PNL-6203
UC-350

18 2

# The Quality of the ELCAP Engineering Data Set - Background Issues

R. S. Crowder III
N. E. Miller

June 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
Operated for the U.S. Department of Energy
by Battelle Memorial Institute

Battelle

## 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 | | |

PNL-6203
UC-350

# THE QUALITY OF THE ELCAP ENGINEERING DATA SET - BACKGROUND ISSUES

R. S. Crowder III
N. E. Miller

June 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

The Bonneville Power Administration (Bonneville) began the End-Use Load and Consumer Assessment Program (ELCAP) in 1983. Prior to beginning ELCAP, there was an abundance of information regarding total power consumption for residential structures in the Pacific Northwest, through billing records for example, and limited information regarding power consumption by various end uses (such as hot water, heating and cooling). This program, conducted for Bonneville by the Pacific Northwest Laboratory[a], involves collecting and analyzing hourly end-use data in commercial and residential buildings in the Pacific Northwest.

The purpose of this document is to provide background information to analyses that may use ELCAP data. In general, the ELCAP data set is extremely high in quality, but analysts should be aware of potential problems that could exist with a data set of this size.

This report describes the quality of the ELCAP data and emphasizes the guidelines for data review along with limitations and suggestions regarding engineering and characteristics data including missing data values, procedures for time-stamp assignments, and incomplete integration periods.

[a] The Pacific Northwest Laboratory (PNL) is operated for the Department of Energy (DOE) by Battelle Memorial Institute (BMI) under Contract DE-AC06-76RLO 1830.

## CONTENTS

## FIGURES

# INTRODUCTION

The two data types contained in the End-Use Load and Consumer Assessment Program (ELCAP) collection are engineering and characteristics data. This document emphasizes the engineering data, which is the record of the actual energy consumption for each installed channel in a given logger over a specified period of time. This document also emphasizes meteorological data collected in conjunction with the engineering data from selected sites.

The second type of data, the characteristics data, are used to control and interpret the engineering data. This type of control data includes 1) what each channel represents, 2) the relations that describe which individual channels should be combined into one end use, 3) which channels represent redundant measurements that can be used for quality control checks, and 4) the gross verification status of the data.

## DATA ANALYSES

This section describes how engineering data are acquired, processed, stored, and verified from the perspective of the analysis affected. Separate topics include guidelines pertaining to missing data values, procedures for time-stamp assignments, data conversion from digital signals to engineering units, and implications of hardware design and installation processes.

The basic engineering data set is stored as a huge collection of time-series records. The data for each logger are stored in individual files containing a week's worth of data for all sensors in that logger. Within each file, data are stored in a two-dimensional array with each column representing the time-series data for one sensor channel, and each row corresponding to the readings of all sensors for a particular time period. These files are stored using a special compressed format that allows for approximately six times as much information as would be possible using the standard American Standard Code for Information Exchange (ASCII) format. This special identity format still affords access times comparable to the standard ASCII format. The data records are recorded with an integration period of 5, 15, or 60 minutes. The vast majority of the data are collected using a 60-minute integration period. The energy consumption data are presented in units of average watts (W), and the meteorological data is reported as the average of the appropriate metric system unit.

### LOGGER POWER MEASUREMENTS

Electrical power is the product of voltage and current. The ELCAP loggers explicitly measures both of these quantities and multiplies them together for each monitored channel; the resulting value represents the actual amount of true power consumed by the monitored devices. The current is measured by a current transformer which produces a signal proportional to the current flowing through the monitored wires. The voltage is measured with a voltage transformer that, in turn, produces a signal proportional to the line voltage. The current is measured separately for each sensor channel. The voltage is measured only once for each phase of the electrical service for the entire logger.

Because a logger reports only 256 different values or counts, it is important that we be able to configure the instrumentation in such a way that each count would represent a different number of watts, depending on the types of circuits monitored. There are three ways to alter the individual channel sensitivity in an ELCAP logger. The first is to choose a current transformer that matches the rated capacity of the monitored circuits. The second is to change the scaling resistor used to condition the signal produced by the current transformer. The third is to change the transformer used to sample the voltage of the building's electrical service.

## LOGGER AVERAGES

Once the sensors have been installed, the logger will record a reading once every second and convert it into the correct number of counts. These 1-second readings are then combined for the duration of the integration period. For instance, if 3600 readings are combined for an hourly reading at the end of the integration period, the binary sum can be truncated so that it is less than eight digits long. This procedure is the same as dividing X-number-of-digits truncated by two. In a situation using hourly records, 12 digits are truncated, or the total of the 3600 readings is divided by 4096. Averaging by truncation introduces a small amount of systematic error but never exceeds -2%.

