

End-Use Load and Consumer Assessment Program:

Daytyping Methodologies for Analyses of Commercial Building End-Use Loads

Z. T. Taylor

September 1990

**Prepared for
the Bonneville Power Administration
under a Related Services Agreement
with the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830**

**Pacific Northwest Laboratory
Operated for the U.S. Department of Energy
by Battelle Memorial Institute**



DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST LABORATORY
operated by
BATTELLE MEMORIAL INSTITUTE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC06-76RLO 1830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831;
prices available from (615) 576-8401. FTS 626-8401.

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.

NTIS Price Codes, Microfiche A01

Printed Copy

| Price Code | Page Range | Price Code | Page Range |
|------------|------------|------------|------------|
| A02 | 1- 10 | A15 | 326-350 |
| A03 | 11- 50 | A16 | 351-375 |
| A04 | 51- 75 | A17 | 376-400 |
| A05 | 76-100 | A18 | 401-425 |
| A06 | 101-125 | A19 | 426-450 |
| A07 | 126-150 | A20 | 451-475 |
| A08 | 151-175 | A21 | 476-500 |
| A09 | 176-200 | A22 | 501-525 |
| A10 | 201-225 | A23 | 526-550 |
| A11 | 226-250 | A24 | 551-575 |
| A12 | 251-275 | A25 | 576-600 |
| A13 | 276-300 | A99 | 601-Up |
| A14 | 301-325 | | |

DAYTYPING METHODOLOGIES FOR ANALYSES
OF COMMERCIAL BUILDING END-USE LOADS

Z. T. Taylor

September 1990

Prepared for
the Bonneville Power Administration
under a Related Services Agreement
with the U.S. Department of Energy
Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352

SUMMARY

As an outgrowth of the End-Use Load and Consumer Assessment Program (ELCAP) being conducted for the Bonneville Power Administration, Pacific Northwest Laboratory researchers have developed several techniques for dealing with the extensive data set. Among these are daytyping methodologies-- techniques for identifying groups of days in which a commercial building's operation is reasonably uniform. Two daytyping techniques--one based on ordered daily averages, another based on hierarchical clustering--are described and discussed in this document.

CONTENTS

| | |
|--|-----|
| SUMMARY | iii |
| 1.0 INTRODUCTION | 1.1 |
| 2.0 RATIONALE | 2.1 |
| 2.1 WHAT IS DAYTYPING? | 2.1 |
| 2.2 WHY DAYTYPE? | 2.1 |
| 3.0 DAYTYPING METHODOLOGIES | 3.1 |
| 3.1 DAYTYPING BASED ON ORDERED DAILY AVERAGES | 3.1 |
| 3.2 DAYTYPING BASED ON HIERARCHICAL CLUSTERING | 3.3 |

FIGURES

| | | |
|-----|---|------|
| 3.1 | Clustering by Ordered Daily Averages | 3.2 |
| 3.2 | Distance Measurement | 3.5 |
| 3.3 | Day-by-Day Hierarchical Clustering | 3.6 |
| 3.4 | Day-of-Week Hierarchical Clustering | 3.8 |
| 3.5 | Sample Daytyping Matrix Used for Visual Inspection | 3.10 |
| 3.6 | Sample Matrix of Daytype Means Used for Visual Inspection | 3.11 |

1.0 INTRODUCTION

It is often theorized that the schedule on which a commercial building operates is the primary determinant of that building's end-use loads. In this paper, we describe a methodology for determining the impact of commercial building schedules on the loads of various end uses.

Through the End-Use Load and Consumer Assessment Program (ELCAP) being conducted for the Bonneville Power Administration, Pacific Northwest Laboratory (PNL) researchers have collected a vast amount of electric energy consumption data from a large sample of commercial buildings in the Pacific Northwest. The commercial building category includes a wide variety of building types. ELCAP staff have collected data from nine of those types: office, dry good/retail, grocery, restaurant, warehouse, school, university, hotel/motel, and other.

