Energy and Buildings 85 (2014) 165–179

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

A statistical approach for post-processing residential building energy simulation output

S. Valovcin [a], A.S. Hering [a,*], B. Polly [b], M. Heaney [b]

[a] *Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO, United States*
[b] *Buildings and Thermal Systems Center, National Renewable Energy Laboratory, Golden, CO, United States*

ARTICLE INFO

Article history:
Received 21 February 2014
Received in revised form 2 July 2014
Accepted 27 July 2014
Available online 6 September 2014

Keywords:
Clustering
Cross-validation
Model calibration
Principal component analysis
Residential building energy simulation

ABSTRACT

Residential building energy simulation (RBES) software plays an important role in evaluating the energy consumption and efficiency potential of homes. These physics-based models are commonly used to assess the energy performance of homes and to predict benefits of making energy-saving improvements to homes *a priori*. However, software may produce biased estimates of energy consumption for a variety of reasons, including: errors in the measurement and observation of building characteristics; differences in the assumed versus actual occupant behavior; and errors in the physical models and algorithms used in the software. In order to evaluate and improve the accuracy of RBES software, the National Renewable Energy Laboratory (NREL) has assembled a set of approximately 1,250 U.S. homes for which measured energy consumption and audit-collected household energy characteristics are available. Algorithms have also been developed that automatically translate the data from each home into RBES input files so that model predictions of annual electricity and natural gas consumption can be compared to measured values. To assess and improve upon the accuracy of these predictions, we first cluster the homes using weighted, independent linear combinations of these variables and then build multiple linear regressions within clusters of similar homes to model the difference between measured and predicted energy consumption based on the recorded features of the homes. The statistical post-processing techniques that we develop for RBES models have the following benefits: (1) they can identify variables and algorithms that may be causing inaccuracies in the RBES process and (2) they can be used to adjust and improve the RBES predictions.

1. Introduction

In 2010, the residential sector accounted for 23% of the United States' total energy consumption, which is 4.4% of the world's energy consumption [1]. Efforts to understand and model this consumption are critical to reducing it. Residential Building Energy Simulation (RBES) tools typically model the energy consumption (e.g., electricity and natural gas) of a home at hourly or sub-hourly time steps using a physics-based approach.[1] Both federally and privately funded organizations have developed such tools for use in home energy assessments that help energy efficiency practitioners and homeowners evaluate the economic benefit of investing in home improvements designed to reduce energy consumption.

One particular quantity of interest predicted by an RBES program is the pre-retrofit annual energy consumption of a home. Then, the annual energy consumption of a home under proposed retrofits can be projected, and the estimated energy savings can be compared with the cost of the retrofit. However, physics-based models can provide biased predictions of the annual energy consumption of the home. The top panels of Fig. 1 show a sample of homes for which RBES predictions of annual energy consumption are plotted against the weather-normalized measured annual energy. Natural gas predictions align better with the line of perfect agreement than the electricity predictions do, but the RBES model still tends to overpredict consumption for this sample of homes. Both electricity and natural gas consumption can be highly driven by occupant behavior, but the errors in electricity predictions are thought to be even more variable due to the fact that a large fraction of electric end use is almost entirely driven by occupant preferences and choices (e.g., miscellaneous electric loads, appliances, and lighting). However, in this study, standard occupant

* Corresponding author. Tel.: +1 3033842462.
E-mail addresses: svalovcin@gmail.com (S. Valovcin), ahering@mines.edu (A.S. Hering), ben.polly@nrel.gov (B. Polly), michael.heaney@nrel.gov (M. Heaney).

[1] Some RBES tools may use larger time-steps or seasonal approaches for modeling; so long as the tool is able to predict annual consumption, the statistical techniques applied in this study are relevant.