The truncated number is stored in the logger memory and subsequently transmitted to the laboratory where it is converted into engineering units by multiplying by the number of watts per count. Correctly converting the transmitted number into engineering units is of the utmost importance. The resolution of each channel depends on the current transformer size, the scaling resistor, the sample voltage transformer, and the ratio of the truncation to seconds during the integration period. Figure 1 is a schematic of a typical sensor channel.

An example may be helpful to further explain this concept. Suppose that there was a hot water heater attached to both legs of a single-phase, 50-ampere (A) circuit breaker. The ELCAP instrumentation would be installed using a 100-A circuit transformer for each phase. The circuit transformer would be attached to two different channels, and both channels would be scaled

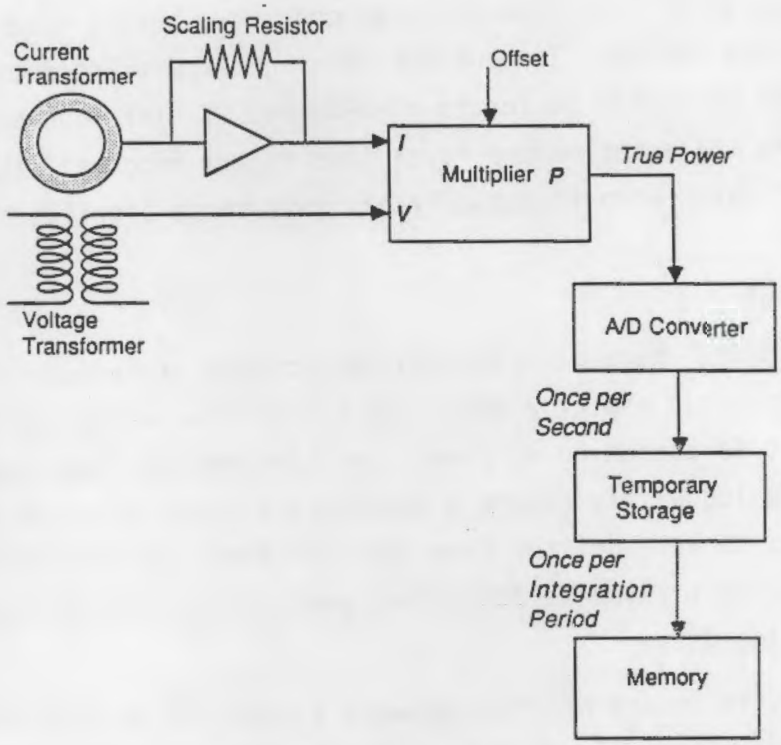


FIGURE 1. Typical Sensor Channel Schematic

to read full-scale at 50 A. The channel resolution would be 21 W/count for hourly data (110 V X 50 A = 5500 W; 5500 W/256 counts = 21 W/count). The resolution could be made even more precise by scaling the channels for a full-scale reading of 30 to 40 A. However, ELCAP installation protocol requires the channel to be scaled to the maximum circuit breaker rating, even though the resulting channel resolutions are larger than may be absolutely necessary.

Another important logger aspect that affects engineering data quality is the offset applied to each channel within the logger. This offset assures the proper function of the electronics. The offsets are individually calibrated for each channel before the logger is sent to the field.

All meteorological data (with the exception of wind speed) is recorded as an analog signal within the logger and accumulated and truncated to form the data record, with the greatest concern being temperature data accuracy. Prior to March 1986, temperature data was reported in integer degrees Kelvin (K).

Data collected after that date has been converted taking advantage of the full precision of the sensor. To continue storing temperature data in an integer format, it was converted to tenths-of-a-degree K. This conversion caused problems; data collected before March 1986 is now reported with four significant figures when it actually has only three significant figures.

## HARDWARE DESIGN

During initial data verification inspections, a hardware design flaw was noted. The logger's use of a soft zero has a major effect on data quality. When a circuit is consuming no power, and the data has been correctly converted, the logger may report a consumption count of +1, 0, or -1. If the offset appears to have drifted from the laboratory calibration to a different value, a new offset value is determined and used to convert future data and correct existing data.

