Applied Energy 144 (2015) 106–113

Contents lists available at ScienceDirect

Applied Energy

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

Automated measurement and verification: Performance of public domain whole-building electric baseline models

Jessica Granderson [a,*], Phillip N. Price [a], David Jump [b], Nathan Addy [a], Michael D. Sohn [a]

[a] Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
[b] Quantum Energy Services & Technologies, Inc., 2001 Addison St., Suite 300, Berkeley, CA 94704, USA

HIGHLIGHTS

- All five tested baseline models predict annual energy use with similar accuracy.
- A twelve-month training period leads to errors less that 8.5% in most buildings.
- A shorter training period is sufficient for models that adjust for weather.
- Combining buildings into a portfolio leads to lower relative error in energy use.
- Evaluation methods can be applied to black-box models.

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form 20 November 2014
Accepted 7 January 2015
Available online 17 February 2015

Keywords:
Baseline prediction
Energy savings
Performance accuracy
Whole-building energy
Energy efficiency programs
Energy management and information systems

ABSTRACT

We present a methodology to evaluate the accuracy of baseline energy predictions. To evaluate the predictions from a computer program, the program is provided with electric load data, and additional data such as outdoor air temperature, from a "training period" of at least several months duration, and used to predict the energy use as a function of time during the subsequent "prediction period." The predicted energy use is compared to the actual energy use, and errors are summarized with several metrics, including bias and mean absolute percent error (MAPE). An important feature of this methodology is that it can be used to assess the predictive accuracy of a model even if the model itself is not provided to the evaluator, so that proprietary tools can be evaluated while protecting the developer's intellectual property. The methodology was applied to evaluate several standard statistical models using data from four hundred randomly selected commercial buildings in a large utility territory in Northern California; the result is a statistical distribution of errors for each of the models. We also demonstrate how the methodology can be used to assess the uncertainty in baseline energy predictions for a portfolio of buildings, which is an issue that is important for the design of utility programs that incentivize energy savings. The findings of this work can be used to (1) inform technology assessments for technologies that deliver operational and/or behavioral savings; and (2) determine the expected accuracy of statistical models used for automated measurement and verification (M&V) of energy savings.

1. Introduction

Energy Management and Information Systems (EMIS) span a spectrum of technologies and services including energy information systems (EIS), building automation systems, fault detection and diagnostics, and monthly energy analysis tools [1]. Tools such as EIS have enabled whole-building energy savings of up to 10–20% with simple paybacks on the order of 1–3 years [2,3] through multiple strategies such as: identification of operational efficiency improvement opportunities, fault and energy anomaly detection, and inducement of behavioral change among occupants and operations personnel.

In addition to *enabling* savings, some EMIS also automate the quantification of whole-building energy savings, relative to a baseline period, using empirical models that relate energy consumption to parameters such as ambient weather conditions and building operation schedule [4–8]. Interval meter data enable the use of baseline models that have several advantages over the monthly models that have traditionally been used to characterize whole-building energy performance [9–11]: they can determine the

* Corresponding author at: 1 Cyclotron Rd., MS 90-3111, Berkeley, CA 94720, USA. Tel.: +1 (510) 486 6792.
E-mail address: JGranderson@lbl.gov (J. Granderson).

http://dx.doi.org/10.1016/j.apenergy.2015.01.026
0306-2619/

relationship between temperature and electric load more accurately and with a shorter duration of data, and they can make predictions at a much finer timescale.

Automated baseline models can be used to streamline the whole-building measurement and verification (M&V) process, greatly reducing the cost compared to traditional processes, which require a level of building engineering expertise that limits scalability. However, several questions remain to be answered before energy managers and utility programs can confidently adopt emerging automation capabilities. For example, in energy efficiency applications one objective is to quantify and minimize the uncertainty in reported whole-building savings, which depends on baseline model effectiveness, building predictability, portfolio aggregation effects, and depth of savings being measured [6].

This paper presents an extension of prior research [5] on how to assess the accuracy and usefulness of whole-building energy models by testing predictions of baseline energy use against actual energy use. We demonstrate the method by applying it to a large random sample of commercial building data to answer the following questions of practical importance:

1. What is the *state of public domain models*, i.e., how well do they perform, and what are the associated implications for automated whole-building measurement and verification?
2. How can buildings be pre-screened to identify those that are highly model-predictable and those that are not, in order to identify estimates of building-energy savings that have small errors/uncertainty?

