Title

Uncertainty Estimation Improves Energy Measurement and Verification Procedures

Authors

Walter, Travis
Price, Phillip N.
Sohn, Michael D.

Publication Date

2014-05-14

LBNL-XXXX

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Uncertainty Estimation Improves Energy Measurement and Verification Procedures

Travis Walter, Phillip N. Price, Michael D. Sohn

Environmental Energy Technologies Division

May 2014

This work was performed at the Lawrence Berkeley National Laboratory, operated by the University of California, under DOE Contract DE-AC02-05CH1131.

Abstract

Implementing energy conservation measures in buildings can reduce energy costs and environmental impacts, but such measures cost money to implement so intelligent investment strategies require the ability to quantify the energy savings by comparing actual energy used to how much energy would have been used in absence of the conservation measures (known as the “baseline” energy use). Methods exist for predicting baseline energy use, but a limitation of most statistical methods reported in the literature is inadequate quantification of the uncertainty in baseline energy use predictions. However, estimation of uncertainty is essential for weighing the risks of investing in retrofits. Most commercial buildings have, or soon will have, electricity meters capable of providing data at short time intervals. These data provide new opportunities to quantify uncertainty in baseline predictions, and to do so after shorter measurement durations than are traditionally used. In this paper, we show that uncertainty estimation provides greater measurement and verification (M&V) information and helps to overcome some of the difficulties with deciding how much data is needed to develop baseline models and to confirm energy savings. We also show that cross-validation is an effective method for computing uncertainty. In so doing, we extend a simple regression-based method of predicting energy use using short-interval meter data. We demonstrate the methods by predicting energy use in 17 real commercial buildings. We discuss the benefits of uncertainty estimates which can provide actionable decision making information for investing in energy conservation measures.

Keywords: uncertainty analysis; building retrofit; measurement and verification; building energy; baseline prediction; cross-validation; change-point model; linear regression

Abbreviations

M&V	measurement and verification
ECM	energy conservation measure
HVAC	heating, ventilation, and air conditioning
ESCO	energy service company
IPMVP	International Performance Measurement and Verification Protocol

1 INTRODUCTION

Energy efficiency improvements in buildings are a cost effective approach to reducing energy use. Energy conservation measures (ECMs) reduce energy consumption through the installation of newer, and usually more efficient, equipment and appliances, retrofitting old equipment, and/or modifying operating procedures. For example, typical ECMs in commercial buildings include replacing light fixtures, retrofitting hot water boilers or heating, ventilation, and air conditioning (HVAC) fans, or changing lighting and HVAC schedules. ECMs could also include real-time anomaly detection or participation in demand response programs that aim to reduce electricity consumption at particular times. In the last two decades we have seen the market to provide such services through energy service companies (ESCOs) expand dramatically, the typical business model being reducing energy costs by implementing retrofits. A constant tradeoff for the retrofit market is between the accuracy in the energy savings estimates (and, by extension, in the payback of an ECM) and the wait needed for accurate estimates (due to waiting for the necessary post-retrofit data). This tradeoff is very important in the ESCO business because obtaining data, and the time it takes to gather them, can significantly impact the costs and return on their investment.

The effectiveness of an ECM is defined typically by the amount of energy use that is *avoided*. In other words, the difference between how much energy the building consumed over a given period, and how much it *would have* consumed without the ECM. The latter value is typically referred to as the “adjusted baseline” and the overall process of confirming ECM effectiveness is called “measurement and verification” (M&V). For examples of M&V techniques, see the International Performance Measurement and Verification Protocol (IPMVP) [1]). Critical technical questions facing M&V practitioners include not only estimating how much energy the building would have consumed, but knowing how accurate the energy estimates must be in order to be useful. These questions have important ramifications in the durations of energy data to record before and after the retrofit, the analysis methods employed, and perhaps most importantly, whether the energy saved by a given retrofit will be measurable.

IPMVP includes several classes of methods that attempt to separate the effect of the ECM from other processes that affect the building's energy consumption. These methods include installing electric meters to monitor the energy use of individual components or subsystems, and creating and exercising computer models that mimic the physical processes of the building. One of the accepted IPMVP approaches is to create a statistical model, based on data from before the ECM was implemented, that can adjust for changes in weather (or in other parameters, if known). The model is then applied to the period after the ECM is implemented, to predict the "baseline" energy use. IPMVP methods for creating these statistical models were developed many years ago, when the only whole-building energy consumption measurements were obtained from monthly utility bills.

In recent years, time-resolved "interval data" have become more commonly available; these are data on electric load as a function of time, typically at 15-minute to 1-hour intervals. Interval data provide new opportunities for M&V, including a reduction in the duration of data required to determine the dependence between weather and building energy use. If the only available electricity consumption data are monthly, then an M&V practitioner must wait until there have been both warm months and cool months in order to determine the relationship between outdoor air temperature and energy use. However, if interval data are available then significantly fewer data may be needed (e.g., from just a few hot and cold days, which may even occur within the same month). The use of interval data should therefore allow whole-building M&V to be completed using much shorter pre- and post-install periods than are currently recommended by IPMVP.

Several methods for computing baseline energy that take advantage of interval data are reported in the literature. Mathieu *et al.* [2] provides a good summary of energy prediction methods. Coughlin *et al.* [3] considers methods that average load profiles from the last several days. Granderson *et al.* [4] describes several other methods, including models based on binning, nearest neighbor models, and nonlinear weighted regressions, in which predictions are based on measurements from similar conditions. Claridge [5] and Taylor *et al.* [6] discuss more complex mathematical methods, including autoregressive integrated moving average models, neural network models, exponential smoothing models, and Fourier series models. These methods, and many others (e.g., [7, 8]), have substantially advanced both the state of the art in M&V and the suite of tools that a practitioner may use to confirm energy savings. The success of these methods, coupled with the wide installation of interval meters means it is likely that these methods will be used for many IPMVP-style M&V techniques in the future. We are already seeing a host of ESCOs emerging to provide this service.

While advancing the state of the art, a critical addition to these tools is a better method to estimate uncertainty in the baseline estimates. Uncertainty is important because it provides actionable information for ESCOs, building operators, and portfolio managers. It provides these stakeholders the information necessary to assess the risks of a financial investment [9, 10]. Quantifying uncertainty also allows M&V practitioners to weigh the limitations and benefits of the amount of data used to compute baseline estimates and retrofit savings: the benefit of an ECM can only be definitively demonstrated if the savings are large relative to the uncertainty in the energy use estimates. The savings is the difference between the baseline energy use and the actual energy use, and since the latter is known from the utility meter, the uncertainty in the savings is equal to the uncertainty in the baseline energy use. The baseline energy use is uncertain because it is the amount it *would have consumed* in the absence of the ECM, which is not measurable and therefore must be predicted, and these predictions are subject to uncertainty.

There is often substantial uncertainty in the baseline prediction, due to the fact that the building's energy use varies with weather, occupancy, operating hours, and many other factors, many of which have unknown relationships to the energy consumption. Such details are often not measured or recorded due to costs or time. Uncertainty in baseline estimates result for three main reasons:

1. Energy use in the building varies due to factors not included in the models. For example, more or fewer people may use the building, hours of operation may change, equipment may be replaced or its usage pattern may change, and so on.

2. Input parameters are subject to error. Outdoor air temperature or humidity measurements may be inaccurate or may be measured miles from the building and thus may not accurately represent site conditions.

