Disaggregation and Future Prediction of Monthly Residential Building Energy Use Data Using Localized Weather Data Network

Kristen S. Cetin, Iowa State University
Michael Siemann, Earth Networks, Inc
Christopher Sloop, Earth Networks, Inc

ABSTRACT

While smart meter and AMI infrastructure is installed in the majority of homes in the United States, in many states, the penetration remains at less than 15%. In these locations however, there is still a need to better inform residential customers of their energy use beyond the aggregated monthly energy use data that a traditional energy bill provides. Providing more informed insights to customers about their energy use, peer comparisons, and associated recommendations to reduce energy use has been shown to correlate with energy saving. This work investigates the use of monthly energy use data, combined with a localized weather network of highly granular weather data to disaggregate monthly energy data into end uses, including HVAC, baseload, and variable loads, and to predict future months' disaggregated energy use. This is accomplished through a methodology that uses 5-degree (F) binned, degree-time values to determine the percentage of the total energy use attributed to each end-use. This methodology then uses a simplified thermodynamic model of the building that determines the type of HVAC system in use, and predicts future energy use based on the future month's weather forecast. This methodology was verified using multiple datasets of over 400 homes in multiple climate zones. The results are used in a monthly scorecard provided to residential customers, which also includes targeted energy savings recommendations driven off peer comparisons.

Introduction

In the United States, buildings account for approximately 40% of energy use, over half of which is consumed by the residential building sector (US EIA 2015). Residential buildings are also responsible for over 37% of electricity consumed, and have also been found to contribute to over half of the peak electricity demands on the electric grid, particularly in warm climates (Wattles 2012). Residential energy use can be attributed to a number of different types of end use categories, including the heating ventilation and air conditioning system (HVAC), baseloads that are consistently in use over time, and variable loads that are more inconsistently used or vary in use based on the occupants' behaviors such as plug loads, and appliances that are intermittently used. Depending on the climate zone, often the largest portion of residential energy use is utilized by the HVAC system (Figure 1). The percentage of energy used by the HVAC depends on the climate conditions in which the residential building is located, ranging, on average, from less than 40% to over 50% annually. Baseloads that are utilized consistently month-to-month can include refrigerators and water heaters, and other consistently plugged in appliances and electronics. Of these, refrigerators and water heaters account for 5-7% and 16-23% respectively, of residential energy use (Figure 1). Finally, other/miscellaneous loads such as plug loads, lighting, etc, account for 25-38% of the total energy use (Figure 1). These loads may behave as baseloads or variable loads, depending on the occupants and home studied.

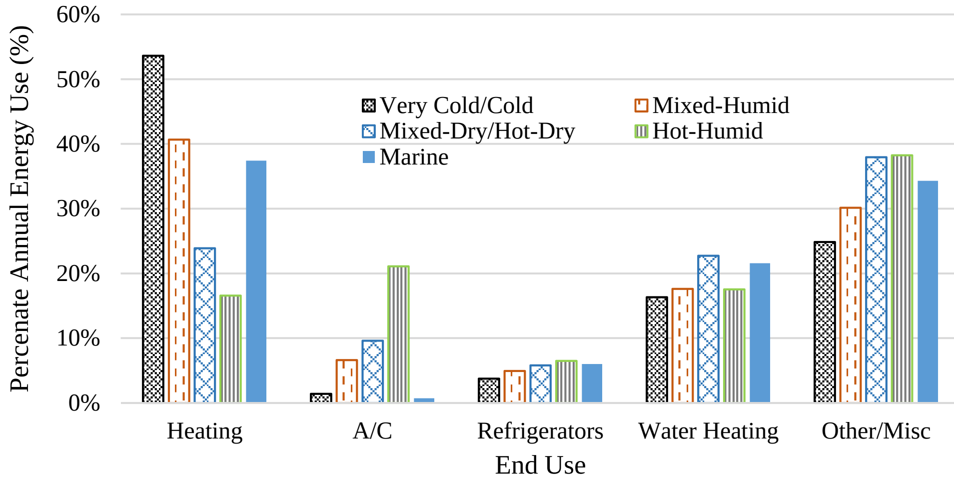

Figure 1: Annual site energy consumption (%) of residential buildings by climate zone and end use (US EIA 2009)

In recent years, an increasing focus on energy efficiency in residential buildings has been spurred by a number of factors, including concerns regarding climate change and greenhouse gas emissions (IPCC 2015), increasing demands for electricity and energy (US EIA 2015), and the need for improved electric grid reliability. To achieve energy efficiency targets for residential buildings, a number of different strategies have been utilized. Currently, approximately half of all residential buildings have Advanced Metering Infrastructure (AMI)/smart meters (IEI 2014). The implementation of smart meters enables utilization of data-driven methodologies to develop energy use insights determined from the more granular smart meter-provided 15-minute to hourly whole home energy use information. In many states smart meters have been deployed in the majority of residential buildings, however, in other states including many of the Midwestern states, the amount of residential buildings with smart meters is limited, and remains below 15% penetration (IEI 2014). However, even for residential buildings without smart meters, encouraging energy efficiency remains an important goal.