The negative impact of this hardware limitation on specific analysis tasks could be quite profound yet easily minimized by redesigning the analyses to be less sensitive to the minimum-time series value. Given this limitation, an example of a poor analytical approach for looking at air-conditioner usage would be to assume that any time an air-conditioner end use was greater than 0%, the air conditioner was on. Thus, for some loggers in the sample, it would be possible to deduce that an air conditioner was on 20% of the time in the month of January. Closer examination of the data would reveal that the reported load was, at most, only 10 W, which is an indication that the channel was reading +1 count instead of 0% for 20% of the time.

Of course, because the calibrated channel-specific offset value is truly offset and channel applied all the time, if the channel's 0 was off by one count, then all readings for that channel would be off by one count during the period of upward drift.

The engineering data itself is accurate from 2% to 4% of the full scale (255 counts). This maximum error figure integrates the accuracy of the sensors, electronics in the hardware, the truncation in the logger's firmware, and the conversion to engineering units in the laboratory by the central computer. This accuracy varies from channel to channel, depending on the number of watts represented by an individual count. In terms of precision, the average error between the main feed to an electrical panel and the sum of its feeder channels is actually close to one-half of 1% of full scale.

## DATA VERIFICATION

Verification procedures are designed to address most aspects of the engineering and control data systems. The emphasis is placed on proper equipment installation and appropriate conversion parameters for individually metered channels. These procedures have proven to be highly accurate and reliable within their own limitations. Once analysts are made aware of these limitations, they can initiate appropriate safeguards against data bias.

Data verification is accomplished through internal comparison of redundantly metered data. For each electrical panel or set of electrical panels, the main feed is explicitly metered, as are the individual circuits. Figure 2 shows a simple two-phase electrical panel using the ELCAP protocol. By taking the difference between the readings for the main channel and the sum of the readings for all the feeder channels, through a process called sum checking, we can determine if the instrumentation has been installed properly. The difference should be close to 0 for a properly installed site. The difference is not exactly equal to 0 because of the data digitized from individually metered channels. The sum-check equations used for the example panel are

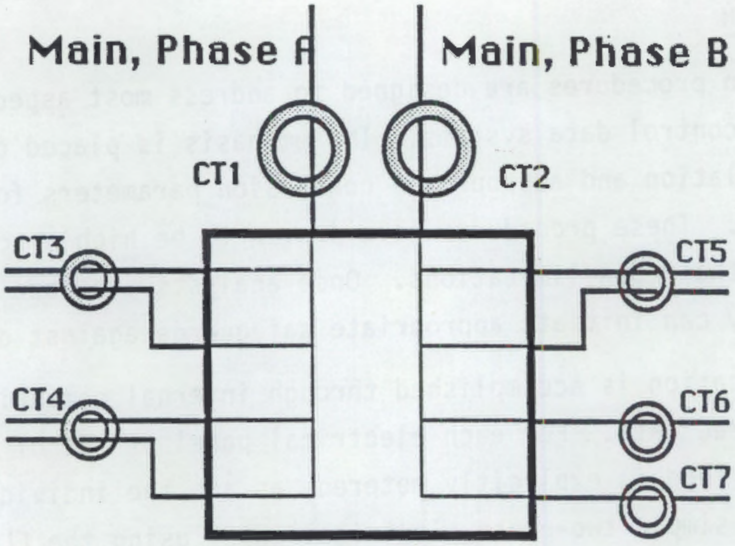


FIGURE 2. Simplified Two-Phase Electrical Panel Using ELCAP Protocol

$$|P1 - (P3 - P4)| < \text{tolerance-A Phase}$$
$$|P2 - (P5 + P6 + P7)| < \text{tolerance-B Phase}$$

where Pn represents the power measured by CTn and tolerances A and B are determined by the resolution of channels 1 and 2, respectively.

If the difference is not close to 0, the verification team attempts to identify the reason for the discrepancy. The cause of the error is then corrected, either by software modification or by a site visit, before data are made available to the analysts. The ability to identify errors in the data depends on the actual data installation.

In addition to making sure that the engineering data are internally consistent, the verification team checks to make sure that equations used to produce the end-use data from the channel data are correct. This ensures that the labels assigned to the individual channels agree with the measurement plan, and that the parameters used to convert the digital data to engineering units are correct.

While this approach has proven to be an incredibly powerful tool, some limitations are noted. Because of physical restrictions, the main power circuits were not monitored; thus, it was not possible to compare the feeder channels to any redundant measurement. Sites where it is impossible to compare redundant measurements for all channels are denoted with a different status. About 8% of the loggers have one or more channels that cannot be sum checked. Despite these limitations, one group of residential data analysts, who were very careful in screening their data, were able to use over 95% of the data that the verification team had labeled as being ready for analysis.

