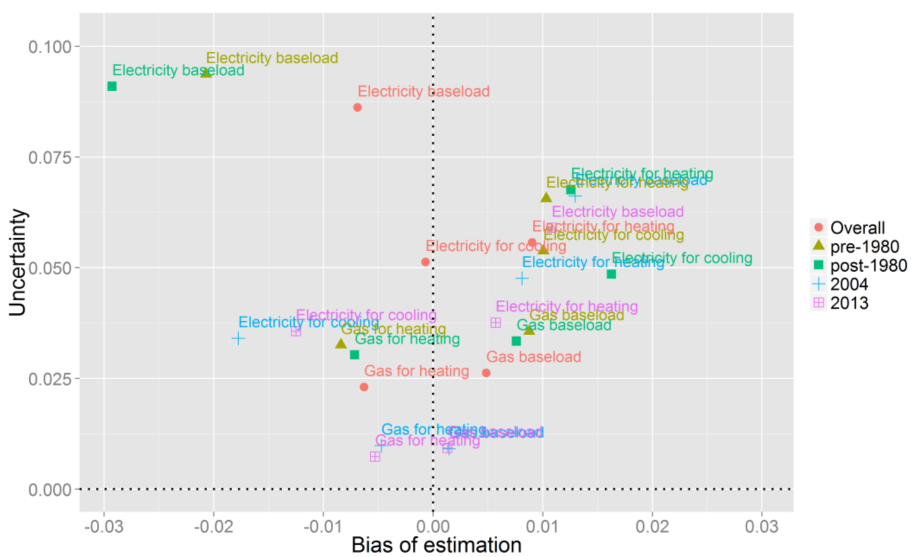


Figure F-35. Bias and uncertainty of cost ratio estimates for standalone retail buildings under different vintages

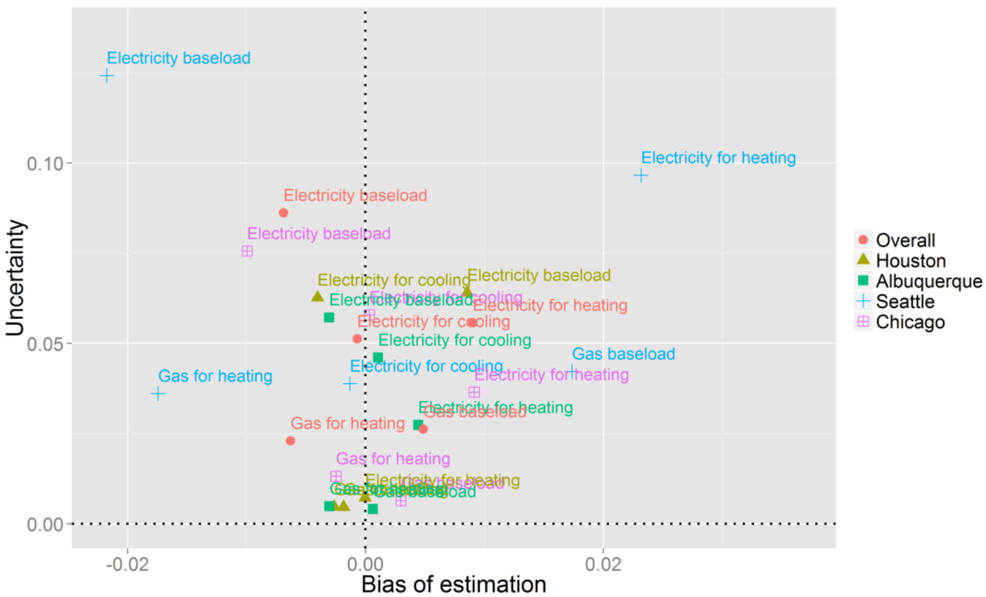


Figure F-36. Bias and uncertainty of cost ratio estimates for standalone retail buildings under different climate zones

