PNL-SA- 17700

PURIFYING MIXED-USE ELECTRICAL CONSUMPTION DATA

Z. Todd Taylor and Robert G. Pratt
Pacific Northwest Laboratory

ABSTRACT

This paper describes several analytical techniques for obtaining pure end-use load information from mixed end-use consumption data. This process is frequently necessary to make metered data useful to those involved in electric utility load forecasting and conservation assessment. Analyses based on traditional thermal models can be greatly augmented by these data sets if the measured entities correspond to those for which modeled estimates are necessary.

We present two scenarios in which greater end-use resolution was needed than was available in existing data. The first involves segregating measured total HVAC consumption data into its heating, cooling, and ventilation constituents. The second discusses a technique to separate measurements of mixed equipment consumption into equipment type categories. These techniques were successfully applied to a large number of metered commercial buildings. We conclude with suggestions for extending these techniques to applications involving high-time-resolution building total data.

INTRODUCTION

Recent advances in metering hardware have resulted in a proliferation of end-use metering projects among electric utilities. However, the use of such data in the analysis of many individual end uses of electricity is hampered because of the lack of detail available in the data. Appropriate analyses of the data require that measured end uses of electricity correspond to those used in existing economic and thermal models. However, the collection of finely segregated end-use data is often cost-prohibitive. The detail of end-use data is often limited by the end-use resolution available in the electric distribution panels. Packaged HVAC units, for example, are often served by a single electric circuit although they provide three distinct functions--heating, cooling, and ventilation. Segregating the three functions in the metering hardware requires taking measurements within the sealed enclosure of the systems (which may void manufacturers' warranties) or taking additional ancillary measurements such as supply air temperatures. Analogous situations exist when lighting and plug loads are served by a single circuit.

This paper discusses techniques that have been used to increase the end-use resolution of existing mixed-use data. Two scenarios are examined: (1) high-time-resolution (hourly) mixed HVAC data and (2) low-time-resolution (monthly) mixed lighting and equipment data. We conclude with a cursory discussion of the potential for extending these techniques to disaggregate high time resolution building total data, such as that collected under the requirements of the Public Utility Regulatory Policy Act of 1978 (PURPA 1980). While this is undoubtedly the most prolific type of mixed-use data available, it is also the most problematic to segregate.

The disaggregation techniques discussed here were developed in response to specific analysis requirements of data collected under the Bonneville Power

Administration's (BPA) End-Use Load and Consumer Assessment Program (ELCAP). The techniques were successful in obtaining reasonable estimates of pure end-use consumption from the mixed end-use data (Taylor and Pratt 1989).

HIGH-TIME-RESOLUTION MIXED HVAC DATA

The technique for segregating mixed HVAC load data into its heating, cooling, and auxiliary components involves obtaining an empirical model of the mixed data as a function of outdoor temperature. That model is then used to "predict" the segregated end uses. The actual (measured) HVAC energy is then partitioned according to the proportions of the predicted HVAC energy attributable to heating, cooling, and ventilation/auxiliary functions.

Description of the Technique

Figure 1 illustrates the process of obtaining the empirical relationship. For a given building, the average daily consumption for the mixed HVAC end use is plotted against the average daily outdoor temperature. Superimposed on the scatter plot is a curve representing a statistical best fit to the data. The curve was obtained using a nonparametric curve-fitting procedure known as lowess (locally-weighted regression scatter plot smoothing). Lowess is a robust (resistant to outliers) curve-fitting technique that involves a series of linear regressions to segments of the data.