In this study we evaluated only public-domain whole-building baseline energy models for which outdoor air temperature is the only predictive variable. However, the methodology that we used can be applied to evaluate any model, including models that make use of additional data such as occupancy levels, business types, or building types.

While resources such as ASHRAE Guideline 14 and the International Performance Measurement and Verification Protocol (IPMVP) [12,13], establish procedural and quantitative requirements for baseline model construction, goodness of fit to data during the model training period, and rules of thumb for model application given different expected depths of savings, they do not provide a general means of assessing model performance during a *prediction* period. They also provide little guidance on using interval data as opposed to monthly data. The methodology presented in this work extends the principles in these existing resources to quantify model predictive accuracy after the training period, and suggests key performance metrics to quantify model accuracy in the context of whole-building M&V. Lengthy periods of interval meter data from several hundreds of buildings are collated to form a 'test' data set, and statistical cross-validation is performed to gauge performance relative to the M&V-focused metrics and time scales of interest.

This methodology shares important similarities to the approaches used in the ASHRAE 'shootouts' of the mid and late 1990s [14,15]. In both cases, cross-validation is used to determine model error, and in both cases, normalized root mean squared error is included as a performance metric. However, the ASHRAE shootouts were limited to data from a total of two buildings, and the cross-validation was conducted only for short subsets of the model training period.

An important feature of this work is that the methodology can be used to objectively assess the predictive accuracy of a model, without needing to know the specific algorithm, or underlying form of the model. Therefore, proprietary tools can be evaluated while protecting the developer's commercial intellectual property. The findings of this work can be used to (1) inform technology assessments for EMIS products and other technologies that deliver operational and/or behavioral savings; and (2) set a floor of performance of automated M&V, that can be used to consider requirements for utility or corporate efficiency programs, including the tradeoffs between cost, and accuracy.

2. Baseline model performance assessment methodology

Baseline energy use models characterize building load or consumption according to key explanatory variables such as time of day, and weather. These baseline models are used for a variety of purposes in EMIS, including near real-time energy anomaly detection, and near future load forecasting, as well as quantification of energy or demand savings [2,4].

Baseline model accuracy is critical to the accuracy of energy savings that are calculated according to the IPMVP. For both whole-building and measure isolation approaches (IPMVP Options B and C) the baseline model is created during the "pre-measure" period, before an efficiency improvement is made. The baseline model is then projected into the "post-measure" period, and energy savings are calculated based on the difference between the projected baseline and the actual metered use during the post-measure period [13]. Therefore, the error in reported savings is proportional to the error in the baseline model forecasts.

2.1. General methodology

Prior work established a *general* 4-step statistical procedure that can be used to evaluate the performance, i.e. predictive accuracy, of a given baseline model [5].

(1) Gather a large test data set comprised of interval data from hundreds of commercial buildings.
(2) Split the test data from each building into model training and model prediction periods. These periods can be tailored according to the specific application or use case of interest, e.g., energy efficiency savings, demand response load reductions, or continuous energy anomaly detection. For this study, the focus was measurement and verification of energy savings at the whole-building level.
(3) For a given set of baseline models, generate predictions based on the training period data, compare those predictions to the data from the prediction period, and compute statistical performance metrics based on the comparison. Again, the models of interest, and the specific performance metrics can be tailored to according to the specific application or use case.
(4) Assess relative and absolute model performance using the performance metrics that were computed in Step 3.

This process is illustrated in Fig. 1, which shows daily average loads. The first several months constitute the "training period," from which the load data and outdoor air temperature data are used to create a statistical model that predicts load as a function of the time during the week and the temperature. This model is then used to predict the load during both the prediction period and the subsequent training period.

The subject building for Fig. 1 has several features that are typical of commercial buildings: the load is temperature-dependent, and load on weekends is substantially lower than the load on weekdays. At a finer timescale, the building also has a nightly minimum load that is much lower than the daytime maximum, but of course this cannot be seen on this plot of daily averages. (Plotting 10 months of hourly or 15-min data would create a plot with so many vertical oscillations that it would be impossible to interpret).

Furthermore, sometime around the beginning of May 2011 the building's energy behavior changes: both the weekday and week-