The ELCAP data set is very detailed. Data are collected in highly disaggregated form (every panel circuit is monitored in each sample building) and at high time resolution (hourly). The availability of hourly load data for each end use allows explicit consideration of hour-by-hour schedules in explaining the measured energy consumption. Day-by-day usage patterns are also explicitly considered as part of ELCAP. Because the sample of ELCAP buildings is so large, PNL researchers have developed automated techniques to identify days of similar building operation so that these schedules could be incorporated.

In this document, we describe several techniques used successfully in ELCAP analyses of end-use consumption data. In Section 2, we present our definition of daytyping and discuss the reasons for using a daytyping methodology in end-use load analysis work. The techniques are then described in Section 3.

2.0 RATIONALE

In this section, we explain what daytyping is. We also discuss the reasons for using daytyping techniques in analyses of end-use load data.

2.1 WHAT IS DAYTYPING?

Daytyping is the process of identifying groups of days in which a building's operation is reasonably uniform. Uniformity can be defined in several ways, but generally means similarity of daily energy consumption and/or load shapes.

2.1 WHY DAYTYPE?

Commercial buildings are, by their nature, highly scheduled in terms of their operating schedules. Electricity consumption is strongly driven by operating schedules. These schedules have components by hour of day, day of week, and sometimes week or month of year. Seldom is the hourly schedule identical for all days of a typical week, making analyses that are sensitive to occupancy and use levels difficult to formulate without explicit knowledge of building schedules. However, typical schedules differ widely among the various building types and even within building types. Many retail and grocery stores, for example, are open for business on weekends, whereas most offices are not. Some commercial concerns close business on one or more weekdays. In retail outlets, routine restocking and janitorial tasks are often performed at night, perhaps one or two days per week, making the load profiles for those days significantly different from those for other business days.

Clearly, any analysis of commercial load shapes requires knowledge of operating schedules. It has been suggested that ELCAP offers a unique opportunity to verify schedule assumptions used by forecasters in building simulation analyses. To obtain prototypical schedules, it is necessary to know which days are similar and can be combined. Similarly, interpretation of peak loads is facilitated by an understanding of the temporal use patterns of buildings.

Not as obvious, but perhaps equally important, is the need for schedule information in analyses that do not involve load shapes. For example, an analysis of daily heating, ventilating, and air-conditioning (HVAC) consumption as a function of outdoor temperature is not valid unless the operating conditions are taken into account. A building's thermal behavior cannot be expected to be consistent between occupied and unoccupied days or between days with different hours of occupancy.

Conventional wisdom maintains that operating schedules are the dominant determinants of energy consumption in commercial buildings. To verify this assumption using ELCAP data requires easy access to day-by-day operating schedule information for all metered buildings. The daytyping methodologies discussed in Section 3 are intended to fulfill that requirement.

3.0 DAYTYPING METHODOLOGIES

Assigning days of the week into groups of similar character appears at first to be an easy task. However, the diversity of the commercial sector in combination with the sheer size of the ELCAP sample makes automated analysis of daytypes more difficult. Depending on the analysis needs, several methodologies can be employed. The complexities of these methods and the details of the results vary. Some procedures simply group days of the week, while others group individual days irrespective of days of the week. For example, some methods are based on the assumption that all Mondays are alike, while others are able to detect an unusual Monday and group it with Saturdays.

In this section, we document two methods that have been applied to all the commercial ELCAP buildings. The first is a conceptually simple technique that groups days based on total daily consumption of key end uses. One limitation to this technique is that it does not enable researchers to detect differences in schedule if total daily consumption is not significantly different. The second method uses a more sophisticated clustering algorithm that is capable of differentiating days with similar total consumption but different load shapes.