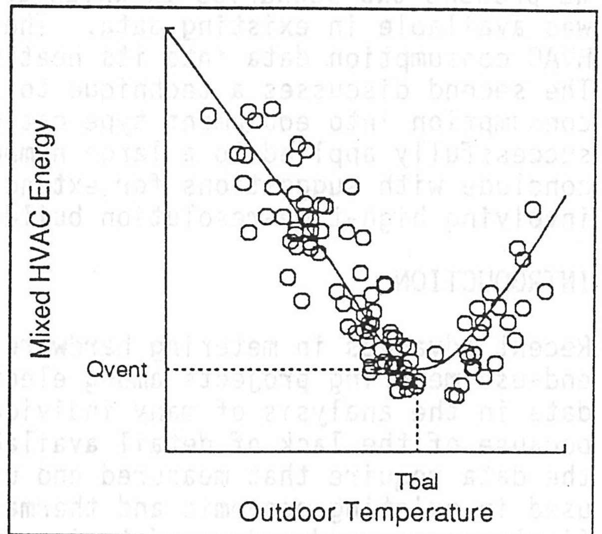

Figure 1 - Illustration of Lowess Fit

If it is assumed that the lowess curve adequately represents the behavior of the mixed HVAC end use, then the disaggregation of the data is straightforward. The point on the abscissa at which the lowess curve reaches its minimum value (T_{bal}) represents the "balance" temperature at which the building changes from heating mode to cooling mode. The ordinal coordinate of this point (Q_{vent}) can be assumed to represent the base system load when neither heating nor cooling is required. This is presumably a combination of ventilation fans, control circuits, and any other auxiliary equipment such as crankcase heaters. Ventilation and auxiliary energy is assumed to be constant at the value of Q_{vent}. On any day during which the average outdoor temperature exceeds the balance point, all consumption in excess of Q_{vent} is assigned to the cooling end use. On days when the temperature is below T_{bal}, the energy is assigned to heating.

This procedure is straightforward and simple, but it does have some drawbacks. First, notice the scatter plot around the minimum of the lowess curve. Clearly, the energy consumption on some days was well below the assumed baseline ventilation rate. On other days it was higher. Second, notice that the region of the lowess curve that contains the minimum is somewhat flat, indicating that there is a dead band on the temperature scale in which neither heating nor cooling is required. An understanding of typical commercial building operations can explain some of these phenomena. The HVAC systems of many buildings operate on strict schedules, lowering thermostat set points and/or reducing ventilation at night. Cool nighttime temperatures combined with a thermostat setback can result in large heating loads in the early morning hours, even in warmer months when cooling may be required in the afternoons. Days during which both heating and cooling are required contribute to the wide scatter around the minimum of the lowess curve. The net impact on the disaggregated loads is that ventilation and auxiliary loads tend to be overestimated.

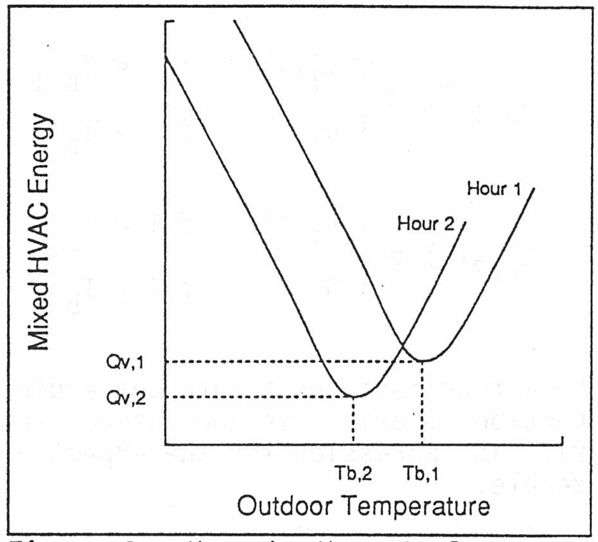

Figure 2 - Hour-by-Hour Analysis of Mixed HVAC Loads

To minimize the errors associated with days of dual HVAC modes, the process described above is modified in two ways. First, each day is classified as either a work day or a nonwork day and a separate analysis is conducted for each group of days. This eliminates scatter resulting from different schedules of operation. Second, to account for the hour-to-hour schedule of each building, hourly loads are analyzed instead of simple daily averages. By fitting a separate lowess curve to each hour's data, the changes in building operation can be captured. This is illustrated in Figure 2, which shows two separate lowess curves, one for each of two sample hours. Notice that the abscissa is again *daily average* outdoor temperature, even though each line represents the average of a single hour. The use of daily temperature data to correlate hourly loads is based primarily on considerations of computational efficiency. However, using daily temperatures also better captures true weather effects in buildings with significant thermal mass, as nighttime loads are sometimes strongly dependent on temperatures earlier in the day.