3. The model is misspecified. Any statistical model includes assumptions, some of which will not be perfectly accurate. For example, ordinary linear regression assumes that model errors are independent, identically-distributed draws from a normal distribution, but in predicting building electric load the errors are often not independent, not identically distributed, and not drawn from a normal distribution. Uncertainty estimates provided by such models are often reported [11–13] but are sometimes not accurate.

The remainder of this paper is organized as follows: In Section 2 we present a regression-based model for estimating baseline electric load and an algorithm that uses cross-validation to quantify uncertainty in baseline predictions. In Section 3 we illustrate the regression model and the uncertainty algorithm using real data from 17 commercial buildings. Finally, in Section 4 we discuss the application of these methods to the M&V process.

2 METHODS

The method for predicting the statistical distribution of the baseline electric load is a two-stage process. In the first stage, we predict the expected electric load. In the second stage, we complete the characterization of the distribution by predicting the uncertainty bounds of the predicted electric load. In this paper, we select a particular model for the first stage, but the uncertainty quantification method used in the second stage can be applied to any model that predicts expected load. In addition, the methods presented here can be used to easily compare the performance of competing models.

2.1 Load Prediction

In this section we describe a linear regression model for predicting whole-building electric load. This approach can be applied to end-uses with sub-metered data, but a more common application is predicting whole-building load. We use a linear regression model based on the time of week and the outdoor air temperature to predict the expected baseline. This model is robust, easy to use and interpret, is computationally efficient, and provides a good fit to the data (both objectively, and when compared to other prediction methods [14]). In addition, it requires relatively little data to be effective. Since M&V practitioners are often faced with very limited data, this is an important benefit. Not only does the regression model rely on only two easily measured values (time and temperature), but it may require shorter measurement periods than are traditionally used (e.g., a few months, rather than a full year). Short-interval data allows the model to be fit to many measurement values, but does not require a long time to collect these data. In addition, fitting the model using short interval data (hourly or sub-hourly) allows the model to extract information about the relationship between energy use and outdoor air temperature that would be obscured if the model were fit to monthly data. This model provides an M&V practitioner with an accurate model of electric load while requiring minimal investment in measurement equipment and monitoring time.

In commercial buildings, it is typical for load to be high during afternoons (when the outdoor air temperature is high and the building is heavily occupied) and low during the nights and weekends (when temperatures are low and/or the building is unoccupied). Many office buildings have an "occupied" mode during which the indoor air temperature is maintained at a comfortable level and an "unoccupied" mode during which the indoor air temperature is either uncontrolled or is maintained only within a broad band. In a typical commercial building, the dependence of load on temperature is a nonlinear function of temperature, and depends on which mode the building is in. In occupied mode, it is common for load to be positively correlated with outdoor air temperature at high temperatures (when using energy for cooling), negatively correlated at low temperatures (when using energy for heating), and relatively uncorrelated at moderate temperatures (when not using energy for cooling or heating). In unoccupied mode, load typically has little correlation with outdoor air temperature. With this knowledge, we selected a model structure that is limited to the "time of week" and the outdoor air temperature as predictor variables. A similar model is described in more detail in [2]. This development and demonstration of this research applies to any general forecast model, such as one that includes additional explanatory variables (e.g., humidity, occupancy). However, since these data are not commonly recorded, we did not select such a model.

Consider K measured data points, where data point k is from time t_k and includes a temperature measurement T_k and a load measurement L_k, for $k = 1, \ldots, K$. We model the load as the sum of a time-dependent portion and a temperature-dependent portion $\hat{L}_k = \hat{L}_{k,time} + \hat{L}_{k,temp}$.

We model the time-dependent portion of load in a way that captures patterns such as lower load at night than during the day, lower load on weekends than on weekdays, and lower load on Friday afternoon than on other weekday afternoons. We model time-dependence by dividing the week into 168 one-hour intervals and assign an indicator variable and coefficient to each interval. The time indicator variable $\tau_{k,i} = 1$ if t_k is in interval i and $\tau_{k,i} = 0$ otherwise, for $i = 1, \ldots, 168$. The time-dependent portion of the predicted load is computed by summing the product of indicators and coefficients over all 168 time intervals $\hat{L}_{k,time} = \sum_{i=1}^{168} \alpha_i \tau_{k,i}$. The time indicators serve to select which coefficient contributes to the predicted energy use. For a given data point, one of the 168 coefficients is multiplied by one and added to the predicted load, and the other 167 coefficients are multiplied by zero and have no effect.

We model the temperature-dependent portion of load so as to describe the behavior of a typical building's heating and cooling system. We model temperature-dependence using a piecewise-linear and continuous function. In order to achieve this functional form, we divide the temperature range into four intervals, and assign a temperature component and coefficient to each interval. The temperature is written as the sum $T_k = \sum_{j=1}^{4} \theta_{k,j}$ where the temperature components $\theta_{k,j}$ are the portion of the temperature T_k in temperature interval j. For example, if the temperature intervals are $20°F - 40°F$, $40°F - 60°F$, $60°F - 80°F$, and $80°F - 100°F$, and the temperature is $T_k = 75°F$, then the temperature components are $\theta_{k,1} = 20°F$, $\theta_{k,2} = 20°F$, $\theta_{k,3} = 15°F$, and $\theta_{k,4} = 0°F$. The temperature-dependent portion of the predicted load is computed by summing the product of components and coefficients over all temperature intervals $\hat{L}_{k,temp} = \sum_{j=1}^{4} \beta_j \theta_{k,j}$.

The predicted load for data point k is the sum of the time-dependent portion and the temperature-dependent portion $\hat{L}_k = \hat{L}_{k,time} + \hat{L}_{k,temp} = \sum_{i=1}^{168} \alpha_i \tau_{k,i} + \sum_{j=1}^{4} \beta_j \theta_{k,j}$. The regression coefficients α_i and β_j are computed using ordinary least squares by minimizing the sum of the squared error $\sum_{k=1}^{K} (L_k - \hat{L}_k)^2$.

To allow the dependence of load on time and temperature to be different when the building is in occupied and unoccupied modes, we model the two modes separately. We first split the data into two disjoint subsets of the original data set, one for occupied mode and one for unoccupied mode. We compute one set of regression coefficients that best fit the occupied data, and another set of coefficients that best fit the unoccupied data. To predict the load at a particular time and temperature, we apply the corresponding regression coefficients.

2.2 Computing Uncertainty

In this section we describe a general method for quantifying uncertainty in baseline energy predictions. The approach can be used to compute uncertainty on any time interval, but we demonstrate the approach by computing uncertainty bounds on monthly energy totals. This is the time scale at which energy predictions are commonly preferred by M&V practitioners because building owners making decisions to invest in ECMs are interested in estimates of energy savings computed over time scales of months or years.

Consider the relationship between the model error and the amount of data used to fit the model. Model error is typically reduced by using more data to fit the model, but there is a limit to this effect because (1) when stochastic variability is present, any model will eventually cease to improve even when more data are collected, and (2) building energy behavior changes over time, so knowing how the building performed in the distant past does not predict how it will perform in the future. For example, over a period of months or years the base load on weeknights is likely to change, so that data from weeknights long ago will not improve the prediction of the next weeknight. In other words, the model must have enough data to characterize a wide range of load and temperature relationships, and to distinguish between the building's average behavior and inherent stochastic variability. However, data from too far back in time can be useless or harmful because those data no longer reflect the building's current behavior.

