DEALING WITH DATA IN LARGE END-USE LOAD METERING PROJECTS-DATA QUALITY AND DATA ACCESS

Mark A. Halverson, Craig C. Conner, Julia T. Caplinger
Pacific Northwest Laboratory

ABSTRACT

Hourly end-use electrical consumption data collected for the Bonneville Power Administration and the United States Department of Energy as part of the End-Use Load Consumer Assessment Program (ELCAP) are subjected to extensive data processing prior to entry in the data archive. Due to the extremely large size of the data set, all data processing activities must be highly automated. Two of the most important data processing activities are data quality checks and pre-aggregation of the hourly data to daily and monthly levels.

Data quality checks allow both the rapid identification of problems and the encoding of a concise indicator of data quality for individual data records. A detailed procedure based on conservation of energy principles has been developed for data quality checks on data collected as part of ELCAP. The pre-aggregation processing makes the most commonly used aggregations (daily, monthly, monthly profile) directly available for analysis, freeing the analyst to concentrate on the information content of the data. One of the most important lessons learned in ELCAP is that up to 50% of the time and money necessary to produce an analytical product may be spent in just preparing the data for analysis. Automation of the pre-aggregation process significantly reduces the time needed for data preparation.

This paper describes the need for automated data quality checks and pre-aggregation processing and discusses the specific checks that are performed on all ELCAP data as it is collected, processed, and archived for eventual analysis.

DEALING WITH DATA IN LARGE END-USE LOAD METERING PROJECTS-DATA QUALITY AND DATA ACCESS

Mark A. Halverson, Craig C. Conner, and Julia T. Caplinger
Pacific Northwest Laboratory

BACKGROUND

The End-Use Load and Consumer Assessment Program (ELCAP) has been collecting electrical consumption data since late 1984. The monitored sample includes approximately 400 homes, 50 apartments, and 150 commercial buildings. (For a complete description of the original sample, see Parker and Foley, 1985; Windell, 1987; Baker, 1984.) More than 90% of the data were collected at hourly increments, with the remainder collected at 5-minute and 15-minute increments. As of January 1988, more than 330 million data points had been collected, with approximately 16 million more data points being collected each month. The program is sponsored at Pacific Northwest Laboratory (PNL) by the Bonneville Power Administration (BPA) and the United States Department of Energy.

The goal of the project is to produce a very large data set of known and verifiable quality that may be used to address conservation and load forecasting issues for BPA. The project must be able to address these issues in a timely manner to meet the needs of BPA. The analyses based on this data must be scientifically defendable and repeatable due to the large economic impacts associated with some of the issues. To meet these needs, an extensive program of data processing activities has been developed. Two of the most important activities are data quality checking and pre-aggregation of the data.

The very large volume of the data set causes immediate problems involving data storage and access time. On-line mass storage devices provide adequate storage capacity but even these devices are strained to provide quick access time if a significant amount of data is required. The need for rapid analytical access to the data has led to an extensive program of pre-aggregation of the data set. This process builds the data sets which are precursors to many of the analysis activities, saving considerable time for the users of the data who would otherwise have to start from the raw data. The pre-aggregated data is presented to the data user in a standard format using standard data set names and with standard quality assurance applied to all data.

A fundamental law for the routine operation of large metering projects was developed early on in the ELCAP program. The law in its simplest form is:

$$1000 >> 10$$

or to paraphrase the law:

Things that work for 10 sites hardly ever work for 1000 sites.

A corollary to this law is " Never do anything once if you are not prepared to do it a thousand times". These simple observations have profound implications for really large projects. Detailed procedures for processing and analysis that cannot be completed within a reasonable amount of time are essentially useless. In a large metering project, there is just not enough time to treat every site as an individual case study. This requirement for efficiency in processing dictates that routine operations must be automated. All the operations described in this paper automated to the largest extent possible.

The need for repeatable and scientifically defendable analyses is the basis of our extensive data quality activities discussed in the next section of this paper. This project (or any other metering project) needs to be able to determine data quality. Data of unknown quality is also data of unknown value. Experience on this and other metering projects has demonstrated that metering projects, like everything else, are subject to error. Faulty installations, hardware and software failures, and actual changes at the metered site should all be expected and should be detectable by a good metering protocol.