Thus, a lowess curve that relates the mixed HVAC response to daily average temperature is developed for each hour, and this set of 24 curves is developed for each of the two day types (work days and nonwork days, if present.) The process of disaggregating the mixed load into its components is essentially identical to that described above, but involves a few additional steps. Consider the example in Figure 2, which represents a fictitious day composed of only two hours (to simplify the explanation.) Suppose hour 1 represents a daytime hour and hour 2 represents nighttime. The lowess curve for hour 1 has a minimum value ($Q_{v,1}$) that represents the daytime ventilation load. The curve for hour 2 has a minimum ($Q_{v,2}$) that represents nighttime ventilation loads. In the example, nighttime ventilation loads are lower than daytime loads. Also, the lower ventilation load, possibly combined with a thermostat setback strategy, results in a lower balance temperature for the nighttime loads. This is illustrated by the points marked $T_{b,1}$ and $T_{b,2}$.

To explain the disaggregation process, the following functions are defined:

$L_1(T)$ ≡ hour 1 HVAC load at daily temperature T

$L_2(T)$ ≡ hour 2 HVAC load at daily temperature T

$L_1'(T) \equiv L_1(T) - Q_{v,1}$ *(the mixed HVAC load above the minimum)*

$L_2'(T) \equiv L_2(T) - Q_{v,2}$ *(the mixed HVAC load above the minimum)*

$$L_{h,1}(T) \equiv \begin{cases} L_1'(T), & \text{if } T < T_{b,1} \\ 0, & \text{if } T \geq T_{b,1} \end{cases} \qquad L_{c,1}(T) \equiv \begin{cases} 0, & \text{if } T \leq T_{b,1} \\ L_1'(T), & \text{if } T > T_{b,1} \end{cases}$$

$$L_{h,2}(T) \equiv \begin{cases} L_2'(T), & \text{if } T < T_{b,2} \\ 0, & \text{if } T \geq T_{b,2} \end{cases} \qquad L_{c,2}(T) \equiv \begin{cases} 0, & \text{if } T \leq T_{b,2} \\ L_2'(T), & \text{if } T > T_{b,2} \end{cases}$$

These functions constitute the empirical model of the mixed HVAC end use as a function of daily average outdoor temperature. Using these functions, we can write an expression for the expected mixed HVAC load for the 2-hour day in our example:

$$\begin{aligned} Q_{mix}(T) &= L_1(T) + L_2(T) \\ &= Q_{v,1} + L_1'(T) + Q_{v,2} + L_2'(T) \\ &= Q_{v,1} + L_{h,1}(T) + L_{c,1}(T) + Q_{v,2} + L_{h,2}(T) + L_{c,2}(T) \qquad (1) \end{aligned}$$

where $Q_{mix}(T)$ = the daily total mixed HVAC load when the daily average outdoor temperature is T.

Equation (1) contains all the information needed to estimate the heating, cooling, and ventilation/auxiliary fractions of the measured mixed HVAC end use. Rearranging that equation helps to visualize the process:

$$\begin{aligned} Q_{mix}(T) &= [Q_{v,1} + Q_{v,2}] && \textit{(ventilation \& auxiliaries)} \\ &+ [L_{h,1}(T) + L_{h,2}(T)] && \textit{(heating)} \\ &+ [L_{c,1}(T) + L_{c,2}(T)] && \textit{(cooling)} \qquad (2) \end{aligned}$$

Using Equation (2), the measured daily mixed HVAC loads can be separated into heating, cooling, and ventilation/auxiliary components. This is done with the ratios indicated by that equation:

$$\text{HeatFraction} = \frac{[L_{h,1}(T) + L_{h,2}(T)]}{Q_{mix}(T)} \tag{3}$$

$$\text{CoolFraction} = \frac{[L_{c,1}(T) + L_{c,2}(T)]}{Q_{mix}(T)} \tag{4}$$

$$\text{VentFraction} = \frac{[Q_{v,1} + Q_{v,2}]}{Q_{mix}(T)} \tag{5}$$

Finally, the actual equations used to disaggregate the measured mixed HVAC end use involve summations over all 24 hours:

$$Q_{mix}(T) = \sum_{i=1}^{24} \left\{ Q_{v,i} + L_{h,i}(T) + L_{c,i}(T) \right\} \tag{6}$$

$$\text{HeatFraction} = \frac{\sum_{i=1}^{24} L_{h,i}(T)}{Q_{mix}(T)} \tag{7}$$

$$\text{CoolFraction} = \frac{\sum_{i=1}^{24} L_{c,i}(T)}{Q_{mix}(T)} \tag{8}$$

$$\text{VentFraction} = \frac{\sum_{i=1}^{24} Q_{v,i}}{Q_{mix}(T)} \tag{9}$$

Equations (6) through (9) are applied to the measured daily mixed HVAC data to obtain the disaggregated heating, cooling, and ventilation/auxiliary end uses. The resulting daily data can be further aggregated to monthly and annual averages as needed.