We now define an algorithm to compute the probability distribution of the residuals (the error between the measured data and the model predictions). The uncertainty algorithm is based on cross-validation (i.e., partitioning the data into subsets, fitting the model to one subset, then validating the model with another subset). We separate the data set (e.g., one year of data) into many shorter time intervals (e.g., one month). We fit the model to the data in one interval, then use the model to predict the data in the next interval. We then compare those predictions to the measured data during the prediction interval and compute the residuals. We repeat this process of computing residuals for each interval in the data set. We suspect the statistical distribution of the resulting set of residuals can be used to estimate the uncertainty in the model predictions.

The algorithm is defined as follows:

Define the sequences of measured load data $L = \{L_1, \ldots, L_K\}$, time data $t = \{t_1, \ldots, t_K\}$, and temperature data $T = \{T_1, \ldots, T_K\}$. Start by separating the data set into M smaller intervals, one for each month, i.e., $\{L\}^m$, $\{t\}^m$, and $\{T\}^m$ are the load, time, and temperature time sequences for month m, where $m = 1, \ldots, M$. For $m = 1, \ldots, M-1$, the residuals samples are computed with the following algorithm.

1. Fit a model to the data from month m by using $\{L\}^m$, $\{t\}^m$, and $\{T\}^m$ to compute the model parameters for month m. Any model can be used, but here we use the model described in Section 2.1. In this case, the model parameters for month m are the regression coefficients $\{\alpha\}^m$ and $\{\beta\}^m$.

2. For the following month, month $m+1$, make load predictions $\{\hat{L}\}^{m+1}$ using the model parameters from month m.

3. Compute the measured energy consumption I^{m+1} and the predicted energy consumption $\hat{I}^{m+1}$ in month $m+1$ by summing the actual loads and predicted loads over of the time intervals in the month.

4. Compute the residual $R^{m+1} = I^{m+1} - \hat{I}^{m+1}$ in month $m+1$.

When the algorithm finishes, the set of residuals $\{R\}$ contains $M-1$ residual samples, each representing the error in the energy consumption prediction during a different month.

We would like to answer the question: If we fit a model using several months of data, how accurately can the next month's energy consumption be predicted? We propose to answer this question by assuming that the set of errors, $\{R\}$, has the same statistical distribution as the error in the next month's energy consumption prediction. We test the validity of this assumption empirically in Section 3.

The set of errors, $\{R\}$, was generated by fitting the model using *one* month of data and using it to predict the energy used in the following month, which is somewhat different from the situation of eventual interest, in which the model is fit to several months of data. On one hand, the errors $\{R\}$ might tend to be too large in magnitude because a model fit to a single month of data may be subject to more stochastic variability than a model fit to several months of data. On the other hand, the errors $\{R\}$ might tend to be too small in magnitude because fitting each month separately allows the model to adjust to features of the data that are incorrectly assumed constant when fitting the model to several months of data (e.g., changes in base load or temperature sensitivity). In the next section, we investigate the extent to which the set $\{R\}$ represents the statistical distribution of errors in the situation of interest.

3 RESULTS

In this paper, we analyze whole-building electric load data from 17 government and commercial office buildings from various locations and climates throughout the United States. Most of the buildings have measured load at 15 minute intervals, one has data at 10 minute intervals, and the remainder have data at 1 hour intervals. Roughly half of the buildings have 27 months of data and roughly half have 12 months of data. The majority of the buildings provided outdoor air temperature data measured on site. For the rest of the buildings, outdoor air temperature data from a nearby weather station were acquired from `http://www.wunderground.com`. Missing outdoor air temperature data were interpolated linearly when only a few hours of data were missing. When more temperature data were missing, the temperature and load data from that time interval were excluded from the data set. We start by illustrating the modeling technique described in Section 2.1 by focusing on measured data from one particular office building. The same modeling technique is used for each building. We then apply the algorithm described in Section 2.2 by utilizing data from all 17 buildings.

Figure 1 shows one week of temperature and load data. We believe that this building, like most office buildings, operates in occupied and unoccupied modes. We assumed the building mode depends on the time of day and the day of the week, and determined these times by inspection. The temperature exhibits a clear pattern of high temperatures during the day and low temperatures at night, and shows gradual variation throughout the week as well. Similarly, the load is high in the afternoons and low at night, but is

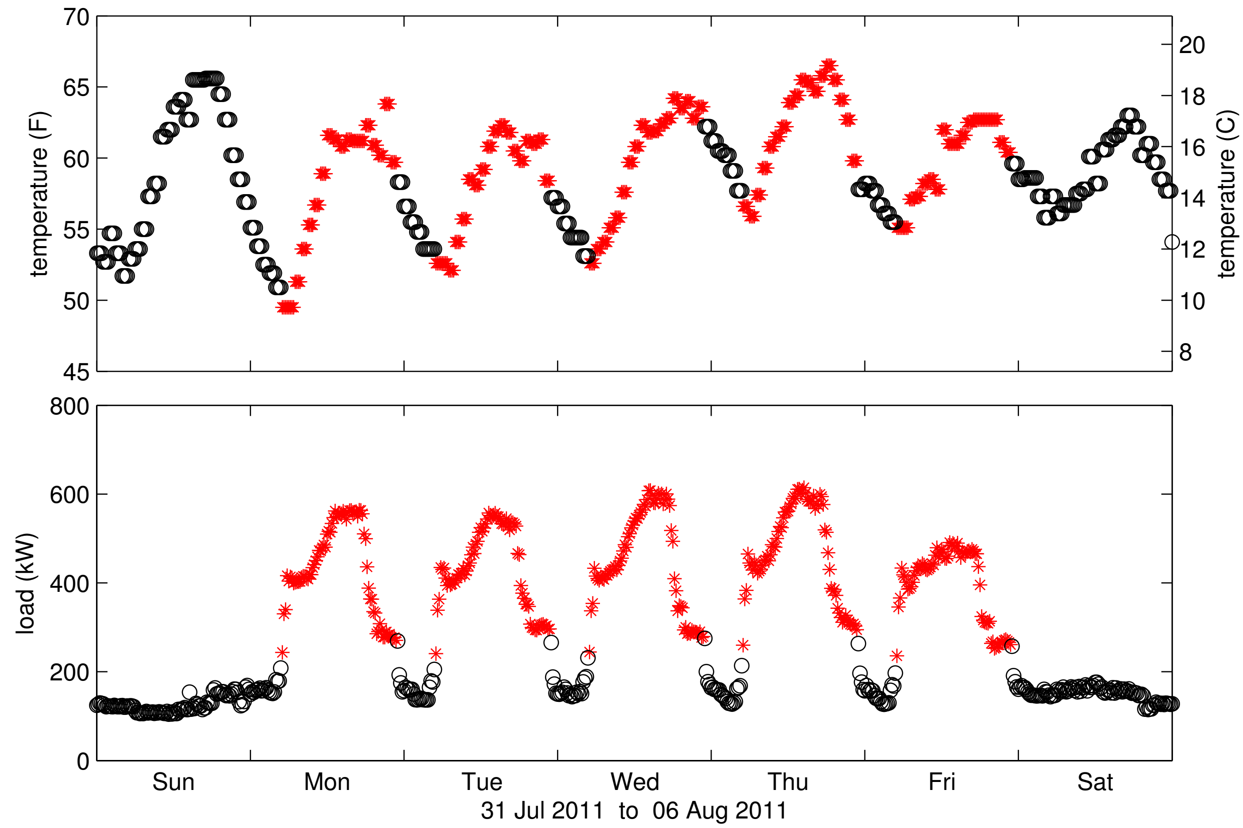

Figure 1: Temperature and load vs. time in occupied mode (red stars) and unoccupied mode (black circles). Occupied hours are Monday-Friday 5am-11pm. Data are for Building 5.

also low throughout the entire weekend. The shape of the load curve during the day is different than that of the temperature curve (e.g., the peak in load at the start of the occupied period), indicating the dependence of load on more than just temperature. In addition, load is lower on Friday afternoon than on other weekday afternoons, indicating the dependence of load on both time of day and day of week. These observations support our choice of a load prediction model that depends on both temperature and hour of week.