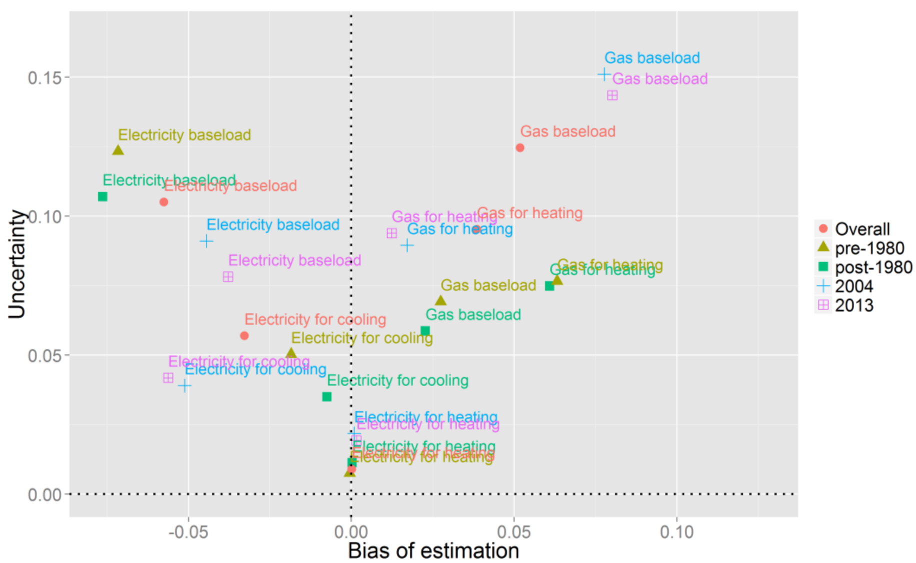


Figure F-37. Bias and uncertainty of cost ratio estimates for primary school buildings under different vintages

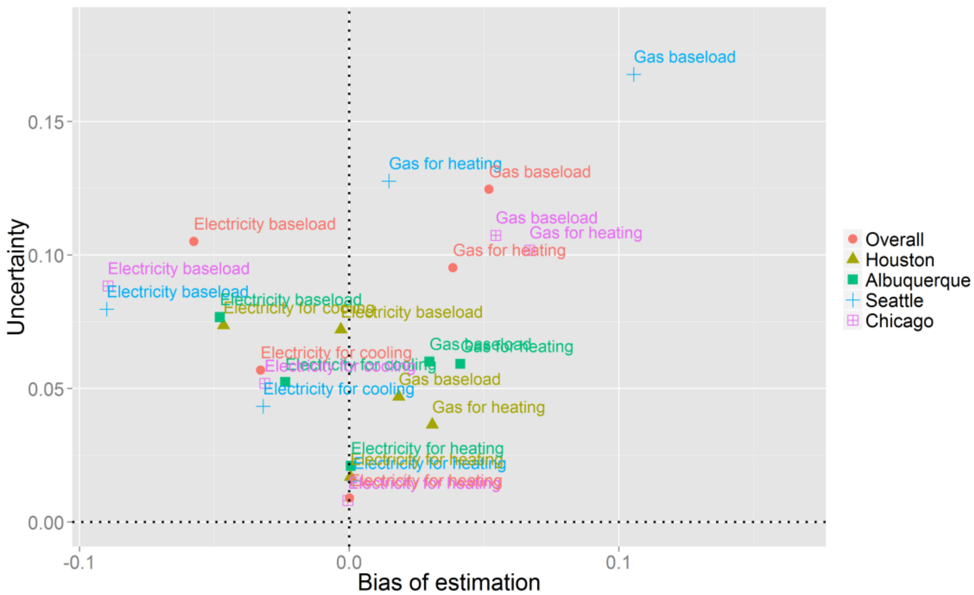


Figure F-38. Bias and uncertainty of cost ratio estimates for primary school buildings under different climate zones

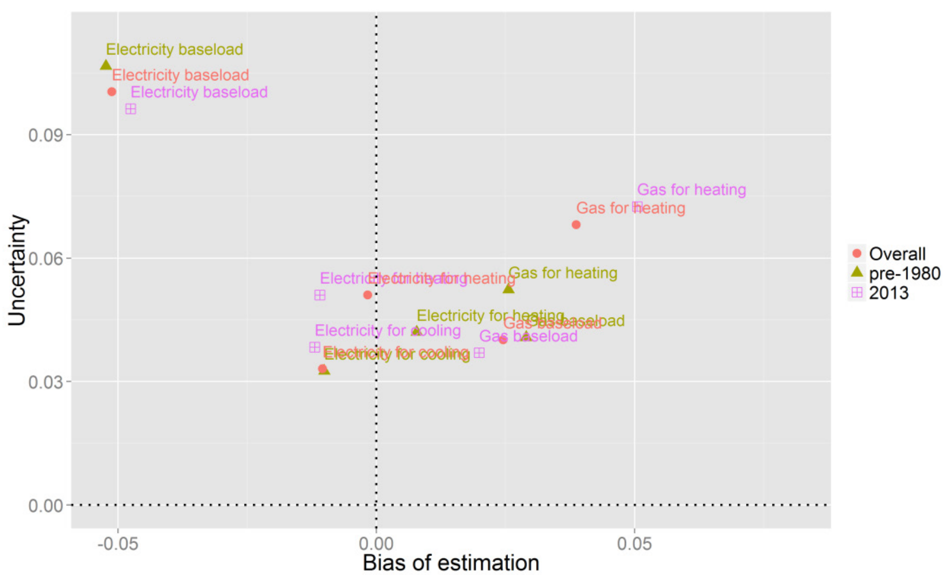


Figure F-39. Bias and uncertainty of cost ratio estimates for main-street buildings under different vintages