DATA QUALITY ACTIVITIES

Data quality is a function of proof and not of assertion. A metering project can use state-of-the-art equipment, well-trained installation teams, the finest computer equipment available, and still collect data that are essentially garbage. What is worse, in many cases the project would not even know that some or all of the data were bad. Without some way of performing a large portion of the data quality checks as the data is collected, the project runs an enormous risk of collecting useless data. Without a detailed procedure for metering the sites and a detailed procedure for determining the quality of that installation, many problems would have occurred.

Figure 1 is a schematic diagram of some of the data quality activities that take place in ELCAP after a site is recruited into the program. Starting from the time of measurement plan development (essentially the design of the metering strategy for that site) and continuing on through the installation and verification of the site and the aggregation to end-use data, data quality issues are extremely important. The discussion below on data quality issues in ELCAP will help to define the quality assurance functions described on the right hand side of the figure. Figure 1 is also a simple flowchart for the path the data takes from the time of data collection until it is archived and made available for analysis.

ELCAP Metering Protocol and Sum-Check

Consider a single data point collected as part of an end-use metering project. The point might represent the total electrical usage for hot water heating at a particular site at a particular hour. By itself, this data point is meaningless. If the metering protocol is known, some estimate of the error associated with this value might be assumed, but there is really no way to prove that the value is correct. One possible way to estimate the "correctness" of the data point is to compare it to many other single points. The data point could be compared to other points at the same site at different hours or to points at different sites at the same hour, but neither of these comparisons give you a "correctness", only a "reasonableness", which demonstrates only that the data point falls in the range of possible values. The only way to determine whether or not you actually have the "correct" value is to compare the value to one or more independent measurements of the same load. If the two measurements are more or less the same, you could assume that you had measured the "correct" value. At the very least, you could assume that your measured value was more "reasonable".

ELCAP Metering Protocol. A complete discussion of the ELCAP metering protocol would take more space than is available here. (See, for instance, Pearson *et al*, 1985.) This protocol is not necessarily the only protocol that will give good results, but it has shown great success in ELCAP and is a valuable example of what can be done to ensure data quality. Consider a breaker panel located in a normal residence in the Pacific Northwest. ELCAP metering protocol calls for each phase of the main feeder cables to the breaker panel to be monitored independently and for each wire leading from a breaker to be monitored. Figure 2 shows a simplified breaker panel monitored according to the ELCAP protocol. The breakers may or may not be monitored independently, depending on the size and complexity of the breaker panel. In the case where breakers are not monitored independently, the protocol calls for only breaker wires of the same phase to be combined. A sum-check process can then be applied to the single value for one phase of the power going into the breaker panel and the sum of the measured values for the breakers. (The breaker level data is eventually aggregated to the end-use level, but most of the data quality checking in ELCAP is done with the breaker level data.) The sum-check process is essentially an application of the first law of thermodynamics to a building electrical system. The procedure compares the sum of a series of energy measurements to a single redundant measurement of the total load for these measurements. The sum-check difference is not zero, but should be within some relatively small range for a given measurement protocol and given hardware. This sum-check process is the basis of many of the ELCAP data quality checks.

The concept of channels of data rather than end-uses is also important in ELCAP. Channels are best thought of a pieces of end-uses that can be later combined to form end-uses. In most residences, it is relatively unimportant whether or not the data collected is in channels to be combined into end-uses or directly as end-uses. There are relatively few electrical panels, and all breakers for a particular end-use tend to be close together. In a large commercial building, the story is different. Panels may be separated by several floors, breakers associated with the same end-use, such as heating, may be found in 10 different panels. In order to eliminate long cable runs and reduce the work involved in an installation, ELCAP protocol calls for multiple data loggers to be used in those cases where widely separated panels are to be monitored. These individual data loggers of necessity then collect channels, rather than whole end-uses. The channel data are aggregated during the processing of the data after collection and archiving.