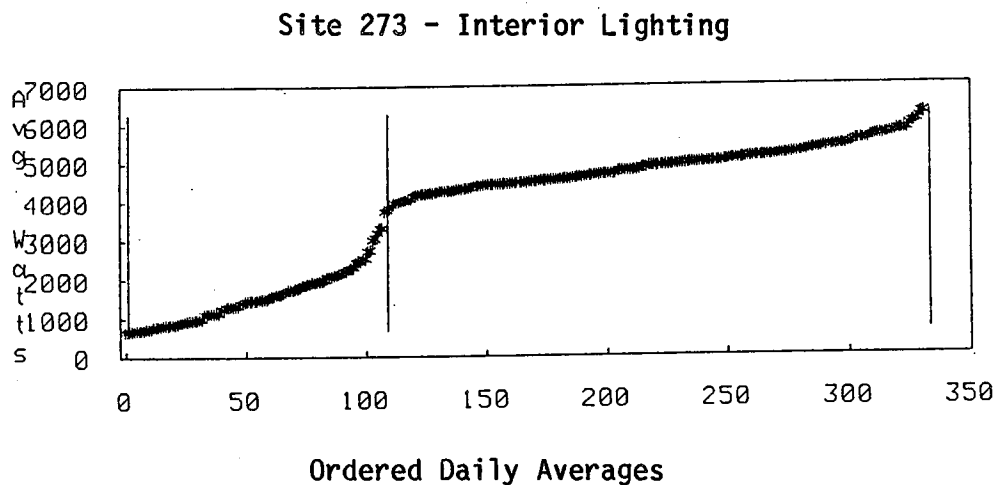
These techniques have been employed to meet specific ELCAP analysis needs. Thus, they do not necessarily represent the "best" techniques overall or the only ones that should be pursued in all analyses of commercial end-use load data. Nevertheless, we offer them here in hopes that other researchers involved in similar work will find them of value in their analyses.

3.1 DAYTYPING BASED ON ORDERED DAILY AVERAGES

Certain analyses are daytype sensitive with respect to overall usage only. That is, whether a day was a workday or non-workday is important, but the details of work schedules are not important. For analyses of that type, we devised a simple technique that differentiates days by daily average consumption. In this technique, it is assumed that a key end use will exhibit consistently higher consumption on workdays than on non-workdays. Similarly,

consistently higher consumption on workdays than on non-workdays. Similarly, partial workdays will exhibit consumption levels that fall somewhere on the continuum between those of workdays and those of non-workdays.

This method of daytyping involves searching for clusters of days with similar consumption levels. The technique is illustrated in Figure 3.1. All daily average consumption values for a key end use (usually lighting) are sorted and plotted in order. If the building is operated in a reasonably consistent manner and if the selected key end use is not sensitive to weather or seasonal variations, then the plot should exhibit a number of plateaus representing various daytypes. By searching the curve for the points of



Site 273 - Interior Lighting

| DayType | Sun | Mon | Tue | Wed | Thu | Fri | Sat | Hol |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 1 | 1 | 3 | 7 | 2 | 3 | 1 | 2 |
| 1 | 48 | 0 | 0 | 0 | 3 | 0 | 48 | 9 |
| 2 | 0 | 44 | 48 | 43 | 44 | 42 | 0 | 3 |

FIGURE 3.1. Clustering by Ordered Daily Averages

steepest change, the breaks between plateaus can be identified. Figure 3.1 shows a building with two plateaus, delimited by the vertical lines drawn at the points of steepest slope.

After the points associated with the different daytypes have been identified, the days can be mapped back to a calendar to help explain the meaning of the daytypes. In this example, daytype 1, with the lowest consumption, represents non-workdays. These are shown to be primarily weekends and holidays. Daytype 2 encompasses the higher consumption days, which are primarily weekdays. Daytype 0 represents days for which data are unavailable.

The example in Figure 3.1 reflects an unusually consistent building schedule. Although this case is somewhat simplified, it highlights the important characteristics of the ordered daily average technique. First, days are clustered without any day-of-week information. This allows special occasions, such as weekend workdays or weekday non-workdays, to be handled properly. Second, only the level of daily consumption is important. For example, if the building is open for business for 9 hours each weekday but opens and closes 1 hour later on Fridays, the ordered daily average technique will cluster Fridays with all other weekdays because overall energy consumption will be the same.

3.2 DAYTYPING BASED ON HIERARCHICAL CLUSTERING

Some analyses are sensitive not only to daily consumption levels, but to daily schedules as well. Any analysis dealing with load shapes will fall into this category, as will any analysis concerned with time of use.