Caveats

We caution that this methodology provides only an estimate of the mixed HVAC components and is vulnerable to errors under certain conditions. One major vulnerability involves a building with multiple, independently-controlled zones and/or multiple HVAC systems. For example, consider a building with two separately-controlled HVAC systems serving two zones with different balance temperatures. A lowess curve could be fit to each of the systems' energy consumption data. However, if the energies of the two are metered on a single mixed HVAC channel, separate curve fits are not possible. The lowess fit to the summed data would resemble the sum of the two theoretical individual curve fits. This is illustrated in Figure 3. The minimum value of the fit to the summed HVAC load does not correspond to the sum of the minima of the two individual curves, resulting in an overestimation of ventilation/auxiliary loads. Also, the simultaneous heating and cooling that takes place between $T_{b,1}$ and $T_{b,2}$ would not be properly recognized by the summed curve with its single balance point, $T_{b,sum}$. The significance of this error depends on the width of the region between the two individual balance points and the fraction of the year the outdoor temperature falls in that range.

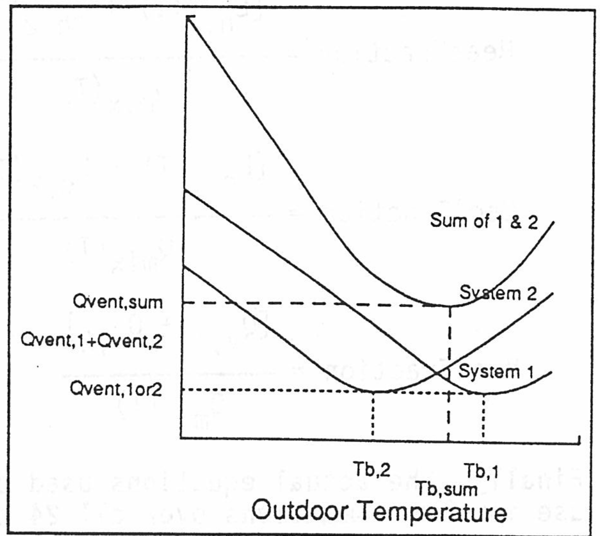

Figure 3 - Effect of Multiple HVAC Systems on Lowess Fit

Another weakness of the method shows up when both heating and cooling occur in a single hour, as might be the case if the sun emerges from behind a neighboring building in mid-morning, quickly warming a small building with a large glazed area. This, too, would have the effect of raising the minimum of the lowess fit and overestimating the ventilation and auxiliary energy fraction.

A final error relates to the ventilation strategy used in a building. Generally, fresh (outside) air must be provided to a building continuously whenever occupants are present. Thus, the fans will run constantly, while the heating or cooling plant cycles on and off as temperature conditioning is required. This strategy is referred to as constant ventilation. However, when the building is empty, the owner may cease to ventilate as an energy conservation strategy. When operating in this mode, the fans are off until heating or cooling is required. This strategy, called demand ventilation, saves energy in two ways. First, the fans use no energy until heating or cooling is needed. Second, because the fans are off, no outside air is being introduced to the building, reducing the amount of heating or cooling that will be needed.

This situation is illustrated in Figure 4. The top (solid) line represents the mixed HVAC load in an example building operating in constant ventilation mode. As the outdoor temperature decreases below the heating balance point ($T_{bal,h}$), the HVAC load increases until the heating equipment is operating at full capacity (Q_{cap}). As the outdoor temperature increases above the cooling balance point ($T_{bal,c}$), the HVAC load increases as cooling energy is consumed. This is completely analogous to the heating side of the curve, though in this illustration the cooling load never reaches or exceeds the cooling equipment capacity.

Between the heating and cooling balance points is a dead band--a range of outdoor temperatures in which neither heating nor cooling is required in the building. It is a result of the typical thermostat strategy that maintains indoor temperatures within a tolerable range rather than at a single predetermined temperature. Notice that within the dead band, energy is still being consumed by the ventilating equipment (fans). Because the fans run constantly, they effectively increase the building's thermal conductance by a constant amount. Thus, the slopes of the heating and cooling curves are constant.