Figure 2 shows load plotted against temperature. In unoccupied mode, temperature appears to have little correlation with load. In occupied mode, temperature is positively correlated with load, particularly at high temperatures. This behavior supports our choice of a load prediction model that fits load to a function of temperature, and that fits separate models for occupied and unoccupied modes.

Figure 3 illustrates the piecewise-linear and continuous portion of the model. The time-dependent portion of the modeled load, $\hat{L}_{k,time} = \sum_{i=1}^{168} \alpha_i \tau_{k,i}$, is subtracted from the measured data L_k, and the result is plotted against the temperature T_k. The temperature-dependent component of the modeled load,

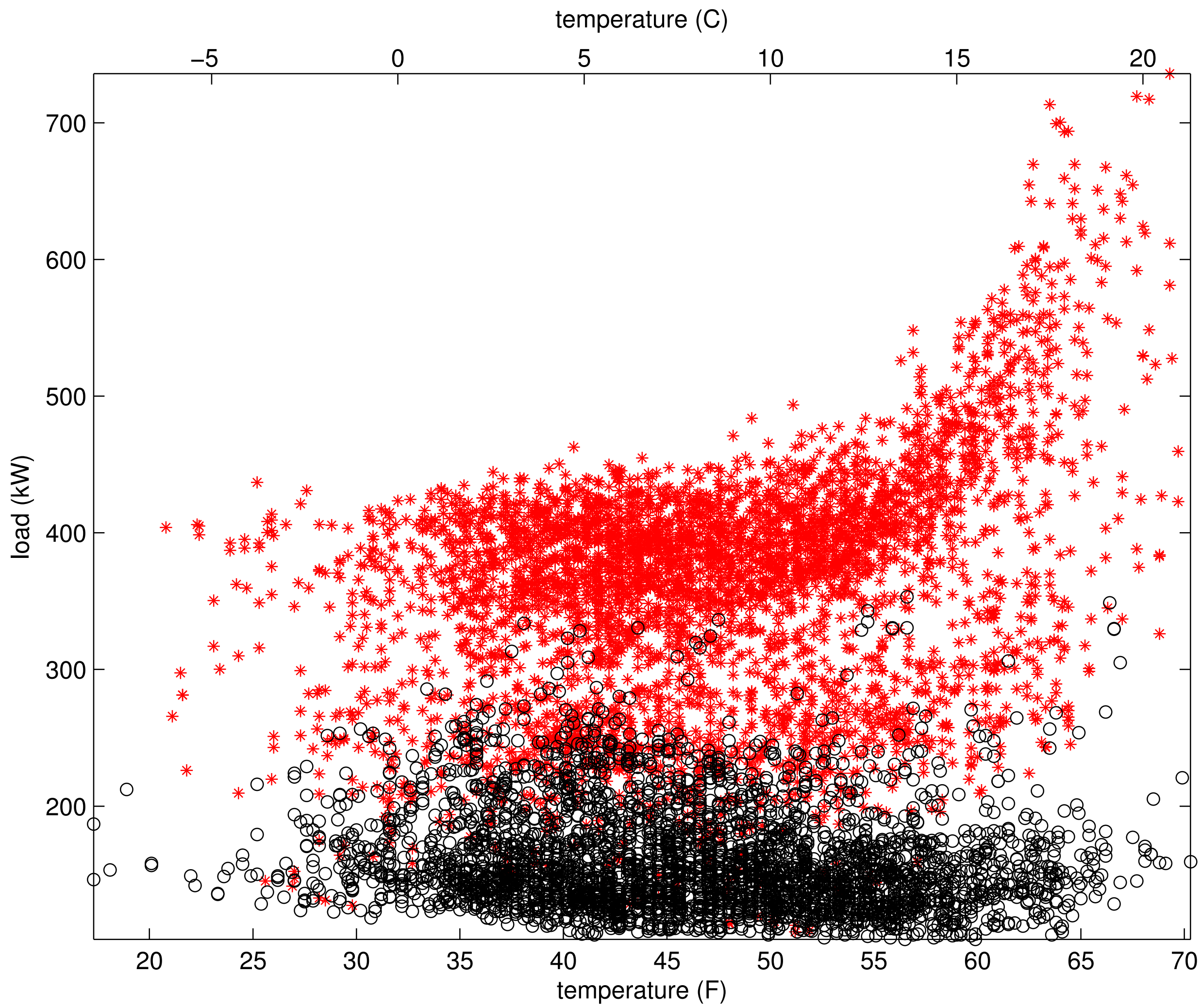

Figure 2: Load vs. temperature in occupied mode (red stars) and unoccupied mode (black circles), showing random subset of 10% of data. Data are for Building 5.