Clustering days based on similarity of both daily total consumption and schedule requires a more sophisticated sorting methodology than that just described in Section 3.1. The method used here is based on the statistical technique of hierarchical clustering. With this technique, observations of a number of variables are collected into statistically similar clusters.

The first step in the clustering process is to calculate "distances" between each observation and all other observations. A number of distance

metrics are possible, but the one used here is the euclidean distance, which is defined as the root sum-of-squares of distances for all variables. That is, the distance between two observations i and j , when N variables are considered, is

$$\text{dist}_{i,j} = \sqrt{\sum_{k=1}^N (v_{k,i} - v_{k,j})^2} \quad (3.1)$$

where $\text{dist}_{i,j}$ = euclidean distance between observation i and j
 $v_{k,i}$ = i 'th observation of variable k
 $v_{k,j}$ = j 'th observation of variable k
 N = the number of variables.

When distances among all observations have been calculated, they form an N by N matrix. The hierarchical clustering algorithm then searches the matrix for the two most similar observations (i.e., the two with the smallest distance between them). These two are then combined to form a cluster.

At the next step, two other points might be clustered together, or a third point might be added to the first cluster if the distance between that point and the cluster is smaller than any other distances. This continues until all points have been combined with a cluster and all clusters have been combined with other clusters.

Consider the following example. Site 444 is a small bank building. Each day's 24 hourly values constitute 24 variables to be observed. Each day represents one observation of these 24 variables. Because interior lighting is known to be largely schedule-driven, with little seasonal variability, it was selected as the key end use for daytyping the building. A distance matrix was constructed according to Equation (3.1):

$$\text{dist}_{i,j} = \sqrt{\sum_{h=1}^{24} (\text{ILT}_{h,i} - \text{ILT}_{h,j})^2}$$

In this example, the i and j are indices for two days, and the h is an index for hour of day. The ILT stands for the ELCAP metered values of interior lighting (average watts). Figure 3.2 illustrates the difference values that go into the distance measurement for a sample pair of days. In this example, Day 1 is a Thursday and Day 2 is a Friday.

After clustering from the resulting distance matrix, the results may be graphically depicted as a tree, as shown in Figure 3.3. The euclidean distance between any two clusters may be read from the y axis at the height of the merger of the two. The "leaves" of the tree are ordered such that, at any merger, the cluster on the left has a lower level of average energy consumption.