Associated with the ELCAP load data archive is an extensive set of control data which defines how the channel data from the various loggers is to be combined to form end-uses for particular buildings. One of the useful features of this storage procedure is that end-uses can be re-defined should an error be discovered in the original assignments of end-uses or should changes occur at the site. This is especially useful when comparing data across projects with different end-use definitions. For example, some projects will define separate heating,cooling, and ventilation end-uses, while others rely on a single HVAC end-use. Different categories of loads may be characterized as "Other" in different studies. Typically, the channel information collected as part of the ELCAP commercial sample is zoned, either by heating systems, floors or tenants. This makes the channel information very useful to those analysts who are more interested in interactions between parts of buildings than in the building as a whole. The channel data are not available in the pre-aggregated data set, but are available to the analyst willing to spend the time to use the extraction programs developed for ELCAP.

Sum-Check Uses. The sum-check procedure allows problems to be detected, both at the time of initial installation and at any time after installation. Data for each site is examined in detail after the initial installation and after every major maintenance site visit. A simple one-time sum-check of the load data is one of the data quality checks used by the installation team. For a study like ELCAP, where installations were made by 5 different local electrical contractors fielding 9 installation teams in 4 states with only minimal supervision from project staff, the comparison process provides a crucial site installation verification function. It is surprising just how many ways an installation can be faulty, even when trained staff are involved. With over 600 installations, many problems were bound to occur and did indeed occur.

Each data record is sum-checked shortly after it is collected. This often allows problems to be caught not long after they occur at the site. After four years in the field, even the finest electrical equipment is subject to hardware failure. Solder joints loosen on circuit boards, transistors burn out, and power supplies gradually fail. Many of these failures have a characteristic appearance in the data and can be caught with the sum-check procedure. Another problem that typically occurs with greater frequency than is generally anticipated is an owner modification to the electrical service. In the sample of commercial buildings monitored in ELCAP, approximately one half of the sites were at least partially rewired in the last two years. (Stoops, 1988.) Once again, the sum-check spots many of these problems.

The list below shows some of the problems that have been found during the sum-check process in ELCAP.

- Improper installation of primary metering equipment
- Improper calibration of metering equipment at time of installation
- Incomplete metering at site, e.g. missed loads, panels or switches
- Owner modifications to site, e.g. new loads, moved loads, reworks
- Hardware Failures, e.g. primary sensors, data logger
- Control data entry errors, e.g. data processing errors

These problems should be anticipated in any end-use metering project. The frequency of any particular problem depends on several things. A building with a high turn-over in occupants can be expected to have a high rate of owner modifications. Even buildings with no occupant turn-over can be expected to change over time. The rate of data processing errors and installation errors is entirely dependent on the size of the project and the capabilities of the staff. The quality of the actual metering equipment also influences greatly the rate of hardware failures.

Development of the ELCAP Data Quality Criteria

The sum-check compares the difference in two measurements to a pre-determined tolerance in order to determine data quality. It is clear that the measurements can only be as good as the monitoring hardware used in the project. Any attempt to make the sum-check difference less than the inherent error of the metering equipment is doomed to failure.

Given the error terms associated with the primary sensing devices, the error terms associated with the electronic components present in the data logger, and any error terms associated with round-off or truncation of data records in the data logger, it is mathematically possible to develop some criterion or set of criteria for a "good" sum-check. In reality, it is easier to base the definition of data quality on a series of test installations rather than on first principles. This is exactly what was done in ELCAP. A series of 40 test installations were made and the homes monitored for a month. The criteria for "good" were developed after the data for the test homes were evaluated. These criteria were then applied to the remaining installations.

Three components of the ELCAP data logger have identifiable error terms. (Tomich and Schuster, 1985). The components with error terms are the current transformers used as the primary metering device, the 8-bit analog-to-digital converter used to produce a binary signal from the analog output of the current transformer, and the data storage algorithm used in the memory of the data logger. In addition, the ELCAP data logger has a relatively "soft" zero, i.e. the logger may read -1,0 or +1 counts for a true zero reading. A complete description of how these error terms propagate through the system would take a separate paper. (See for instance Crowder and Miller, 1987). The important concept is that the error terms and therefore the criteria for "good" are highly dependent on the hardware and software in use in project. One would not expect the ELCAP "good" criteria to be applicable to other project, nor would the criteria from other projects be applicable in ELCAP.

Additional Uses of the Sum-Check. The sum-check process described above not only finds problems in the data, but is also used to code a data quality flag for each data record collected. The overall quality of the installation has already been determined, but individual data records may vary considerably. Each record is given a value indicative of the quality of that record compared to the best ELCAP records as determined by the sum-check. The flags are integers from 1 through 9, with the higher numbers indicating less useful or less accurate data.