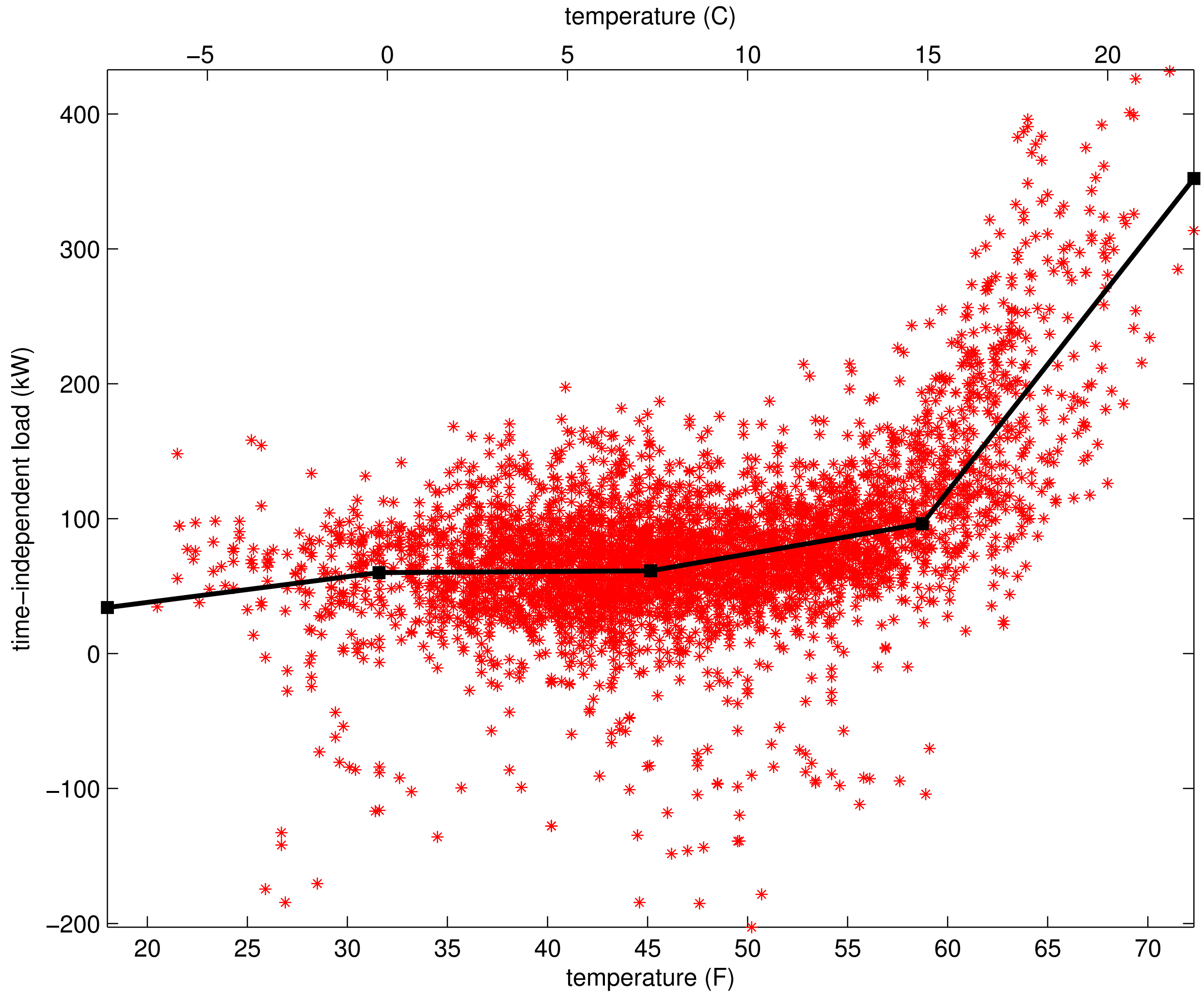

Figure 3: Time-independent load vs. temperature in occupied mode, as measured (red stars) and modeled (black lines), showing random subset of 10% of data. Data are for Building 5.

$\hat{L}_{k,temp} = \sum_{j=1}^{4} \beta_j \theta_{k,j}$, is superimposed. For lower temperatures, temperature has little effect on load, but at high temperatures, temperature is correlated with load. The agreement between the modeled and measured values of time-independent load justifies our choice of a load prediction model that fits load to a piecewise-linear and continuous function of temperature.

Figure 4 shows the measured and predicted load for three separate weeks in different seasons. Overall, in each of the three weeks, the predicted load is very close to the measured load, despite the variability of outdoor air temperatures and daily load shapes with season. In June, there is a peak in load in the late afternoon on Monday and Tuesday, while in May, the late afternoon peak is more pronounced later in the week; the load in January shows no such peak. The model does not capture this peak well because it is

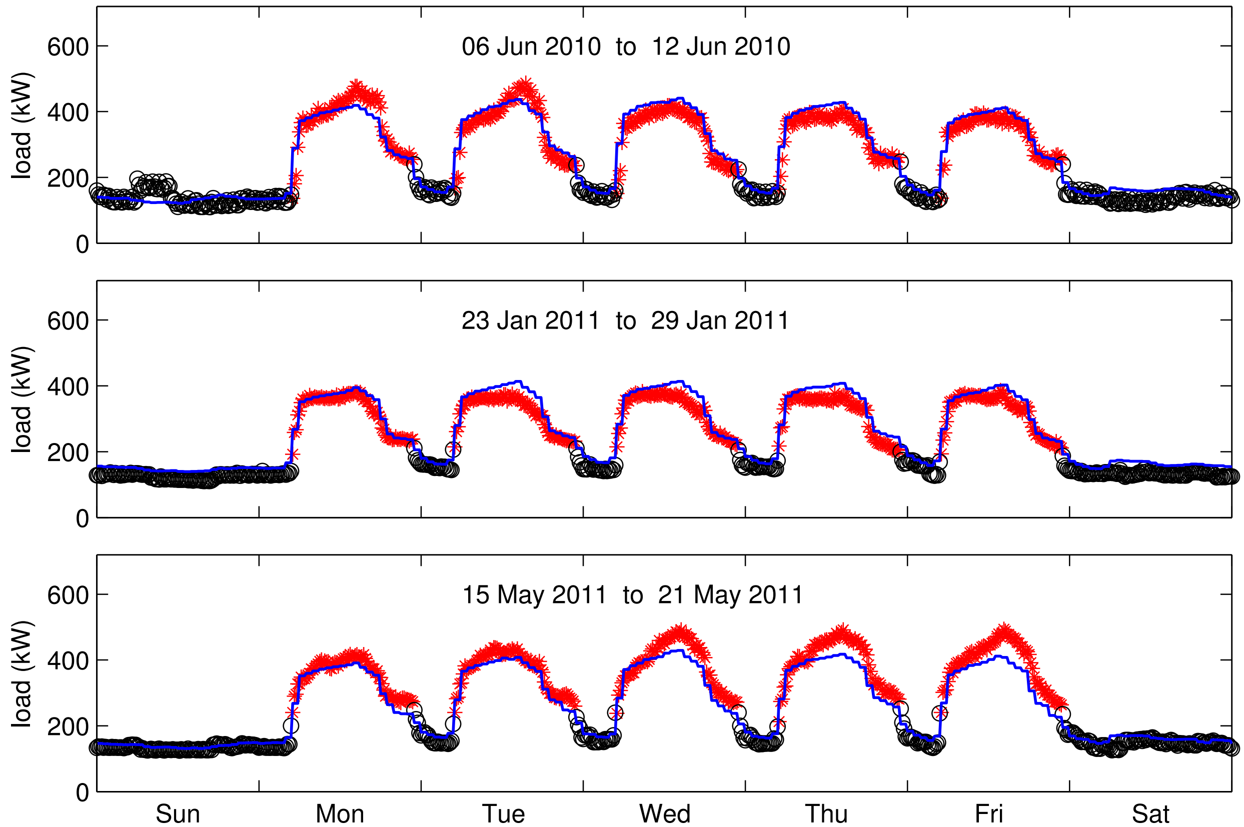

Figure 4: Load vs. time, as measured and predicted (blue line). Occupied mode (red stars) and unoccupied mode (black circles) modeled separately. Data are for Building 5.

averaging behavior over many weeks, and most weeks do not exhibit this peak. A similar argument explains the model underpredicting load for a short interval on Sunday morning in June. While predictions may be high (e.g., the middle of the week in January) or low (e.g., the end of the week in May) for short periods, predicted totals on longer time scales (which are of interest to M&V practitioners) are accurate.

In Figure 5, measured load is plotted against predicted load, illustrating reasonable agreement between the model and the data. There is larger variation at high loads than at moderate loads. The linear regression coefficients are computed to reduce error at moderate loads, which are very common, at the expense of allowing larger error at high loads, which are much less common. Since these high loads occur relatively infrequently, their impact on monthly energy totals will be minimal, and the errors at high loads will not be problematic to M&V practitioners interested in long time scale predictions.