For those homes without smart meters, the method of manual energy use meter reading on a monthly basis has been a standard practice throughout the United States for many years. This manual reading provides the monthly electricity (kWh/month) and/or gas (ccf/month or therms/month) use values given on utility billing statements. Similar to insights developed from smart meter data, residential customers with monthly data can also benefit from additional insights on how energy is used in their home, and to motivate them to reduce energy use.

Previous studies have found that providing energy use information and feedback to the consumer can achieve up to 10-20% reduction in energy use, as summarized in Faruqui et al. (2009) and Ehrhardt-Martinez et al. (2010). "Indirect" feedback programs, which provide energy consumption information to the consumer after the energy is consumed, such as in a monthly, weekly or daily statement showing whole-home energy use and feedback on what can be done to reduce energy loads, have achieved energy savings of up to 8.4%, as summarized in Armel et al. (2013). The greatest energy savings has been shown to occur if submetered data is available to consumers (Neenan and Robinson 2009). Unlike whole-home data, disaggregated data allows for isolation of individual appliances and high energy users. The most important reason cited for this increase in energy reduction is due to the enhanced ability to provide personalized recommendations for energy reduction strategies (Armel et al 2013). Peer comparison of energy use between similar homes has also showed significant savings (e.g. Ayres et al. 2009).

Recently, data-driven building energy use data analysis has focused on grey-box and black-box modeling techniques in which an inverse model of each individual building is created (Dong et al 2015, Cross et al 2014, Kim et al 2015, Hallinan et al 2011). However, this family of models, sometimes classified as change point models, typically only consider whole-home energy use prediction and do not consider the development of insights of disaggregated energy use by end use, or future prediction of energy use of residential buildings. Thus, in this research, we present a new methodology that employs monthly residential electricity billing data and highly granular historic and forecasted weather data from a localized weather network of weather stations to disaggregate monthly electricity data into three end use categories, and to predict the future billing month's whole home energy use. Utilizing these insights, the information is presented in a digital scorecard report format that is sent to residential customers. This scorecard also includes energy savings tips and provides a peer comparison of a particular residential building to its neighbors.

Methodology & Results

The proposed methodology utilizes weather data and electricity use billing data as inputs, and ultimately outputs a scorecard report that is provided to residential customers. In between, a simplified grey-box inverse model is implemented which uses the weather data and electricity billing data to disaggregate the monthly energy use into end uses, and to predict electricity use for the future billing month. Figure 2 summarizes the inputs and outputs of the proposed model and a summary of the steps of the methodology's framework. Each of these steps in the framework are discussed in further detail in this section, including the (1) inputs, (2) model, (3) outputs, and (4) final product.

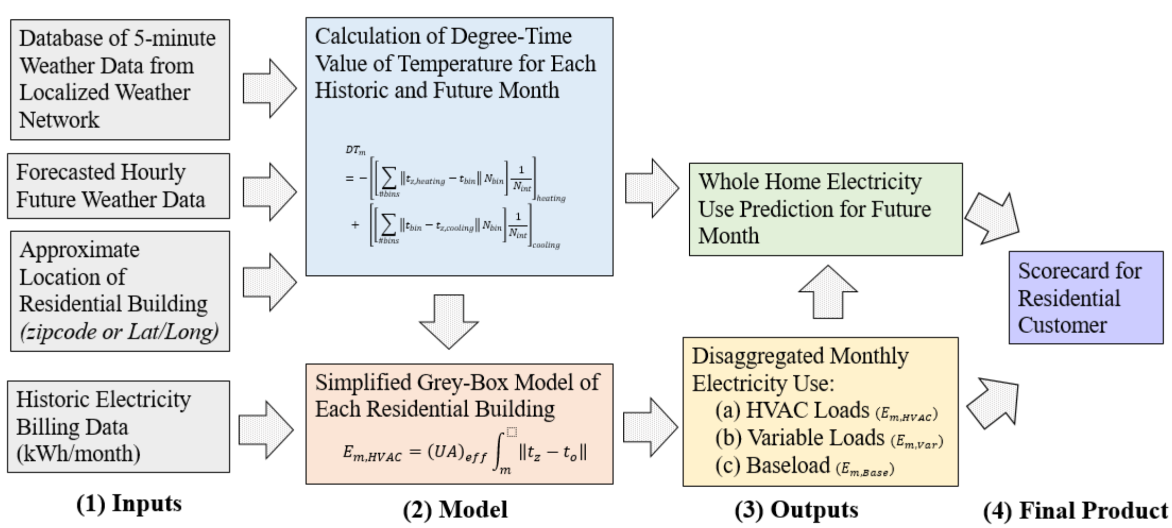

Figure 2: Flowchart diagram of disaggregation and future monthly energy use prediction methodology

(1) Inputs

The required data input for this methodology includes (a) weather data (outdoor temperature), (b) forecasted hourly future weather data for the month following the current billing cycle (outdoor temperature), (c) the approximate location of the residential building of which a model is being developed, and (d) historical electricity use billing data for the residential building being considered.