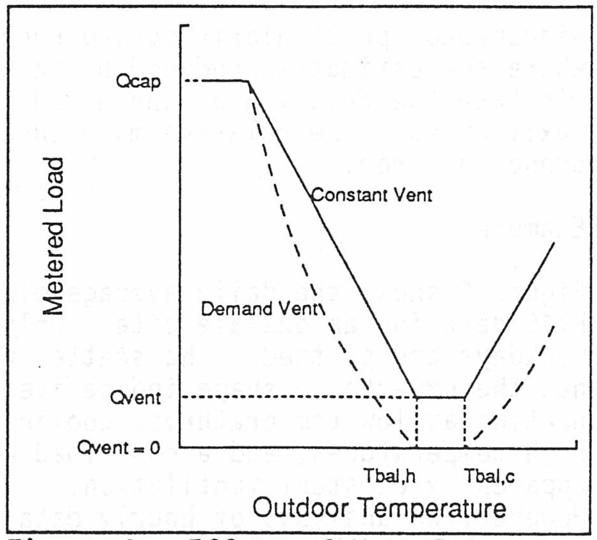

Figure 4 - Effect of Ventilation Strategy on Lowess Fit

The bottom (dotted) line on Figure 4 shows a typical demand ventilation operating curve. Within the dead band, the mixed HVAC load is zero, because neither heating nor cooling is required and the fans are turned off. As the outdoor temperature decreases below the heating balance point[1], the HVAC load increases. However, unlike the constant ventilation scenario, the slope of the heating curve is not linear. This is a direct result of the demand ventilation strategy. As more heating is required, the fans must run more to deliver the heat to the building. As the fans run more, more heating is required to condition the outside air being drawn in by the fans. Thus the curve slopes upward until heating capacity is saturated. At this point, the heating equipment, and hence the fans, run constantly, and the two strategies are identical.

Because the demand ventilation curve is zero in the dead band, the HVAC disaggregation method described here will assign none of the mixed HVAC to ventilation, even when the load is at full capacity. The result is an underestimation of the ventilation/auxiliary load.

In summary, multiple systems on a single measured end use and heating and cooling occurring in a single hour tend to cause an overestimation of ventilation and auxiliary loads. Demand ventilation causes an underestimation of the ventilation and auxiliary loads. Which of these effects dominates varies from building to building. Generally, the problems associated with multiple zones are minimized in reality because large, multizone buildings tend to have large, segregated systems that are easily monitored separately. Packaged HVAC units and other mixed systems tend to be present in smaller, typically single-zone buildings. Likewise, when only a single zone and system are involved, the problems of simultaneous heating and cooling in a single hour are rare. Demand ventilation, on the other hand, is likely to be

---

[1] Generally, demand ventilation operation will also reduce the heating balance point because of the reduced heating load due to outdoor air. In this example, it has not been moved, to simplify the graphics and focus on the phenomenon of interest.

widespread, particularly during unoccupied hours. It should be noted that where the estimates produced by this method are in error, they will tend to misstate the fraction of the mixed HVAC load that is due to ventilation and auxiliaries. The relative magnitudes of heating and cooling loads are less prone to error.

Example

Figure 5 shows the daily average mixed HVAC data for an example site. Only weekdays are plotted. The scatter plot has the expected U shape indicative of heating at low temperatures, cooling at high temperatures, and a base load of apparently constant ventilation. However, an analysis of hourly data shows that the base load is not constant. Figure 6 presents the 24 hourly curve fits. The hour represented by each curve is printed at each end of that curve. Notice that for each daytime hour, the minimum value of the curve is roughly 2500 W. Night hours (after midnight), however, bottom out at nearly 0 W, indicating that ventilation is not constant, but is turned off at night.