Another use of the sum-check is in the initial evaluation of an installation. Typically, a week to 2 weeks worth of data is examined, or verified, by project staff immediately after an installation or major rework of the installation. This verification procedure takes up to 2 days for complicated commercial sites, including actual verification time and the required documentation. This process is facilitated by the use of a knowledge-based expert system developed at PNL (Laufmann and Crowder, 1987) solely for the purpose of reducing verification time. A problem requiring a site visit can be identified and a new visit made to the site in a couple of weeks, but only if the site is given top priority in the verification queue. More typical is a time lag of a month or more before anyone really looks at the data in detail. This is to be expected in large projects.

The inevitable time lag between a verification and a site visit (if required) leads to another use for the sum-check. The redundancy required by the sum-check procedure provides a means of recovering lost data. In the case where a hardware failure leads to loss of data from particular monitored breakers, the sum-check equation may be "solved" for any one or more of the individual feeder loads by subtracting the other feeders from the main load. One consequence of this action, called reconstruction, is that the error term is now combined with the calculated load, leading to greater uncertainty in that particular load. Reconstructed data is given a status which indicates that all normal ELCAP sum-checks were not possible, but that otherwise the data is thought to be good.

End-Use Checks on ELCAP Data

The majority of analysts are more interested in the whole building end-uses rather than data at the breaker level. This is especially true in the case of residences. The first step of the pre-aggregation process is therefore to extract the data from the channel archive and create the end-uses. This is simple conceptually, but takes a long time due to the large amounts of available data.

All the data quality checks discussed so far have been based on the channel level data and the sum-check process. A second series of tests is used by the verification staff to determine the "reasonableness" of the end-use loads being monitored. Small amounts (2-3 weeks) of data are examined for abnormal usage patterns or obviously incorrect end-use assignments. An example might be a residential channel labelled "refrigerator" that draws

5000 watts for half the night. Most loads do not exhibit such dramatic abnormalities and therefore the reasonableness checks are not as powerful a test as the sum-check.

Once the data have been extracted into hourly end-use data, three additional kinds of quality assurance checks are performed; range checks, repeated value checks and end-use sum-checks. The range checks are essentially reasonableness checks. Simple upper and lower limits have been defined for end-uses such as interior temperature, humidity and wind speed. For outdoor temperature a range of possible values have been defined for each of the four seasons in each of several climate zones. The most complex range check is performed for the pyranometer readings where a range has been defined for each hour of the day for each season of the year. Data that do not pass the range checks generally indicates failed hardware, poor location of the sensors, or data processing errors. Repeated value checks are performed for indoor and outdoor temperature, humidity, and wind speed. Data values which are found to repeat more than a specified number of consecutive hours are flagged. This procedure will catch certain types of hardware problems. The third type of error checking is end-use sum-checking (not to be confused with the electrical panel sum-checking discussed previously). End-use sum-checking compares an individual end-use to the sum of its components. For ELCAP residential data, the building total should equal the sum of the HVAC, Water Heating and OTHER end-uses, within the tolerances of the installation. The OTHER end-use should also equal the sum of its component end-uses, such as Lights and Plugs, Refrigeration, Range, etc. After completion of the end-use level quality assurance procedures, the data is ready to be added to the pre-aggregated analysis-ready data set. Data failing these checks are made missing in the pre-aggregated data set, then flagged and sent back to verification for inspection and possible correction.

THE ELCAP PRE-AGGREGATED DATA SET

The ELCAP data are presented to the data user in several forms. The data is "pre-aggregated" at four temporal levels and into several types of data described below. The goal of the "pre-aggregated" data is to present the data user with data in as useful a form as is practical. Specific rational methodologies for aggregating the data, checking for errors and dealing with missing values need to be employed. For data users who have unusual requirements, the data should be presented in a form allowing them to efficiently create special data sets. From an operational standpoint, it is important to construct software producing a compact data set which can be efficiently maintained and accessed. These operational problems are compounded by the dynamic nature of the data set and the size of the data set.