The weather data used in this research is provided by matching the house location to the closest Earth Networks WeatherBug weather station (Earth Networks 2014). The average distance between each house and weather station is 5.9 km, with the house and the weather station often found within the same zipcode. These weather data are gathered and stored in either 5-minute or hourly intervals. The weather network measurements of temperature are made using crystal oscillator thermometers.

The utility electricity billing data input is the meter reading of the monthly energy use provided by the electric utility company. This value is in kWh of electricity use per month. As residential customers sign up for electricity billing at different times, the billing cycle of each individual residential buildings varies in terms of start and end date of each month.

(2) Thermodynamics-Based Model

Utilizing these inputs, a simplified thermodynamics-based model is used as a basis for modeling of residential energy use due to heating and cooling (HVAC). The other two end use loads (baseloads and variable loads) are also determined, and are discussed in Step 3a below. The model utilized in this research is a modified version of the *Bin Method* (Kuhen et al 2001), which has several advantages over the more commonly used degree-day methodology, including the ability to account for a range or multiple values of indoor temperature setpoints. Steady-state conditions are assumed in which the most influential factor in the determination of the energy use of the residential building is driven by conduction. Assumptions of this method include that internal gains and other heat gains, do not significantly affect energy use in comparison to those due to conduction. Given the long timescale, thermal mass is also assumed to be negligible. These assumptions are generally better for residential buildings, in which the energy use is more governed by the outdoor temperature alone, as compared to commercial buildings which are more strongly influences by HVAC type and control strategy (ASHRAE 2013). Residential buildings are also generally light-frame construction, making their thermal mass less significant than that of commercial buildings. .

Equations 1a-c describe the thermodynamics-based relationship utilized for this model. The amount of heating/cooling energy use required, $E_{m,HVAC}$, over a month, *m* (Eq. 1a), is determined by integrating the net heat transferred over the month time period, where the double vertical lines indicate that only the positive values are included in the integral. The integral term is called the *degree-time* value, DT_m, for month *m*, as shown in Eq. 1b, and is approximated using Eq. 1c. In these equations, *E* is the energy use per month *m* due to heating and/or cooling, $(UA)_{eff}$ is the effective heat transfer value for the building, t_z is the zero-load indoor temperature at which the heat gains in the building are assumed to equal to the heat losses, and t_o is the outdoor temperature values over the period of time studied (1 month). t_{bin} is the average temperature for each bin of temperatures, N_{bin} is total the number of bins of temperature, and N_{int} is total the count of the temperature values from the weather data input over the month, *m*, being evaluated.

For the approximation of the integral value in Eq. 1a, the *degree-time* value indicated by Eq. 1c is used, where outdoor temperatures occurring over the time interval of interest are divided into bins of 5-degree F increments of temperature, ranging from the lowest to highest temperature data value available in each of the months considered. For example the number of time intervals in which the outdoor temperature falls between 49.9°F and 54.9°F would be included in the 50/55°F bin, with a mean value of t_{bin}= 52.5°F. This equation includes the heating degree-time value: $-\left[\left[\sum_{\#bins} \|t_{z,heating} - t_{bin}\| N_{bin}\right] \frac{1}{N_{int}}\right]_{heating}$ and the cooling degree-time value: $\left[\left[\sum_{\#bins} \|t_{bin} - t_{z,cooling}\| N_{bin}\right] \frac{1}{N_{int}}\right]_{cooling}$, which are summed together to achieve a total degree-time value. If the total degree-time value is less than zero, this indicates that the month is a majority heating month, and if it is greater than zero, this indicates the month is a majority cooling month. The closer to zero the degree-time value, the less heating and/or cooling should occur during that time period.

$$E_{m,HVAC} = (UA)_{eff} \int \|t_z - t_o\| \; d\tau \quad \text{Eq. 1a}$$

$$E_{m,HVAC} = (UA)_{eff} \; DT_m \quad \text{Eq. 1b}$$

$$DT_m = -\left[\left[\sum_{\#bins} \|t_{z,heating} - t_{bin}\| N_{bin}\right] \frac{1}{N_{int}}\right]_{heating} + \left[\left[\sum_{\#bins} \|t_{bin} - t_{z,cooling}\| N_{bin}\right] \frac{1}{N_{int}}\right]_{cooling} \quad \text{Eq. 1c}$$

The zero-load temperatures for heating, $t_{z,heating}$, and cooling, $t_{z,cooling}$, in Eq. 1c are assumed to be equal to 23.8°C (75°F) for cooling, and 12.8°C (55°F) for heating based on analysis of the ELCAP dataset, as discussed in Pratt et al (1993), of residential building energy end uses which breaks down heating and cooling energy into separate end-use loads. As shown in Figure 3a, the cooling and heating degree-time values simultaneously occur within the monthly average temperatures of approximately 23.8°C (75°F) to 12.8°C (55°F). The buildings from which data was utilized to test this model did not have connected thermostats in which the value of the setpoints could be verified, however this is a subject of ongoing research that would enable the use of a range of setpoint values to be entered into this model.