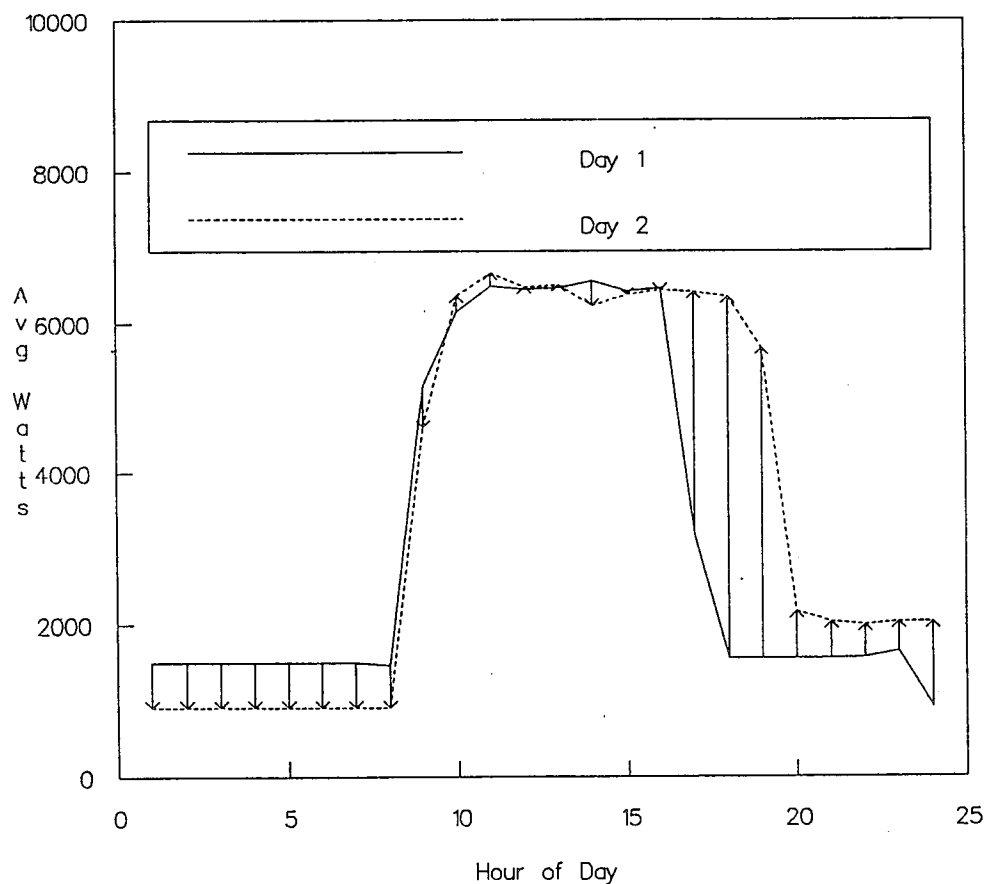


FIGURE 3.2. Distance Measurement

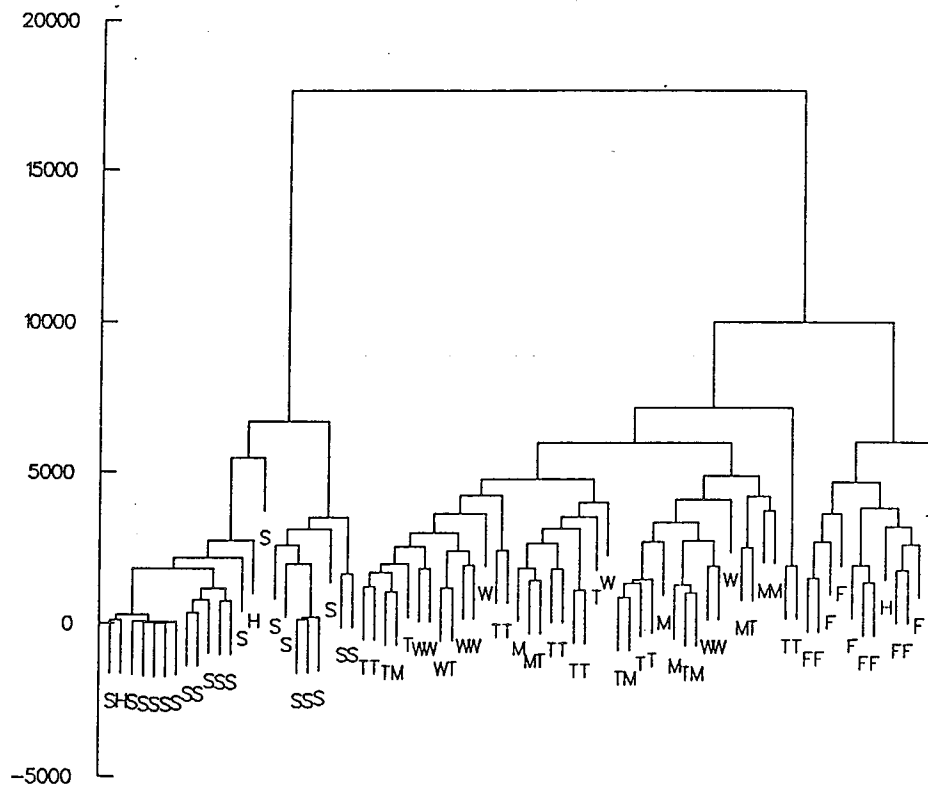


FIGURE 3.3. Day-by-Day Hierarchical Clustering

Two major clusters are identified in Figure 3.3; they merge at a euclidean distance of approximately 18,000. Clearly, from the labels describing each point in the cluster tree, the difference between the two major clusters is that one represents weekends and the other weekdays. The second most significant merger occurs at a euclidean distance of roughly 10,000. This merger represents the difference between Fridays and other weekdays. With the exception of a single day (either a Tuesday or a Thursday), the rightmost cluster, which has the highest consumption of all clusters, consists only of Fridays. From a survey of the building occupants, we learned that the bank stays open on Fridays 2 hours later than on other workdays. This was also observed in the example days shown in Figure 3.2.