http://dx.doi.org/10.1016/j.enbuild.2014.07.060
0378-7788/

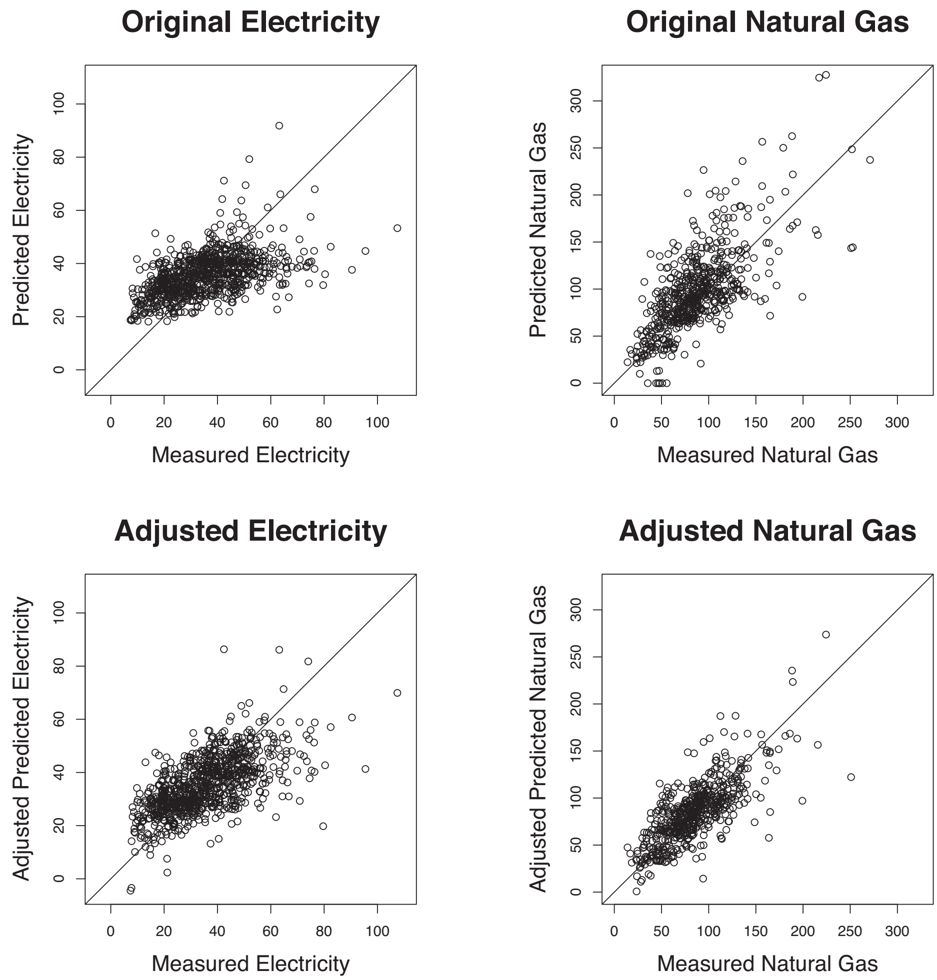

Fig. 1. Top row: Plot of RBES predictions of electricity (left) and natural gas (right) against weather-normalized measured annual consumption of each type of energy in MMBtu. Bottom row: Plot of adjusted RBES predictions versus weather-normalized measured annual electricity (left) and natural gas (right) consumption for the combined clusters with a reference line of perfect agreement.

behavior is used in the software, and the building characteristics data were generally collected for the purpose of an "asset rating," which means specific occupant behavior was not considered. Thus, standard occupant behavior is assumed to provide a fair comparison of the homes' energy efficiency features.

An approach for adjusting predictions to more closely match the actual electric energy consumption in the home would be useful, as shown in the bottom panels of Fig. 1, which will be discussed in more detail in Section 4. Deviations between observed and predicted energy consumption may occur for any of the following reasons: (1) measurement error in data collection; (2) error in utility data normalization; (3) deviations in occupant behavior; (4) coding errors in translation scripts; or (5) deficiencies in the RBES model [2].

This research focuses on the last two types of errors. First, in order for a home to be successfully simulated by an RBES tool, a specific set of simulation inputs describing the home is required. Translation scripts automatically map information in a database for a given home to create a simulation input file, as shown in Fig. 2. Some inputs are computed based on characteristics observed or measured in the home. For example, duct leakage percentages are calculated when the results of a duct pressurization test that is performed during the audit are available. Assumptions, simplifications, and coding errors in the translation scripts can contribute to errors in the RBES predictions. The RBES program may also face certain limitations, like the inability to model homes with multiple hot water heaters or multiple duct systems. In such cases, assumptions and simplifications are necessary to make predictions.

Secondly, there may be errors and inaccuracies in the RBES program itself. Default assumptions made within the program, such as occupant use profiles, may not be accurate for the particular homes being analyzed. Inherent simplifications, such as the use of an isothermal, single-zone model for the conditioned space, could introduce significant error into the model predictions. Modeling algorithms, such as ground-coupling models for foundation heat transfer, may not accurately model the energy performance of a particular component of the home. There may even be coding errors in the RBES program itself, which can contain hundreds-of-thousands of lines of computer code.

Several prior studies have sought to assess various RBES models by comparing predictions to measured energy, but most only assess accuracy on a relatively small sample of homes [2–5]. EnergyPro was found to, on average, overestimate the combined electricity and natural gas savings expected from retrofits by nearly a half in

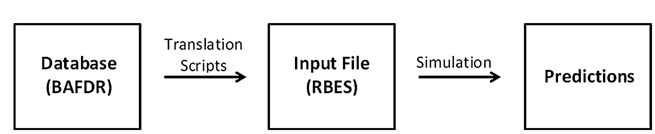

Fig. 2. Outline of general RBES simulation process. The BAFDR is NREL's Building America Field Data Repository that contains data on the homes used in this study.