Figure 6 shows the relationship between model error and the amount of data used to fit the model for 5 of the 17 buildings in the data set. For each building, the model is fit several times using different durations

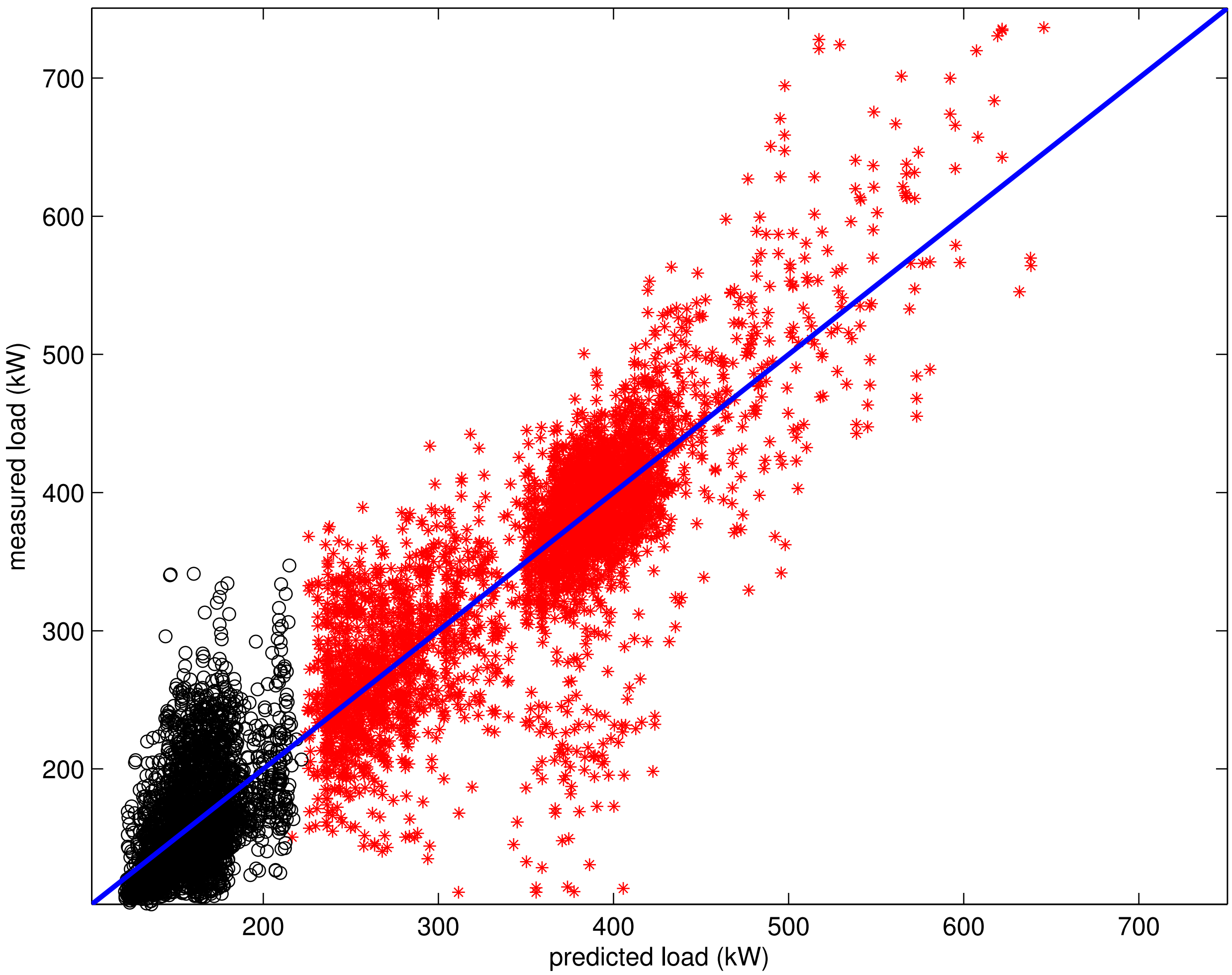

Figure 5: Measured load vs. predicted load (blue line), showing random subset of 10% of data. Occupied mode (red stars) and unoccupied mode (black circles) modeled separately. Data are for Building 5.

of data. Each time, it is used to predict the load during the final month. On the horizontal axis is the length of the interval used to fit the model (in units of days), and on the vertical axis is the normalized difference between the measured load and the predicted load during the final month. In some cases, model error is reduced by using more data to fit the model (e.g., Buildings 4, 9, and 13), but for many of the 17 buildings tested, the error when fitting to only a few months of data is about the same magnitude as when using four or more months of data (e.g., Buildings 5 and 7). This observation suggests that a method that uses the error when fitting to a small amount of data helps to predict the error when fitting to a large amount of data.

In this study, we observed that model error no longer reduces when more than a few months of data are used. However, we should note that this result is specific to the particular model used here. If a model other than the one described in Section 2.1 is used, error may continue to reduce when more data is used. In addition, this result is specific to the data for the buildings in this study; other buildings may illustrate different behavior. An important contribution of this work is developing a practical and empirical approach that M&V practitioners can apply to investigate the relationship between model error and the amount of data for any given model and data set. Practitioners can use the results of such an analysis to decide between multiple competing models, whether a model is suitable, the degree to which additional data are likely to improve the analysis, etc.

Figure 6 also illustrates the type of analysis that can be performed by M&V practitioners to explore how much data is needed for a M&V analysis. Some minimum level of model performance is necessary. However, the right level of accuracy depends in part on the intended application. Collecting more data may improve model accuracy, but it might not be worthwhile if the improvement is small relative to the effort (and therefore costs) to obtain measurements for longer periods (e.g., delaying the installation of retrofits and reconciling retrofit savings). Since different M&V projects have different needs on prediction accuracy, M&V practitioners can balance acceptable model accuracy against additional measurements using an analysis similar to that shown in Figure 6.

To illustrate the method for computing uncertainty at the portfolio level, we applied the algorithm in Section 2.2 to all 17 buildings in the data set. For each building, we estimate the uncertainty of the residuals of the predicted energy use. We do so by separating the final month of data from the data set, and applying the algorithm to all except the final month. In other words, we generate the set of residual samples $\{R\}$ as though the final month of data does not exist. These residual samples serve as an estimate of the uncertainty in the predicted energy use during the final month.