To implement this model, the information needed to determine the binned temperature values DT_m is calculated, and monthly energy use values, E_m, are provided as inputs. Thus the next step is to determine the relationship between these two values to determine the model relationship that represents HVAC energy use, and how this contributes to the total energy use relative to the variable loads and baseloads (disaggregation).

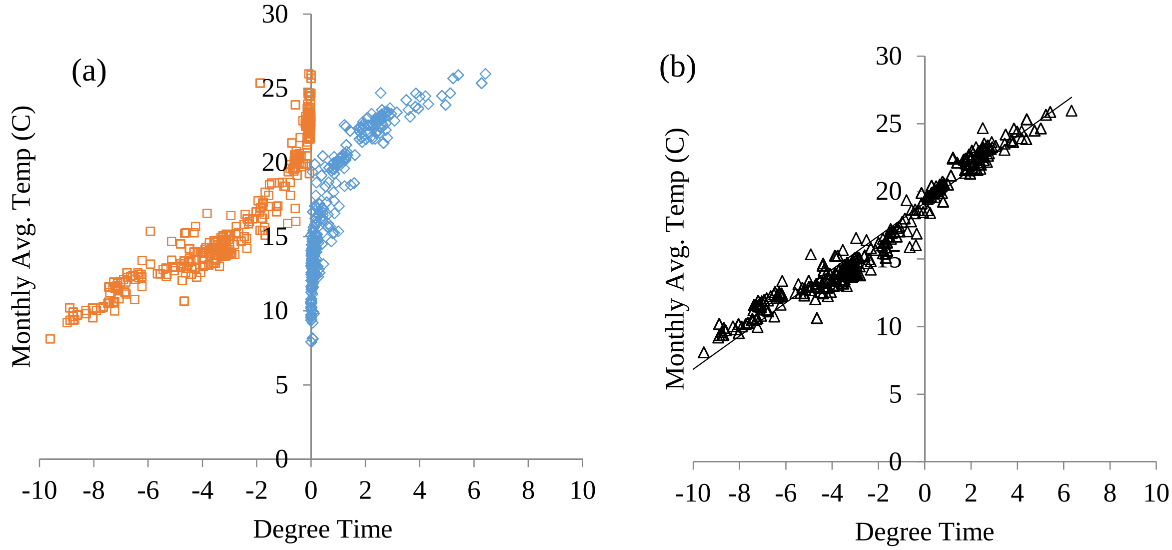

Figure 3: The (a) heating (orange) and cooling (blue) degree-time values and (b) the total degree-time values (black) compared to the average monthly temperatures using the ECLAP dataset, referred to in Pratt et al (1993).

(3a) Disaggregation of HVAC, Variable and Baseload Loads

To disaggregate the monthly energy use into HVAC loads, variable loads and baseloads, the procedure outlined in Figure 4 is followed. This methodology was developed based on the results of analysis of several large datasets of residential energy use data as described below. The steps are summarized as follows:

- *Inputs:* The degree-time value calculated from the previous step and the previous months' electricity billing data.
- *Step i:* Determine the month with the degree-time value closest to zero: $[min\ \{|DT_m|\}]$. Of all of the months considered, the month with the degree-time value closets to zero should have the lowest percentage of total energy use attributed to the HVAC system.
- *Step ii:* Determine the HVAC, and variable and base loads for this low degree-time month $\left(E_{m,\ HVAC}\right)_{min\{|DT_m|\}}$ and $\left(E_{m,Var+Base}\right)_{min\{|DT_m|\}}$, respectively, based on the percentage of energy use in the lowest energy use month from the HVAC, P_{HVAC}, a value determined based on the residential energy use dataset analysis conducted.
- *Step iii:* Determine the HVAC, variable and base loads for each of the months of billing data available, $\left(E_{m,\ HVAC}\right)$ and $\left(E_{m,Var+Base}\right)$ respectively. This final step assumes the baseload varies over time based on the coefficient of variation of the variable and baseloads, $COV_{Var+Base}$, determined from the analysis of several large datasets of residential energy use data. R is a uniformly distributed random variable from 0 to 1.
- *Outputs:* Monthly energy use from the HVAC, variable and baseloads for each month.