The clustering technique properly identified the major daytypes of workdays and non-workdays and also identified a less significant difference between Fridays and other workdays. By "cutting" the tree at successively lower values of euclidean distance, more clusters might be identified that have physical significance. In this case, however, it appears that the merge points of the remaining clusters occur at relatively small distance values, indicating that the major schedule differences are accounted for by the three clusters already identified.

Again, the example represents a fairly simple building with very consistent scheduling. The same understanding of schedules could have been deduced from survey data. However, most buildings are not nearly as consistent as this bank, and the operating schedules of many may change after surveys are taken. The clustering technique properly accounts for anomalous days and for sudden changes in schedules.

The clustering example shown in Figure 3.3 operates on daily load shapes and clusters similar days regardless of day-of-week information. For some analyses, retention of day-of-week identity is important. Examples of such analyses include those in which the metered load profiles are used to calibrate the inputs to hourly load models and those aimed at quantifying the impacts of schedules on load variations of various end uses. In both cases, it is desired to characterize loads by identifiable schedule parameters such as hour of day and day of week. To accommodate these analyses, the clustering technique must be slightly modified.

For most buildings, schedule differences by day of week and hour of day dominate the schedule-attributable variation in loads. Month-to-month variations might be important in some cases (a hotel in a winter resort, for example), but will otherwise usually be small. Hour-of-day differences are accounted for by the distance matrix algorithms described above. The technique used here to characterize day-of-week variations is simple averaging. For example, the hour-by-hour data for all Mondays in a specified period are averaged to yield an average Monday profile. Other days of the

week are treated similarly. Holidays can also be averaged separately if so desired.

Having day-of-week average profiles, the clustering technique can be applied exactly as described before. In this case, however, there are only 7 observations (the 7 days of the week). As before, the 24 hourly values in each profile constitute the variables used in computing distances. An example of the resulting clusters is shown in Figure 3.4, which is based on the same building that was shown in the daily clustering of Figure 3.3. Although there are only 7 leaves in the tree, the technique still identifies weekends and weekdays as distinctly different and identifies Fridays as significantly different from other weekdays. Again, at any merge point on the tree, the cluster to the left has lower average energy consumption than the cluster to the right.