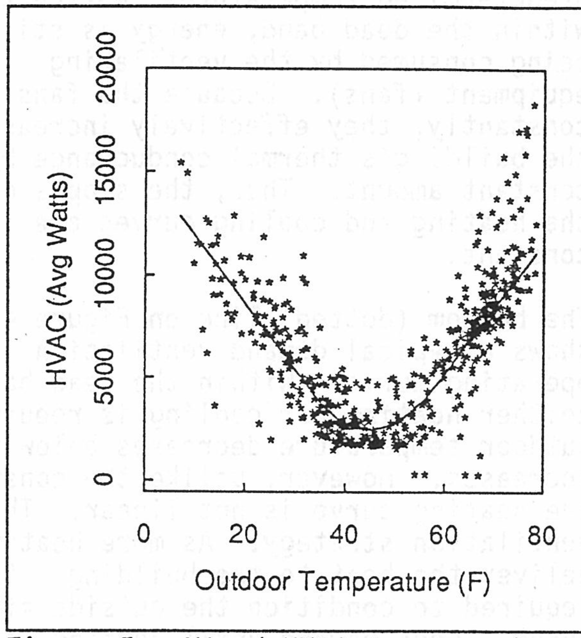

Figure 5 - Mixed HVAC Scatter Plot for Example Building

Because the example site had detailed submetering of the individual HVAC end uses, the end use separations implied by these curves were easily compared with reality. In this case, the estimated ventilation load was higher than actual (about 66%), with concomitant lower estimates of heating and cooling (each about -17%). The ventilation estimate, although high, is about 25% lower than it would have been if the minimum value of the U-shaped curve fit to the daily data had been used to estimate ventilation, indicative that the method accounted for the difference between day and night operations.

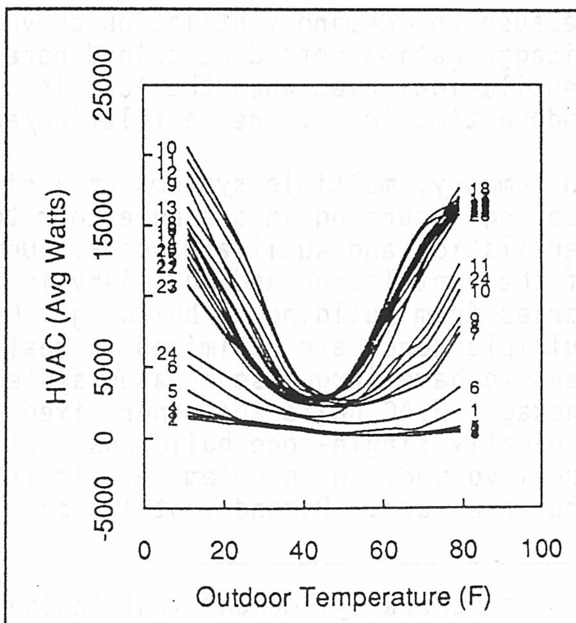

Figure 6 - The HVAC Disaggregation Process

LOW-TIME-RESOLUTION MIXED LIGHTING AND MISCELLANEOUS EQUIPMENT DATA

Separating lighting loads from a mixture of lighting and miscellaneous equipment (plug) loads is considerably more difficult than separating HVAC data. These mixed miscellaneous loads are typically composed of a mixture of end uses that are very highly correlated with one another. Lighting loads are driven primarily by occupancy schedules, but so are plug loads. There is little evidence in the data that differentiates the two components. These loads also show little or no relationship to an external driving force such as temperature, which was used to rationalize the HVAC loads.

The approach we use to disaggregate mixed lighting and equipment loads is unique to ELCAP and other metering projects of similar detail. The approach works only with low-time-resolution (monthly) data. It relies heavily on the availability of detailed information about the lighting and equipment connected to the metering channels. The ELCAP data base includes detailed records of equipment types and capacities metered on each logger channel. The large number of buildings in the ELCAP sample allowed a characterization of the use of each distinct type of equipment in relation to its rated capacity. These characterizations are explained in detail in a forthcoming BPA publication, Commercial Equipment Loads - End-Use Load and Consumer Assessment Program (ELCAP), and we reiterate only the basics of the procedure here.

The equipment use characterizations resulted in utilization factors for various categories of equipment. A utilization factor can be interpreted as the equivalent full-load hours of operation expected when a particular type of equipment is present in a building. That is, by multiplying the rated capacity (kW) of a piece of equipment by its utilization factor, the expected average load (kWa) for that equipment is obtained. The utilization factors were inferred from buildings in the ELCAP sample through regression analyses that related end-use loads to the capacities of the various devices included in the loads. The general form of the regression model was

$$L = \sum_{i=1}^{N} \beta_i * CAP_i \qquad (10)$$

where L = the average annual load for a given metered end use (W)

β_i = the regression coefficient for equipment type i, which is the estimated utilization factor (thousandths)

CAP_i = the rated capacity of equipment category i contributing to the load L (kW)

N = the number of equipment types monitored on the channel.