Pre-Aggregated Data From The Production Perspective

The storage of the ELCAP archive data set at the channel level, with mixed resolution of 5-minute, 15-minute, and 60-minute resolution gives an analyst the greatest opportunity to analyze data in any detail necessary to an analysis. While this gives the greatest possible flexibility to the data analysis, most analyses require the aggregation of the data in two or three ways, including aggregation to the end-use level, aggregation to the hour, day or month level, and aggregation across loggers for a multiple data logger site. Data extraction software has been written for the project allowing a user to custom design the data set extraction to meet the analysis aggregation and data quality requirements. Although the software is flexible, the aggregation process is complex, and requires up to 10 CPU minutes per site year on the project's super mini-computer. Although a significant amount of analysis of the ELCAP data has required the detail provided in the archive data set, most analysis has involved the use of data which has been aggregated in basically the same way. The extraction of the data into the format needed has often consumed more than 50% of the analysis resources for any analytical product. As described in the next section, an analysis-ready data set was designed which meets the needs of a broad range of analysis products. The objectives in the design of this pre-aggregated data set center around the storage of the most usable information for analytical work, maintaining the data set with frequent updates, verifying the quality of the data at the end-use level, and providing fast, but simple access to the data.

The ELCAP data set is stored on-line on magnetic storage media in order to facilitate the easiest access to the data. The huge volume of the data has necessitated that data compression techniques be employed in order to allow all the data to be stored on-line. A modified run-length encoding method has been used which allows the archive data to be stored in one-eighth the space required by ASCII data storage. A further modified version of this encoding method has been used for the storage of the analysis-ready data set.

Typically, ELCAP analysis is done using statistical graphics software. Although the prototype pre-aggregated data set was stored in the format of the statistical graphics software, the data set compression did not meet the on-line storage needs of the project, and was replaced by the modified run-length encoding method for the production data set. The data can now be accessed in two ways, either by a direct conversion to ASCII, or through extensions to the statistical graphics software which allows a vector to be read from the pre-aggregated data set directly into the analysis environment.

The production and maintenance of the analysis-ready data set is done in a highly automated manner. Because of the large amount of data being continually collected for the project, the processing of the data has been done on a schedule. The production of the analysis-ready data set has been placed at the end of the processing cycle after the data acquisition from the logger, data quality checks, and archiving of the data. Any changes made to the archive channel-level data set are captured, and appropriate updates to the pre-aggregated data set are made in-place to the existing pre-aggregated data set. At the end of each calender quarter, a "snapshot" copy of the data set as it exists at the end of the quarter is made and distributed for analytical use on several computers.

Pre-Aggregated Data From The User Perspective

The information derivable from the data is directly related to the aggregation period. The three figures below illustrate this point. All three figures show the water heating load for the same residence. Figure 3 shows the hourly values for one week of data. The days are separated by vertical lines, with the y-axis being power. The tall spikes show periods of large hot water use at specific times on specific days. The small spikes are standby losses or standby losses combined with low levels of water use. Figure 4 shows the monthly profile information for several months from this site. The monthly profile graphic clearly shows the hourly pattern of use characterizing this site's use of hot water. Figure 5 shows the monthly values for 18 months. There is an apparent seasonal variation, with a clear drop in January and February of 1987. Inspection of the "vacancy" data described below shows that the occupants were gone during parts of these months. The combination of the information from different levels of aggregation can give more insight into the actual behavior of the load than the information from any single aggregation period.

Although there are many uses for end-use data, experience has shown that the needs of most data users can be satisfied with a limited number of standard aggregations. The ELCAP data set presents data as hourly, daily, monthly and monthly profile aggregations. These data sets represent time series data containing measured values for each hour, day or month respectively. The monthly profile data sets contain the mean daily profile (24 hourly values) for each month, one profile per calender month. Generally speaking, other levels of aggregations can be created much more efficiently from the aggregated data than from the raw hourly, 5-minute or 15 minute data. For instance a single yearly value can be calculated much more quickly from the 12 monthly values in that year than from the 8760 hourly values. Other aggregations could conceivably be added to the pre-aggregated data (such as weekly, seasonal, or annual totals), but the principle remains the same: the data should be presented to the user in a form as close as practical to the form in which they will ultimately use it. The emphasis in presenting aggregated data is on reducing the volume of data which must be manipulated by the analyst. The reduction in data set size and pre-processing of the data allows quicker analysis and makes possible interactive and exploratory analysis, with the direction of the exploration determined as the work progresses.