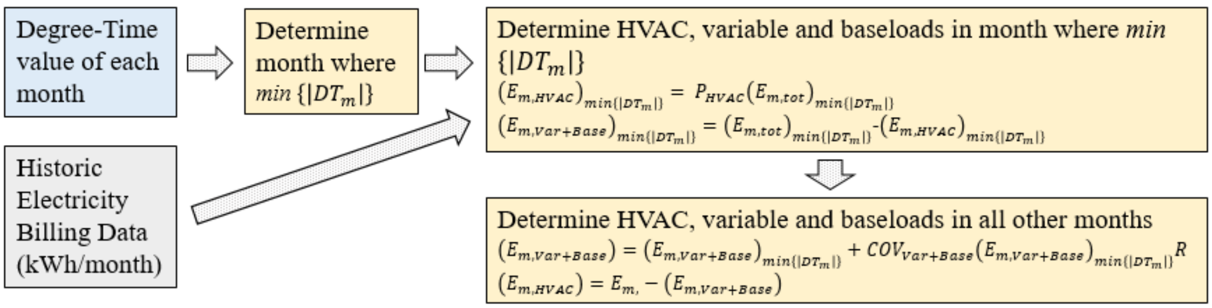

Figure 4: Methodology for disaggregation of energy use into HVAC, variable and baseloads.

For this methodology, the values of P_{HVAC}, and $COV_{Var+Base}$ are determined based on analysis conducted on a total of 645 homes of disaggregated energy use data, including 161 single family residential buildings in Texas, and 484 residential buildings in the Pacific Northwest. Details on the collection of the data in these datasets and their accuracy are discussed in Cetin and Novoselac (2015) and Pratt et al (1993), respectively. For each dataset analyzed, the whole-home energy use and HVAC energy use (including both the indoor air handling unit and the outdoor compressor/condenser) were recorded and binned into monthly periods by the calendar month. HVAC energy use was subtracted from whole-home energy use to determine the total non-HVAC loads.The non-HVAC energy use values, which include both variable loads and baseloads, were found to be fairly constant in total throughout a year-long period for both datasets. Figure 5 shows the normalized monthly energy use over a year-long period for both datasets analyzed, created using the methodology discussed in Cetin et al (2014). The average normalized monthly energy use is shown in red, and the normalized monthly energy use for each house in the dataset (Figure 5a) and each climate zone (Figure 5b) are indicated using solid lines. In both cases this average normalized monthly energy use value ranges from slightly less to slightly more than 1 (0.85-1.15).

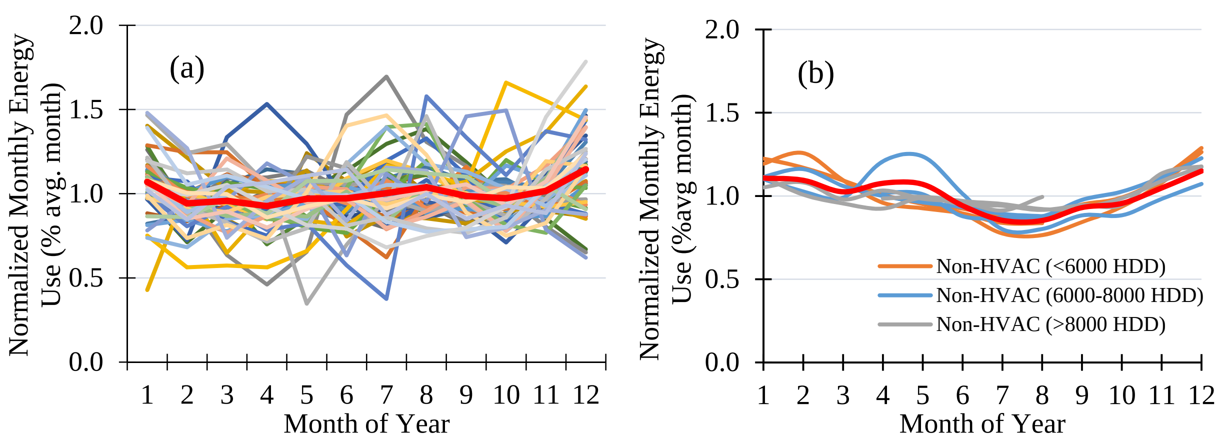

Figure 5: Normalized monthly energy use of non-HVAC loads in single family residential buildings in (a) Texas (n=161) where each line represents a home in the dataset, and the red line represents the average, and (b) the Pacific Northwest (n=484), where each line represents the average of the homes in a particular climate zone defined by the number of heating degree days (HDD) and the red line represents the average.

The coefficient of variation (COV) of the non-HVAC use was approximately 0.12 overall in both residential building datasets. This fairly constant non-HVAC load throughout a year-long period which includes all seasons and transition months is also consistent with assumptions made for residential energy modeling software based on the Building America House Simulation Protocol and Building America Analysis Spreadsheets (Wilson et al 2014). Based on this information, it is assumed that the sum of the variable and baseloads are fairly consistent on a monthly basis with a variation in values of approximately 0.12 ($COV_{Var+Base}$= 0.12) from month to month, and multiplied by the average value of the variable and baseloads $\left(E_{m,Var+Base}\right)_{min\{|DT_m|\}}$.