Observations for the regression represent metered sites. That is, one instance of Equation (10) is obtained for each building with end use L. Separate regressions were run for various combinations of business types, end uses, and equipment types. The results of all regressions were compared to identify the most reliable utilization factors for each business type.

This technique relies on the assumption that the utilization factors, which are developed on groups of buildings, are applicable to individual buildings. While this assumption is clearly inaccurate in many cases, it has been shown to hold for the average across many buildings. (See the publication mentioned above.) The technique also assumes that the utilization factors provide good estimates of monthly loads, even though they were developed with annual data.

This is reasonable for the mixed lighting and equipment loads because there is little correlation with weather beyond (typically) weak relationships between lighting loads and the number of daylight hours. Buildings with heating and/or cooling metered on the mixed channels will be in error. However, the presence of HVAC other than ventilation in the mixed lighting and equipment data is rare in the ELCAP sample and, when it is present, it tends to be a small fraction of the total HVAC load.

Once the utilization factors are known, the mixed lighting and equipment data can be segregated. Consider a building with a mixed end use that measures the consumption of N equipment categories. If these categories are numbered from 1 to N, the fraction of the mixed general load attributable to the ith equipment category is obtained as follows:

$$f_i = \frac{CAP_i * \beta_i}{\sum_{j=1}^{N} CAP_j * \beta_j} \qquad (11)$$

where f_i = fraction of mixed general load attributable to equipment category i
CAP_i = capacity of equipment category i measured as mixed general (kW)
β_i = utilization factor for equipment category i (obtained from regression on equation 10)
N = the number of distinct equipment categories present on the mixed general circuit(s)

Once the equipment load fractions are computed, the mixed data are separated into the pure end uses corresponding to the various equipment categories.

HIGH-TIME-RESOLUTION BUILDING TOTAL DATA

Hourly or 15-minute measurements of building total electric consumption are available from large numbers of buildings metered under the requirements of PURPA. The typical approach to estimating heating, cooling, and other fractions of the total is closely related to the approach we use to disaggregate mixed HVAC data. Figure 7 illustrates the process using daily average building total loads from an example building. Simple linear curve fits to the heating and cooling regions of the scatter plot intersect at the supposed balance point, shown by the vertical dotted line. The building load at this point, shown by the horizontal dotted line, is presumed to consist of lighting and miscellaneous equipment loads and perhaps a constant ventilation load. Loads in excess of this base load are presumed to be heating if the temperature is below the balance point and presumed to be cooling if greater. However, little additional information can be gleaned from the graphic, particularly regarding the portions of the base load consisting of ventilation and internal (lighting and plug) loads.

Akbari et al. (1988) devised a technique involving computer simulations to glean additional end-use resolution. However, knowledge of building characteristics may be insufficient to allow such a simulation. We propose that the techniques of correlating hourly HVAC data to outdoor temperature described above can be applied to building total data to obtain better-informed estimates of end use splits in the absence of detailed building information. Figure 8 illustrates this. Each line represents a best fit (using the lowess procedure described earlier) to the data for a single hour plotted against outdoor temperature. Although not all the hour labels are legible, it is clear that the morning hours are much more likely to exhibit heating loads when outdoor temperatures drop than are afternoon hours. The converse is true for cooling loads. A night, there is very little temperature sensitivity at all and the minimum values of the curves are lower than during the day. These phenomena are indicative of a nighttime reduction in lighting and equipment loads and perhaps a thermostat setback strategy.