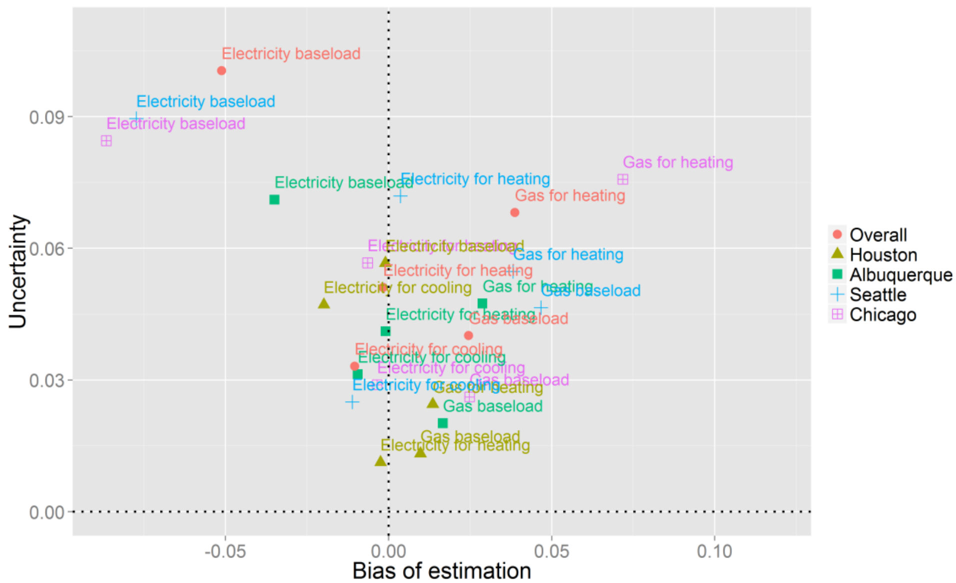


Figure F-40. Bias and uncertainty of cost ratio estimates for main-street buildings under different climate zones

Appendix G. Evaluation Results of the Tool Message Flags

This appendix categorizes the evaluation results of the accuracy of message flags for different types of buildings according to the performance indicators discussed in Section 7.

Table G-1. Evaluation Results of Message Flags for Small Office Buildings

Message Flags	True Diagnosis Rate [%]	False Alarm Rate [%]	False Diagnosis Rate [%]	Weak Diagnosis Rate [%]	True Negative Rate [%]	Accuracy To Classify Buildings Based on Electricity Baseload [%]	Accuracy To Identify if Internal Electricity Load Dominates the Cause of High Electricity Baseload [%]
A, B, C, O, and P	N/A	N/A	N/A	N/A	N/A	91.99	N/A
O and P	N/A	N/A	N/A	N/A	N/A	N/A	65.10
D	62.86	45.09	N/A	N/A	54.91	N/A	N/A
E	0.00	0.92	N/A	N/A	99.08	N/A	N/A
M	13.79	7.60	N/A	10.48	86.22	N/A	N/A
H	28.57	10.00	0	N/A	82.22	N/A	N/A
I	16.28	7.99	N/A	N/A	92.01	N/A	N/A
K	25.81	0.17	N/A	N/A	99.83	N/A	N/A
L	N/A	5.93	N/A	N/A	94.07	N/A	N/A
Q	14.29	7.78	10.00	N/A	82.22	N/A	N/A
N	12.50	0.00	N/A	N/A	100.00	N/A	N/A