To determine the split of HVAC and variable/baseload values (P_{HVAC}), the degree-time value determine in Step (2) was used, where, as discussed previously, the months with the lowest heating and cooling loads are those in which the degree-time value is closest to zero. Thus for this methodology, the month with the lowest energy use was used to determine this split. From analysis of the two previously discussed datasets, the split of HVAC/non-HVAC loads during the transition months in which the lowest HVAC use occurs, is on average, approximately 80% non-HVAC and the remainder being HVAC (P_{HVAC} = 0.2). This is demonstrated in Figure 6 for both datasets, which shows a histogram of the percentage non-HVAC use in the lowest degree-time months.

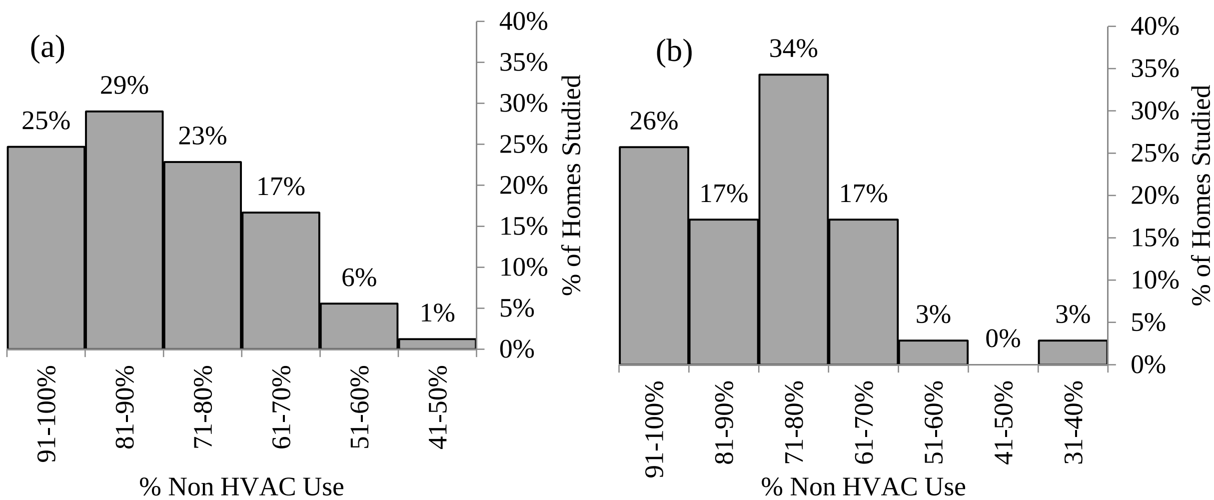

Figure 6: Percent of monthly energy use from non-HVAC end uses during the month with the lowest degree-time values in single family residential buildings in (a) Texas (n=161), and (b) the Pacific Northwest (n=484).

A test of the accuracy of this methodology for disaggregation of monthly energy use by end use found that on average, the prediction of the disaggregated HVAC load was less than 18% error (Figure 7). The red lines in Figure 7 indicate a +/- 20% error. For homes with larger numbers of months of previous weather data, the error in general was found to be lower.

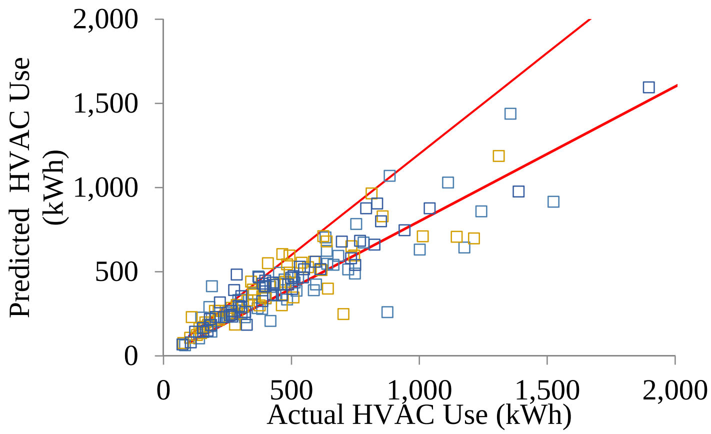

Figure 7: Comparison of the actual HVAC use and the predicted HVAC use each month using the proposed modeling methodology.