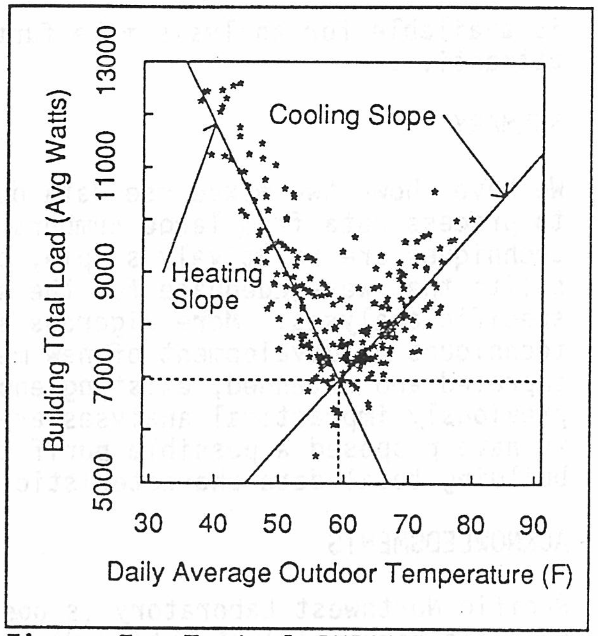

Figure 7 - Typical PURPA Disaggregation

Other telltale patterns that emerge from this type of plot concern the position of the minimum value of each U-shaped curve. Buildings typically have lower lighting and plug loads at night than during the day. The hourly curve fits reveal this by a lower minimum curve value during those hours. The balance temperature will also tend to rise (move to the right) at night because less heat is being generated inside the building. A nighttime reduction of ventilation loads causes a similar shift in the vertical position of the minimum value, but causes a reduction (move to the left) in the balance temperature because less outside air is being introduced to the building. Lower ventilation also changes the slopes of the heating and cooling regions of the curve. By carefully examining the character of these shifts, one can make at least qualitative inferences about the make-up of the base load.

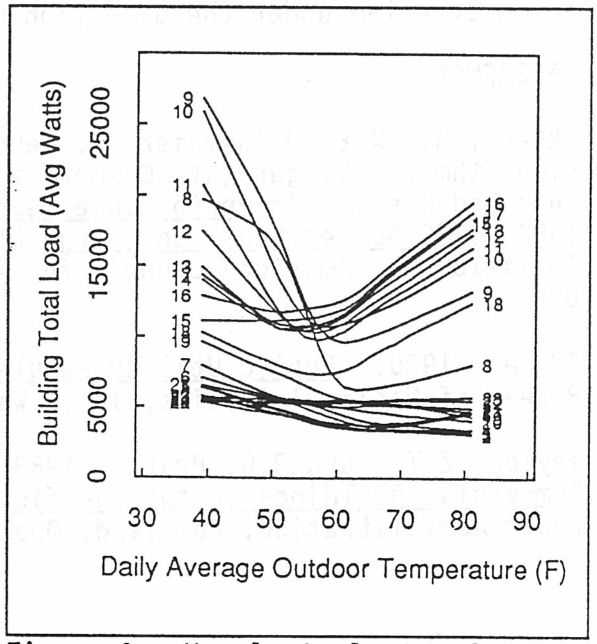

Figure 8 - Hourly Analysis of PURPA Data

Systematic examination of the hourly temperature response curves of the building total data should provide a more rational approach to estimating end use fractions. Introduction of extant knowledge about specific structures might strengthen the conclusions. Although the end-use segregations made possible by this technique are approximations, the wealth of PURPA data that is available for analysis make further pursuit of this and similar procedures attractive.

SUMMARY

We have shown two mixed-use data purification techniques that have been used to process data from large numbers of buildings in a single application. The techniques are relatively simple, providing rough approximations of end-use splits that were adequate for the simple cross-building aggregations of the specific analyses. More rigorous analyses might require extensions to these techniques or development of new methods. As procedures of this type are improved and extended, existing end-use data will increase in value as previously impractical analyses are made feasible. It is in this spirit that we have proposed a possible purification method for high-time-resolution building total data characteristic of PURPA metering.

ACKNOWLEDGMENTS

Pacific Northwest Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under Contract DE-AC06-76RLO 1830.

The work discussed in this paper was funded by the Bonneville Power Administration under the direction of Rich Gillman, End-Use Research Section.

REFERENCES

Akbari, H., K.E. Heinemeier, P. LeConiac, and D.L. Flora. 1988. "An Algorithm to Disaggregate Commercial Whole-Building Hourly Electrical Load Into End Uses." In Performance Measurement and Analysis - Proceedings of the 1988 ACEEE Summer Study on Energy Efficiency in Buildings, Volume 10, pp. 10.14-10.26. American Council for an Energy-Efficient Economy, Washington, D.C.

PURPA. 1980. Public Utility Regulatory Policy Act of 1978. PL 95-617. The Bureau of National Affairs, Inc., Washington, D.C.

Taylor, Z.T., and R.G. Pratt. 1989. Description of Electric Energy Use in Commercial Buildings in the Pacific Northwest. DOE/BP-13795-22, Bonneville Power Administration, Portland, Oregon.