The argument for presenting the data user with data at high levels of aggregation should not be taken as an argument for not collecting data at a low level of temporal aggregation. Most ELCAP data is collected hourly, which is required by some data users. Hourly data is also required for the creation of the monthly profiles. Except for the volume of data which must be processed, error checking is also easier at lower levels of aggregations.

In addition to the variation in temporal aggregation, up to six types of summary data are presented for each end-use in the ELCAP daily, monthly and monthly profile data sets. The types of summary data stored with each end-use depends on the end-use category, such as energy, meterological, or wood-stove sensors. Two types of summary data stored with energy end-uses are mean power and total energy. For end-uses which represent a specific piece of hardware, a third data type of summary data, a count of the number of hours of use in the aggregation period, the "on-count", is presented. For instance, at the monthly level, each RANGE end-use has a data set which gives the on-count for each month. A device is considered "on" only when the hourly value exceeds some pre-determined minimum. The on-count does not therefore include periods of low power. In the case of the RANGE end-use, the low power use might represent a clock on the range but not the use of the range. The fourth data type is an indicator of aggregation periods in which a device uses zero power. The fifth data type is a count of the number of valid observations which went into making up that data value. For instance a particular month of data might have 744 hours in the month, but the monthly total may contain only 740 valid hourly observations. The sixth data summary type is again a mean value, but one reflecting the number of valid observations. Criteria have been set for each level of aggregation to specify the percent of missing values allowed in this mean value summary. For instance a monthly value would be produced for a specific month only if more than 96% of the data is present. Threshold values are also set for the total energy summary data at each level of aggregation. Experience has shown that data users prefer the standard aggregations, with its missing value criteria because of the ease of use.

Two behavioral data sets, "vacancy" and "wood stove use" are derived and stored with the aggregated data. The "vacancy" data set indicates whether the occupants were home on a particular day. For some types of analysis, whether or not the occupant is home is significant. The vacancy indicator was developed from an automated examination of several end-uses which reflect the presence of the occupants; water heating, clothes washer, clothes dryer, range and dish washer (when present). The values of the "vacancy" data set indicate whether the occupant was home, gone or probably home if this can be determined from the data. The other behavioral indicator is "wood stove use". The use of wood stoves is significant for many thermal analyses. For the houses with monitored wood stoves (most ELCAP wood stoves are monitored) this vector indicates whether or not the wood stove was used for each hour.

SUMMARY

The importance of data quality in a large metering project cannot be emphasized enough. The experiences of ELCAP in dealing with large amounts of data illustrate that techniques developed for handling small numbers of sites on a one-by-one basis just do not work. Early implementation of carefully thought out, automated procedures is vital if the metering project does not wish to literally drown in a sea of data.

Equally important is the concept of processing or pre-aggregating the raw data prior to analysis. With up to 50% of an analyst's time consumed in simple data manipulations when dealing with raw data, the advantages of pre-processing the raw data to eliminate this activity are obvious. The value of the raw data lies not in the data itself, but in the information content of the data. Significant emphasis in a metering project should be placed on the development of software that automatically extracts the information content from raw data.

These methods were developed in ELCAP at the cost of years of work and millions of dollars. Although many of the ELCAP data quality checks were developed at the beginning of the project, most of the ELCAP system represents years of development and refinement. While the specifics of the ELCAP system may not be applicable to all large metering projects, the concepts of automated data quality checking and pre-aggregation of data are (or should be) of great interest to all metering projects.

REFERENCES

Baker, M.H. 1984. End-Use Load and Conservation Assessment Program: Commercial Sector Sample Design. DOE/BPA/13795-1. Pacific Northwest Laboratory, Richland, Washington.

R.S. Crowder III and N.E. Miller. 1987. The Quality of the ELCAP Engineering Data Set. PNL-6203. Pacific Northwest Laboratory, Richland, Washington.

Laufmann, S.C. and R.S. Crowder III. 1987. LVA: A Knowledge-Based System for Diagnosing Faults in Digital Data Loggers. In the Proceedings of the Third Conference on Artificial Intelligence Applications, Institute of Electrical and Electronics Engineers, Washington, D.C.