(3b) Future Monthly Energy Use Prediction

To predict the future whole-home monthly energy use of each residential building considered, the procedure outlined in Figure 8 was utilized. Based on previous literature of change-point models for prediction of energy use of buildings, the relationship between the outdoor temperature-based degree-time value and the monthly HVAC energy use, as described in Eq. 1b, is assumed to be linear (ASHRAE 2013). The steps are summarized as follows:

- *Inputs:* The degree-time values for each previous month, the disaggregated energy use data from Step 3a, and the forecasted future weather data.
- *Step i:* Divide the months of disaggregated energy use data and degree-time values into heating months and cooling months.
- *Step ii:* Determine the degree-time value of the future month (Eq. 1c).
- *Step iii:* Determine if the future month will be a heating or cooling month. If the degree-time value for the future month is greater than 0, the month will be a cooling month; if the degree-time value is less than 0, the future month will be a heating month.
- *Step iv:* Using the previous months' HVAC energy use and degree-time values for cooling (if the future month is a cooling month), or heating (if the future month is a heating month) months, use linear regression analysis to determine the slope, and intercept of the relationship between the degree-time (independent variable) and HVAC energy use (dependent variable).
- *Step iv:* Calculate the future month's energy use: Use the linear relationship to determine the future month's HVAC use based on the future month's degree-time value; Sum the future month's energy use and the variable loads and baseloads from the previous month.
- *Output:* Whole-home energy use for the future month.

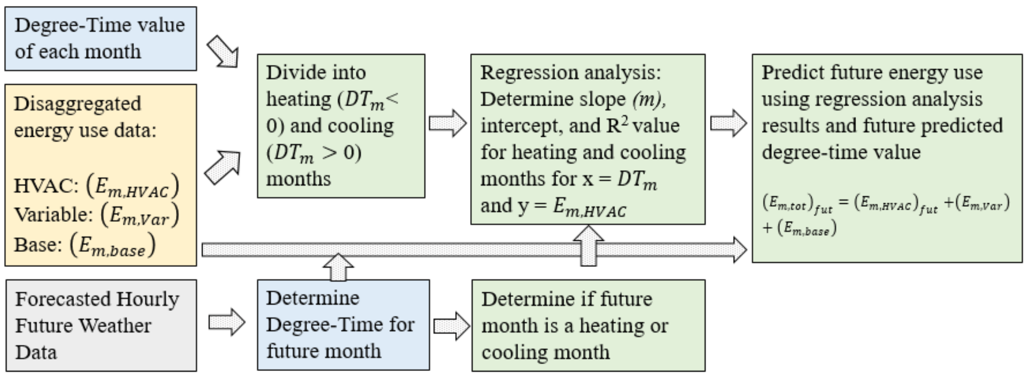

Figure 8: Methodology for determining future whole-home energy use.

This methodology was implemented for 3050 residential buildings for three different future forecasted months, including November, December and January. The accuracy of the future energy use forecast is dependent on both the accuracy of the model and the future weather forecast. For the three months tested, the average error of the whole-home energy use was -2.7%. A parity plot of the actual versus predicted use and a histogram of the future month energy use forecast error are shown in Figure 9a and 9b respectively. The standard deviation is 28.9%. 67% of all homes' future month's forecasted energy use was within 20% error. As the future monthly energy use was measured only on the whole-home level, the accuracy of the individual disaggregated end uses is not compared.

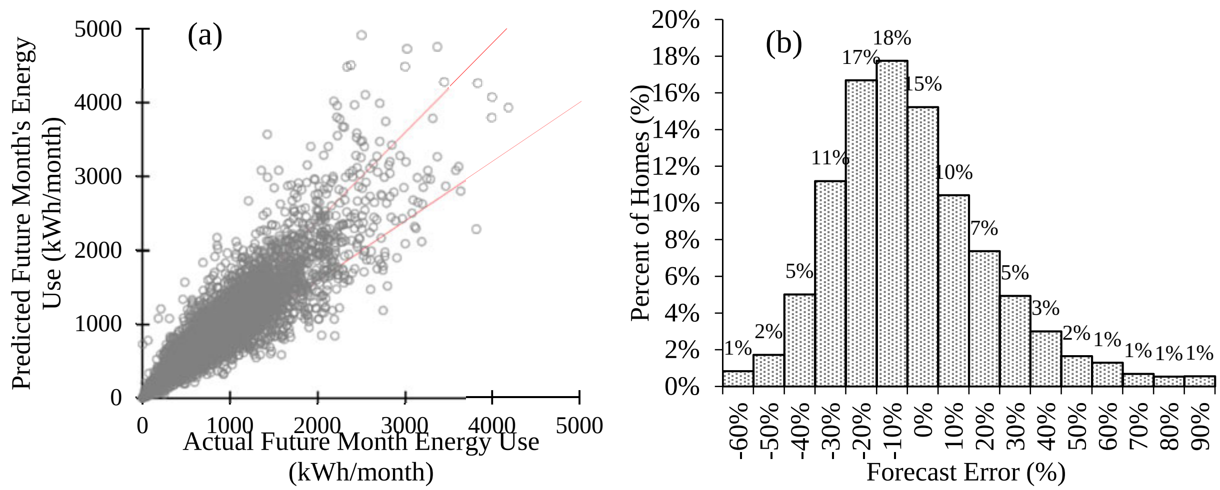

Figure 9: (a) Comparison of the actual future month's energy use compared to the model predicted future month's energy use and (b) a histogram of the future energy use forecast error

Final Product

The results of the disaggregation and future month's energy use are utilized in a scorecard format which is emailed to the residential customer. This scorecard contains the whole-home energy use of the residential building studied, as compared to an average home and an energy efficient home of similar size in the same NOAA climate region (NOAA 2016). It also contains the disaggregated monthly energy use data and a value of the predicted increase or decrease in energy use for the future month, derived from the future month's energy use prediction. The scorecard also provides energy savings tips relative to the season and residential customer's energy use patterns. An example of a scorecard is shown in Figure 10.