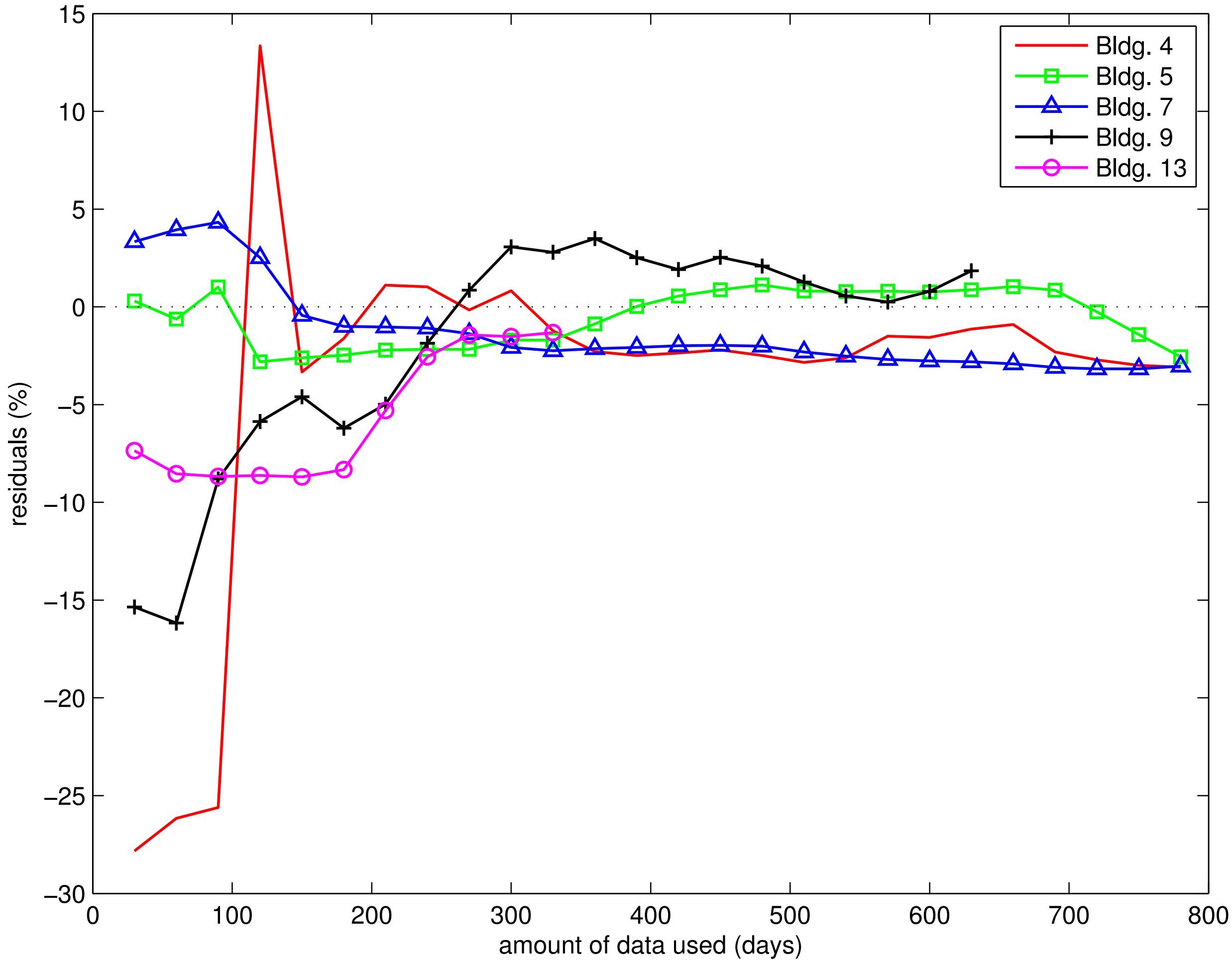

Figure 6: Residuals normalized by measured value vs. number of days of data used to fit model. Data are predicted for the final month.

To assess the validity of the uncertainty estimates, we also compute the actual residual for each building. We first fit the model in Section 2.1 to all of the data except the final month, then use it to predict the energy use in the final month. We compute the actual residual R_{act} by subtracting the energy prediction during the final month from the measured energy during the final month. In a typical application of the algorithm, the actual residuals would not be available. We compute them here as a means to assess the uncertainty algorithm's validity.

As an example, consider a building with 12 months of data from January through December. We separate December from the data set and apply the algorithm in Section 2.2 with $M = 11$ to the January through November data, resulting in a set of residuals $\{R\}$ containing 10 samples. We then fit the model in Section 2.1 to the January through November data and use it to predict the energy use in December. We compute the actual residual R_{act} as the difference between the prediction during December and the measured energy during December.

For each building, this results in a set of samples $\{R\}$ that constitute the estimated distribution of the residuals, and one actual residual R_{act}, of the predicted energy use during the final month. If the uncertainty estimation algorithm provides a good approximation of the model uncertainty, then one would expect the actual residual for the final month to be consistent with the predicted uncertainty. In other words, the actual residual should appear to have been drawn from the same distribution as the residual samples. For example, after many trials (i.e., for many buildings), one would expect that for half of the trials, the actual residuals for the final month are within the 1st and 3rd quartiles of the residual samples, and similarly for other quantiles. If the actual residuals are within the 1st and 3rd quantiles for more than half of the trials, then the uncertainty is likely overestimated. Likewise, if the actual residuals are within the interquartile range for fewer than half the trials, then the uncertainty is likely underestimated.

Figure 7 depicts the distributions of the residual samples $\{R\}$ and the actual residuals R_{act} for the 17 buildings in the data set, and shows that the uncertainty predicted by the estimation algorithm in Section 2.2 is consistent with the actual residuals. The residuals lie within the interquartile range for roughly half the buildings. In addition, very few of the actual residuals lie outside the extreme values of the residual samples distributions. Figure 7 shows that the uncertainty estimates are neither underestimates nor overestimates, and that the distribution of the residual samples is a consistent estimate of the uncertainty in the energy use predictions. For the 17 buildings studied, M&V practitioners can accurately compute the uncertainty in energy use predictions by applying the algorithm in Section 2.2. These uncertainty estimates