Parker, G.B. and L.O.Foley. 1985. Development of the Sample for ELCAP Monitoring in RSDP Homes. DOE/BPA/13795-5. Pacific Northwest Laboratory, Richland, Washington

Pearson *et al*. 1985. EPRI Metering Workshop Papers- End-Use Load and Conservation Assessment Program. DOE/BPA/13795-3. Pacific Northwest Laboratory, Richland, Washington.

Stoops, J.L. 1988. Managing Adequate Data Capture Rates from Remote Sites or How Everything Can Break or Change and What To Do About It. presented at the ACEEE Summer Study on Energy Use in Buildings, Pebble Beach, California, August, 1988. American Council for an Energy-Efficient Environment, Washington, D.C.

Tomich, S.D. and G.J. Schuster. 1985. End-Use Load and Conservation Assessment Program: Description of Field Data Acquisition System for Building Energy Monitoring. DOE/BPA/13795-4. Pacific Northwest Laboratory, Richland, Washington.

Windell, P.A. 1987. End-Use Load and Conservation Assessment Program: Residential Sector Sample Design. DOE/BPA/13795-16. Pacific Northwest Laboratory, Richland, Washington.

Project Activity	Applicable QA or Data Access Function
Site Specific Measurement Plan Development	Internal review of site installation plan
Installation of Metering Equipment	Sum-check on standard load Sum-check on building load
Data Collection	
Initial Verification	Sum-check on building load Reasonableness test on end-uses
Routine Verification	Sum-check on building load Encoding of data quality flag
Aggregation of Channels to End-Uses	Production of end-use data
End-Use Level Quality Assurance Checks	Sum-check of building load Range checks Repeated value checks
Pre-Aggregation Process	Production of standardized user-oriented data sets

Figure 1. Major quality assurance activities in ELCAP.

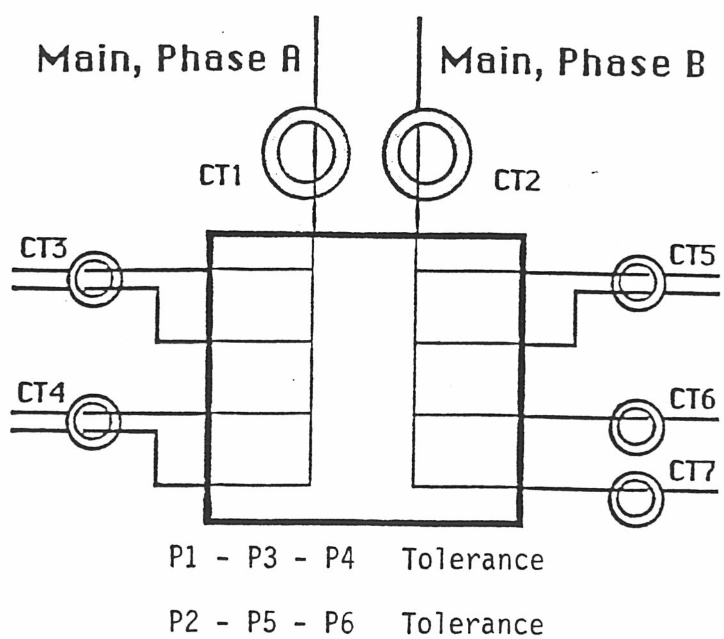

Figure 2. Simplified ELCAP metering configuration.

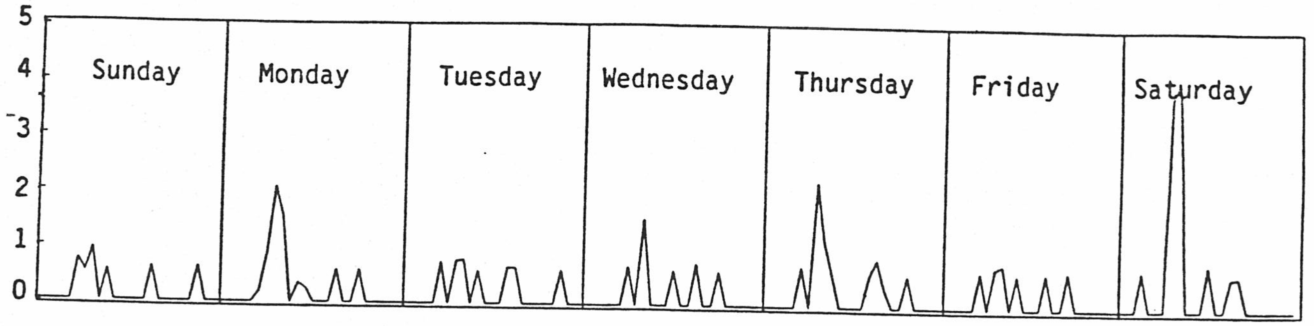

FIGURE 1. Week of Hourly Data (kW)