Figure 10: Example Scorecard provided to residential customers, including whole-home energy use, disaggregated energy use, and future predicted energy use

Discussion and Conclusions

In order to provide additional insights regarding the monthly electricity use of residential building consumers, a methodology was developed to both disaggregate monthly energy use into three different end uses including HVAC, variable and baseloads, and to predict the electricity use of the future month. Given the limited energy use information available to develop the model to determine disaggregation and future energy use, the results of this research indicate that the disaggregation and future energy use prediction algorithms provide reasonable results and level of accuracy, including an average of less than 18% error for disaggregation, and an average of 2.7% error for future energy use prediction. Further testing of the proposed model using residential sub-metered data would be beneficial for further refinement of these models.

The outputs of this model, along with peer comparison of neighboring homes and energy savings tips, are used in a scorecard format provided to the residential energy consumer. The scorecard output provides additional insights for residential customers beyond the typical whole-home monthly energy use information typically provided on a utility bill, along with peer comparison of neighboring homes and energy savings tips, with the ultimate goal of encouraging energy efficient behaviors for the residential building occupants.

References

ASHRAE (2013). ASHRAE Fundamentals Handbook. Chapter 19: Energy Estimating and Modeling Methods. ASHRAE, Atlanta, GA.

Armel KC, Gupta A, Shirmali G, Albert A. (2013). Is disaggregation the holy grail of energy efficiency? The case of electricity. Energy Policy. 52, 213-234.

Ayres I, Raseman S, Shih A. (2009). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage (No. w15386). National Bureau of Economic Research.

Cetin KS, Novoselac A. (2015). Single and multi-family residential central all-air HVAC system operational characteristics in cooling-dominated climate. Energy and Buildings, 96, 210-220.

Cetin KS, Tabares-Velasco PC, Novoselac A. (2014). Appliance daily energy use in new residential buildings: Use profiles and variation in time-of-use. Energy and Buildings, 84, 716-726.

Cross A. (2014). Development of a Black-Box Transient Thermal Model for Residential Buildings. PhD Dissertation. University of Nevada, Las Vegas.

Dong B, Li Z, Rahman SM, Vega R. (2015). A hybrid model approach for forecasting future residential electricity consumption. Energy and Buildings. In Press doi:10.1016/j.enbuild.2015.09.033

Earth Networks (2014). Earth Networks Weather Station Installation, Operation and Maintenance Manuel. http://download.aws.com/manuals/RedBugBoxInstall.pdf

Ehrhardt-Martinez K, Donnelly KA, Laitner S. (2010). Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities. Washington, DC: American Council for an Energy-Efficient Economy.

Faruqui A, Sergici S, Sharif A. (2010). The impact of informational feedback on energy consumption—A survey of the experimental evidence. Energy, 35(4), 1598-1608.

Hallinan, K.P., et al. (2011). Establishing building recommissioning priorities and potential energy savings from utility energy data. ASHRAE Transactions. 117(2)

Institute for Electric Innovation (2014). Utility-scale smart meter deployments: building block of the evolving power grid. The Edison Foundation.

International Panel on Climate Change (2013). Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York, NY.

Kim KH, Haberl JS. (2015). Development of methodology for calibrated simulation in single-family residential buildings using three-parameter change-point regression model. Energy and Buildings, 99, 140-152.

Kuehn TH, Ramsey WW, Threlkeld JL. (2001). Thermal Environmental Engineering, Prentice Hall, New Jersey, USA.

Neenan B, Robinson J, Boisvert RN. (2009). Residential electricity use feedback: A research synthesis and economic framework. Retrieved October, 26(2011), 3.

NOAA (2016). U.S. Climate Divisions. http://www.ncdc.noaa.gov

Pratt RG, Conner CC, Cooke BA, Richman EE. (1993). Metered end-use consumption and load shapes from the ELCAP residential sample of existing homes in the Pacific NW. Energy and Buildings. 19, 179-193.

United States Energy Information Administration. (2009). RECS, Residential Energy Consumption Survey, U.S. Department of Energy, Washington D.C., U.S.A.

United States Energy Information Administration. (2015). Energy Factbook-Energy Use Source. U.S. Department of Energy, Washington D.C., U.S.A.

Wattles P. (2012). Demand Response and ERCOT Grid Reliability. Energy Efficiency Forum – Arlington, TX May 23, 2012.

Wilson E, Engebrecht Metzger C, Horowitz S, Hendron R. (2014). 2014 Building America House Simulation Protocols. National Renewable Energy Laboratory.