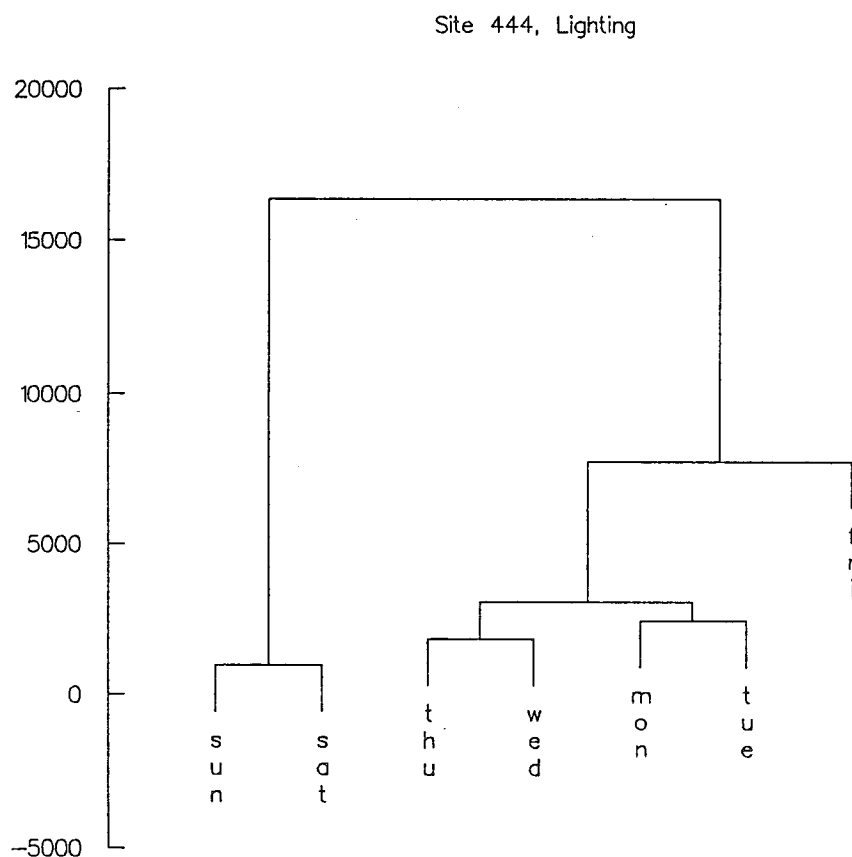


FIGURE 3.4. Day-of-Week Hierarchical Clustering

Because month-to-month variations might be important in some cases, and also to identify any sudden changes in operation schedules, separate day-of-week average profiles were generated for each available month. Therefore a separate clustering analysis was done for each month.

In this example, it is reasonably obvious that three day types adequately describe the building's operation. However, that conclusion is based on foreknowledge of the operating schedules as well as some subjective reasoning.

To automate the technique to work for arbitrary buildings without user intervention, a methodology for "cutting" the cluster trees was necessary. We tested various methods with little success. The problem is that the euclidean distance is not directly comparable to the power levels of the end use, and the scale of the distances relative to the overall magnitude of the data is difficult to ascertain computationally. The question resolves to "How different is different?"

The tree-cutting technique we finally selected is largely empirical, based on visual examination of a number of trial sites, and requires the analyst to visually select the final daytypes. The steps in the technique are as follows:

1. Transform the tree's y-axis (distance) scale to rough units of watts. This is accomplished by "working backward" through Equation (3.1). First, each y value is squared. It is then divided by 24 to obtain an average squared distance. Finally, its square root is taken to transform the distance values to the order of magnitude of watts.
2. Compare the transformed y values to the mean power levels of all observations (days of week). Cut the tree where the transformed distance value equals or exceeds 30% of the mean power level of all observations. This will result in anywhere from one to seven daytypes. The 30% value was found to adequately distinguish day-of-week differences. It is conservative in that it tends to select too many daytypes rather than too few.
3. Map each day of week to the appropriate daytype and print the results for each month. Visually examine the printout to identify persistent patterns and eliminate insignificant fluctuations.

The third step, which requires the analyst's active observation, is very important. A sample printout of daytyping information from site 556 is shown in Figures 3.5 and 3.6. Each row of Figure 3.5 represents a different month. Each column represents a day of the week. The data in the matrix tell which daytype number was assigned to that day of the week during that month. Daytypes are numbered such that the highest consumers are in daytype 1. Each column of Figure 3.6 corresponds to one of the daytypes from Figure 3.5. Rows are again months. This matrix contains the mean power consumption of all days in each daytype. By visually comparing these two matrices, the analyst can distinguish anomalous months from pervasive patterns.

Month 65, for example, was found by Step 2 to have four distinct daytypes. Wednesday, having the highest consumption, appears alone as a daytype. The other weekdays comprise the second daytype. The third and fourth daytypes are Saturdays and Sundays, respectively.