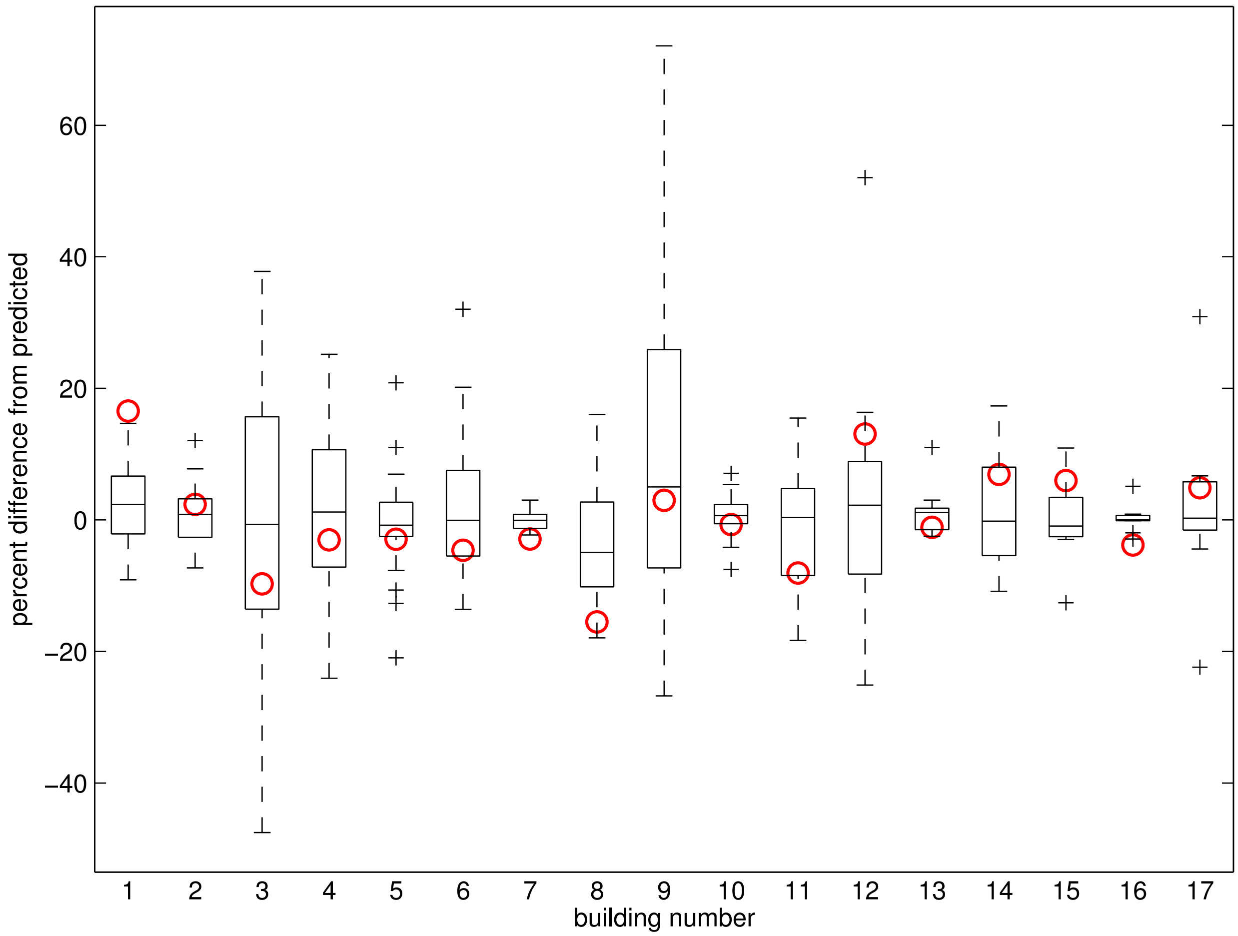

Figure 7: Integrated load residuals for 17 buildings. Red circles are actual residuals (fit model to all except final month, then predict final month). Box plots show median, quartiles, extreme values (dotted), and outliers (+) for residual samples from the uncertainty estimation algorithm.

can then be used to weigh the risks, costs, and benefits of investing in ECMs.

A more robust test of the algorithm in Section 2.2 would be to compare several trials of actual residuals against the proposed distributions for each building, rather than just one (as is shown in Figure 7). In addition, the actual residuals could be compared to the proposed distributions for more than 17 buildings. However, these tests require significantly more data than we had access to and must be the subject of future research.

Figure 7 also shows that the energy predictions for some buildings are much more uncertain than for others. For example, compare the tight distributions of errors in Buildings 13 and 16 to the wide distributions in Buildings 3 and 9. If the owner of Building 16 is planning to implement an ECM that is expected to reduce the building's energy use by 10%, they can be confident that they will easily see the savings using a whole-building baseline approach. In contrast, there is no hope of seeing such an effect in Building 9 because the expected savings are small compared to the uncertainty in the baseline prediction.

The uncertainty in the energy predictions could be due to factors not included in the model (e.g., occupancy), error in temperature measurements, model misspecification, or the duration of data used to fit the model. The fact that these uncertainties can be large for some buildings and small for others illustrates the benefit of the methods presented here: quantifying uncertainty in energy predictions provides actionable information to M&V practitioners.

4 DISCUSSION AND CONCLUDING REMARKS

In this paper, we addressed the problem of quantifying energy savings due to implementing energy conservation measures (ECMs), and isolating those savings from energy differences due to other influences (e.g., weather, time of day, day of the week). We presented an energy prediction technique that uses linear regression on time of week indicator variables and a piecewise-linear and continuous function of temperature. We illustrated the prediction method using actual data from a commercial building. We extended prior research by presenting a method for estimating the statistical distribution of the prediction error and applied the method to actual data from 17 commercial buildings.

The work focuses on providing practical analytical tools and concepts for the M&V practitioner. A method to compute estimates of uncertainty in energy baselines is critical for practitioners and stakeholders to value the tradeoffs between data gathering, duration of pre- and post-ECM analyses, and expected energy savings. We see that simple regression models are likely to be preferred over more complex methods, in the near term, due to the ease of understanding and applicability. The use of such models further emphasizes the need to include uncertainty in estimations.

In the analysis of the 17 commercial buildings, we show that cross validation is suitable as a first approach to quantifying baseline uncertainty for M&V. In this analysis we show that a full year's worth of data to build baseline models (as prescribed by some IPMVP methods) does not necessarily improve the performance of the monthly or annual energy estimates. Moreover, for the buildings considered, uncertainty estimates are consistent with measured values, indicating the viability of the approach.

Future analysis should include other methods of estimating baseline uncertainty, testing using data from additional buildings, and further exploration of the tradeoffs between more complex regression models and the duration of the model training period.

5 ACKNOWLEDGEMENT

This work was performed at the Lawrence Berkeley National Laboratory, operated by the University of California, under DOE Contract DE-AC02-05CH1131. We gratefully acknowledge partial support for this work from LBNL's LDRD funds and from the California Energy Commission.

References

[1] International Performance Measurement and Verification Protocol Committee. International performance measurement and verification protocol. Technical Report DOE/GO-102002-1554, U.S. DOE, March 2002.

[2] Johanna L. Mathieu, Phillip N. Price, Sila Kiliccote, and Mary Ann Piette. Quantifying changes in building electricity use, with application to demand response. *IEEE Transactions on Smart Grid*, 2(3):507–518, September 2011.

[3] Katie Coughlin, Mary Ann Piette, Charles Goldman, and Sila Kiliccote. Statistical analysis of baseline load models for non-residential buildings. *Energy and Buildings*, 41(4):374–381, April 2009.

[4] Jessica Granderson, Mary Ann Piette, Girish Ghatikar, and Phillip N. Price. Building energy information systems: State of the technology and user case studies. Technical Report LBNL-2899E, Lawrence Berkeley National Laboratory, November 2009.

[5] David E. Claridge. A perspective on methods for analysis of measured energy data from commercial buildings. *Journal of Solar Energy Engineering*, 120(3):150–155, August 1998.

[6] James W. Taylor, Lilian M. de Menezes, and Patrick E. McSharry. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22(1):1–16, March 2006.

[7] John Kelly Kissock, T. Agami Reddy, and David E. Claridge. Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings. *Journal of Solar Energy Engineering*, 120(3):168–176, August 1998.

[8] Srinivas Katipamula, T. Agami Reddy, and David E. Claridge. Multivariate regression modeling. *Journal of Solar Energy Engineering*, 120(3):177–184, August 1998.

[9] Morris Hamburg. *Statistical Analysis for Decision Making*. Harcourt Brace Jovanovich, 3rd edition, 1983.

[10] M. Granger Morgan and Max Henrion. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analyses*. Cambridge University Press, 1992.

[11] Margaret F. Fels. PRISM: An introduction. *Energy and Buildings*, 9(1–2):5–18, May 1986.

[12] John Kelly Kissock and Carl Eger. Measuring industrial energy savings. *Applied Energy*, 85(5):347–361, May 2008.

[13] T. Agami Reddy, John Kelly Kissock, and D. K. Ruch. Uncertainty in baseline regression modeling and in determination of retrofit savings. *Journal of Solar Energy Engineering*, 120(3):185–192, August 1998.

[14] Srinivas Katipamula. Great energy predictor shootout II: Modeling energy use in large commercial buildings. *ASHRAE Transactions*, 102(2):397–404, 1996.