Table G-2. Evaluation Results of Message Flags for Medium Office Buildings

Message Flags	True Diagnosis Rate [%]	False Alarm Rate [%]	False Diagnosis Rate [%]	Weak Diagnosis Rate [%]	True Negative Rate [%]	Accuracy To Classify Buildings Based on Electricity Baseload [%]	Accuracy To Identify if Internal Electricity Load Dominates the Cause of High Electricity Baseload [%]
A, B, C, O, and P	N/A	N/A	N/A	N/A	N/A	84.94	N/A
O and P	N/A	N/A	N/A	N/A	N/A	N/A	72.08
D	60.75	39.07	N/A	N/A	60.93	N/A	N/A
E	0.00	3.37	N/A	N/A	96.63	N/A	N/A
M	13.11	13.68	N/A	11.21	80.82	N/A	N/A
H	34.94	14.67	0.00	N/A	79.34	N/A	N/A
I	17.24	13.97	N/A	N/A	86.03	N/A	N/A
K	0.00	0.00	N/A	N/A	100.00	N/A	N/A
L	50.00	7.12	N/A	N/A	92.88	N/A	N/A
Q	8.70	5.98	0.00	N/A	79.34	N/A	N/A
N	6.25	0.00	N/A	N/A	100.00	N/A	N/A

Table G-3. Evaluation Results of Message Flags for Standalone Retail Buildings

Message Flags	True Diagnosis Rate [%]	False Alarm Rate [%]	False Diagnosis Rate [%]	Weak Diagnosis Rate [%]	True Negative Rate [%]	Accuracy To Classify Buildings Based on Electricity Baseload [%]	Accuracy To Identify if Internal Electricity Load Dominates the Cause of High Electricity Baseload [%]
A, B, C, O, and P	N/A	N/A	N/A	N/A	N/A	85.74	N/A
O and P	N/A	N/A	N/A	N/A	N/A	N/A	32.95
D	92.73	76.26	N/A	N/A	23.74	N/A	N/A
E	13.33	2.76	N/A	N/A	97.24	N/A	N/A
M	0.74	3.69	N/A	0.91	95.49	N/A	N/A
H	13.70	3.58	0.00	N/A	73.77	N/A	N/A
I	31.18	18.64	N/A	N/A	81.36	N/A	N/A
K	73.64	12.65	N/A	N/A	87.35	N/A	N/A
L	N/A	2.72	N/A	N/A	97.28	N/A	N/A
Q	33.33	22.64	27.40	N/A	73.77	N/A	N/A
N	18.75	0.00	N/A	N/A	100.00	N/A	N/A

Table G-4. Evaluation Results of Message Flags for Primary Schools

Message Flags	True Diagnosis Rate [%]	False Alarm Rate [%]	False Diagnosis Rate [%]	Weak Diagnosis Rate [%]	True Negative Rate [%]	Accuracy To Classify Buildings Based on Electricity Baseload [%]	Accuracy To Identify if Internal Electricity Load Dominates the Cause of High Electricity Baseload [%]
A, B, C, O, and P	N/A	N/A	N/A	N/A	N/A	78.92	N/A
O and P	N/A	N/A	N/A	N/A	N/A	N/A	0.00
D	93.91	84.29	N/A	N/A	15.71	N/A	N/A
E	1.54	0.81	N/A	N/A	99.19	N/A	N/A
M	5.88	6.16	N/A	0.00	93.84	N/A	N/A
H	32.95	17.72	0.00	N/A	63.37	N/A	N/A
I	54.22	17.52	N/A	N/A	82.48	N/A	N/A
K	100.00	100.00	N/A	N/A	0.00	N/A	N/A
L	9.09	7.36	N/A	N/A	92.64	N/A	N/A
Q	23.08	18.91	32.95	N/A	63.37	N/A	N/A
N	37.50	0.00	N/A	N/A	100.00	N/A	N/A

Table G-5. Evaluation Results of Message Flags for Main-Street Buildings

Message Flags	True Diagnosis Rate [%]	False Alarm Rate [%]	False Diagnosis Rate [%]	Weak Diagnosis Rate [%]	True Negative Rate [%]	Accuracy To Classify Buildings Based on Electricity Baseload [%]	Accuracy To Identify if Internal Electricity Load Dominates the Cause of High Electricity Baseload [%]
A, B, C, O, and P	N/A	N/A	N/A	N/A	N/A	86.86	N/A
O and P	N/A	N/A	N/A	N/A	N/A	N/A	52.38
D	100.00	82.85	N/A	N/A	17.15	N/A	N/A
E	0.00	0.00	N/A	N/A	100.00	N/A	N/A
M	7.69	13.29	N/A	0.00	86.71	N/A	N/A
H	4.55	0.70	0.00	N/A	74.65	N/A	N/A
I	8.57	3.61	N/A	N/A	96.39	N/A	N/A
K	100.00	70.24	N/A	N/A	29.76	N/A	N/A
L	N/A	0.64	N/A	N/A	99.36	N/A	N/A
Q	50.00	24.65	22.73	N/A	74.65	N/A	N/A
N	0.00	0.00	N/A	N/A	100.00	N/A	N/A