| Month | Day of Week | | | | | | |
|-------|-------------|-----|-----|-----|-----|-----|-----|
| | Sun | Mon | Tue | Wed | Thu | Fri | Sat |
| 65 | 4 | 2 | 2 | 1 | 2 | 2 | 3 |
| 66 | NA | NA | NA | NA | NA | NA | NA |
| 67 | 4 | 2 | 2 | 1 | 2 | 2 | 3 |
| 68 | NA | NA | NA | NA | NA | NA | NA |
| 69 | 4 | 3 | 3 | 1 | 3 | 3 | 2 |
| 70 | 3 | 2 | 2 | 1 | 2 | 2 | 3 |
| 71 | 3 | 2 | 2 | 1 | 2 | 2 | 3 |
| 72 | 3 | 2 | 1 | 1 | 2 | 2 | 2 |
| 73 | 3 | 1 | 1 | 1 | 2 | 3 | 3 |
| 74 | 1 | 1 | 1 | 1 | 3 | 2 | 1 |
| 75 | NA | NA | NA | NA | NA | NA | NA |
| 76 | 3 | 2 | 2 | 2 | 2 | 1 | 4 |
| 77 | 3 | 2 | 1 | 2 | 2 | 2 | 3 |
| 78 | NA | NA | NA | NA | NA | NA | NA |
| 79 | 5 | 4 | 2 | 4 | 1 | 4 | 3 |
| 80 | NA | NA | NA | NA | NA | NA | NA |

FIGURE 3.5. Sample Daytyping Matrix Used for Visual Inspection

| Month | Day Type Number | | | | | | | |
|-------|-----------------|----------|----------|----------|---------|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 65 | 72420.6 | 61445.5 | 52608.08 | 43951.87 | NA | NA | NA | NA |
| 66 | NA | NA | NA | NA | NA | NA | NA | NA |
| 67 | 72644.3 | 61898.0 | 55936.08 | 49839.46 | NA | NA | NA | NA |
| 68 | NA | NA | NA | NA | NA | NA | NA | NA |
| 69 | 86020.8 | 77092.8 | 75866.36 | 62526.92 | NA | NA | NA | NA |
| 70 | 85755.2 | 75397.9 | 61929.83 | NA | NA | NA | NA | NA |
| 71 | 85756.7 | 80135.3 | 74456.07 | NA | NA | NA | NA | NA |
| 72 | 88505.9 | 77875.9 | 76599.00 | NA | NA | NA | NA | NA |
| 73 | 126324.0 | 108005.3 | 81480.23 | NA | NA | NA | NA | NA |
| 74 | 107409.9 | 86022.7 | 83786.50 | NA | NA | NA | NA | NA |
| 75 | NA | NA | NA | NA | NA | NA | NA | NA |
| 76 | 103263.1 | 101649.8 | 88085.34 | 68494.17 | NA | NA | NA | NA |
| 77 | 82241.0 | 77515.0 | 61703.75 | NA | NA | NA | NA | NA |
| 78 | NA | NA | NA | NA | NA | NA | NA | NA |
| 79 | 77165.0 | 76007.2 | 72395.50 | 64215.96 | 55789.8 | NA | NA | NA |
| 80 | NA | NA | NA | NA | NA | NA | NA | NA |

FIGURE 3.6. Sample Matrix of Daytype Means Used for Visual Inspection

Examining the entire matrix, a pervasive pattern appears. Wednesdays virtually always have higher consumption than other weekdays, and Saturdays usually have higher consumption than Sundays, although not always. However, beginning with month 73, the pattern changes. Apparently, an operational change occurs at that point. Figure 3.6 confirms this suggestion. Consumption levels rose by more than 40% between months 72 and 73 (December 1986 and January 1987).

This observation highlights an important aspect of the hierarchical clustering methodology. Not only are pervasive daytype patterns easily identified, but sudden changes in operation and/or occupancy may be located as well. The technique will be useful in future analyses in which vacant periods and tenant changeovers must be identified.

DISTRIBUTION

No. of
Copies

OFFSITE

- 2 DOE/Office of Scientific and
Technical Information
- 2 R. A. Gillman
Bonneville Power Administration
End-Use Research Section
P.O. Box 3621-RPEE
Portland, OR 97208
- M. E. Taylor
Bonneville Power Administration
End-Use Research Section
P.O. Box 3621-RPEE
Portland, OR 97208
- W. M. Warwick
Battelle Portland Office
500 NE Multnomah, Suite 650
Portland, OR 97232

No. of
Copies

ONSITE

- 13 Pacific Northwest Laboratory
- Z. T. Taylor
W. F. Sandusky (5)
G. M. Stokes
Publishing Coordination
Technical Report Files (5)