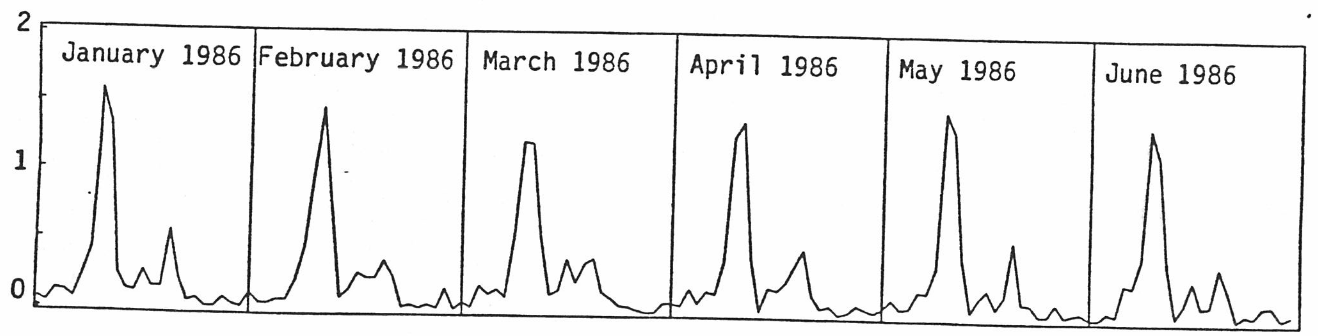

FIGURE 2. Six Monthly Profiles (kW)

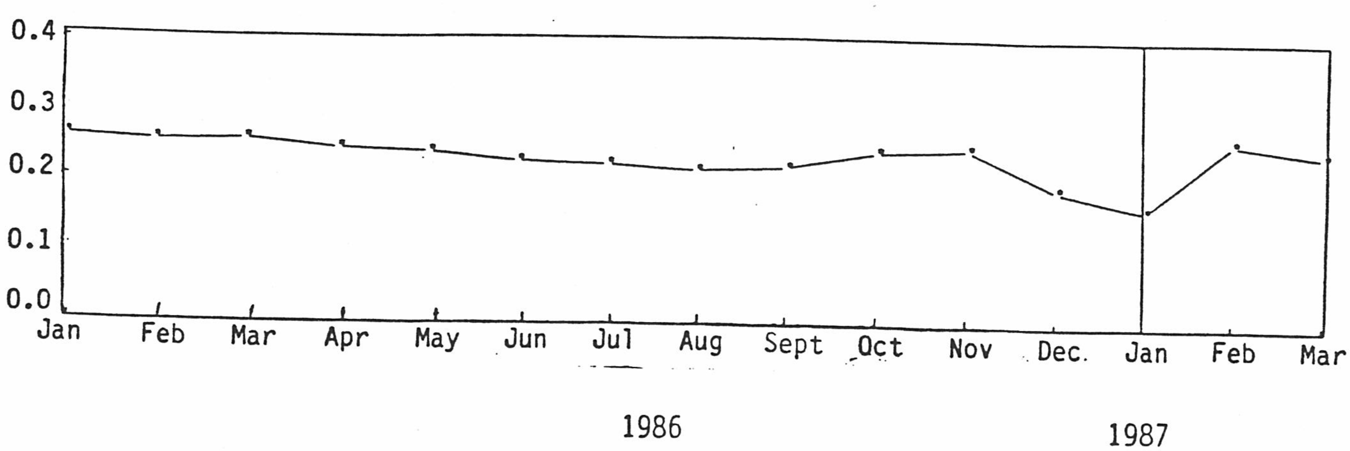

FIGURE 3. Monthly Data (kW)