One of the project's unique features is the built-in redundancy monitoring of both the mains and the feeders in the ELCAP installation protocol. For the bulk of the sites, where such redundant measurements have been feasible, this sum checking facilitates initial data verification and holds the key to ongoing and automated quality data assessment for the energy channels.

## GENERATION OF TIME STAMPS

Every engineering data record has a time stamp denoting when the data were collected. This time stamp records the end of the period that the data value covers. At the beginning of the data logger processing chain, the time stamp is the number of seconds past midnight that occur at precisely the moment the data records are written into the logger's memory. The logger clocks are set according to Greenwich mean time (GMT). Each time the logger is interrogated, its clock is compared to GMT. Should there be a discrepancy of 2.5 minutes or greater, the logger clock automatically resets itself by a command from the data acquisition computer.

Consider an example. Suppose that the logger has parameters sent to it, instructing that a record be collected every hour. The logger will, at the end of each hour, precisely according to its own internal clock, write a record to its memory. If this occurred at logger time 8:00 a.m., the time stamp on the record would read 28,800 seconds. The data record, however, represents the accumulation of data scans from logger time 7:00 a.m. until logger time 8:00 a.m. If the logger clock is keeping GMT time correctly, the data record is synchronized to both its own internal clock and GMT.

## INCOMPLETE INTEGRATION PERIODS

Another small but bothersome problem with the data is caused by an incorrect number of scans (1-second readings) being included in a data record. This can only happen during the integration period when the logger clock is being resynchronized to GMT. The result is that a record is produced that contains too many or too few scans for the integration period. Because the logger software assumes that all integration periods have the correct number of scans, it will truncate the wrong number of digits. If too few scans are included, the record appears to have very low or negative readings. These negative readings occur when the summed reading is less than the offset sum for the full integration period for the energy channels. If too many scans are used, all channels will report values that are considerably higher than the actual consumption.

In May 1986, a program was run to remove all the records suspected of incomplete integration periods from the data set. This process removed less than 0.02% of the total records from the data set; however, the discrepancies in the records with too many scans were large enough to significantly bias the analyses towards those analysts who had not screened their data. Also implemented were changes in the data processing software. These changes allowed records to be removed from the files as data were converted from raw units to engineering units. Any analyses tasks using channel-level data before May 1986 should be reextracted or examined for the removal of potentially contaminated records. Any channel-level data extracted after May 1986 will not manifest this problem.

## POWER OUTAGE EFFECTS

If the power fails at a site, the logger will stop operating. Once the power is restored, the logger clock has had both its day and post-midnight seconds set to 0. Precisely 1 hour after power restoration, (if the logger was on hourly integration), the first record would be written to memory.

Although this record represents an integration of exactly 1 hour, it has a very small chance of accurately representing a sum of scans going from 1 GMT hour to the next. An accurate representation would have only occurred when the power outage was exactly an integral number of hours long. Consequently, for data records written after a power outage, the number of post-midnight seconds will not appear synchronized with GMT clocks.

## TIME STAMP CONSTRUCTION

When the raw data are brought into the laboratory, the time stamp is

- converted to Pacific Standard Time
- used to construct a minute block representing the nearest number of minutes since midnight Pacific Standard Time
- used as a data quality flag encoded with information regarding whether or not this record was made after a power outage. It also indicates how close the record was to GMT synchronization.

In all versions of the computer programs SELECT and EASE, the minute blocks, modified in accordance to the analyst's time zone choice, are part of the output file produced. The post-midnight GMT seconds have never been part of the output file or the EASE 2.0 output. Likewise, the data quality flags have never been part of the output file. EASE 2.0 (or a future version) will take the data quality flag into account when deciding whether or not a record is suitable for aggregation. The specifics of the algorithm and timetable have not yet been worked out.

## MISSING DATA

Missing values are an inevitable result of any data collection effort the size and complexity of ELCAP. To understand how missing values will affect analyses, it is important to understand the cause and frequency of the different types of missing data. It is also important to understand how the various extraction tools treat missing values.

The three different types of missing values in the ELCAP data set are

1. isolated records with missing values - These are usually caused by clock resets, random communication failures, or local power outages.

These blocks of missing data are typically only one or two records long, and the blocks appear to be distributed randomly over time.

2. portions of months ranging from 20 to 200 continuous hours - These blocks of missing data are usually the result of the loss of a data deposit from the logger's memory. The data can be lost because of intermittent modem failures, loss of logger parameters, or communication problems. These types of problems affect less than 0.5% of the data dumps, but they tend to affect some loggers more than others.

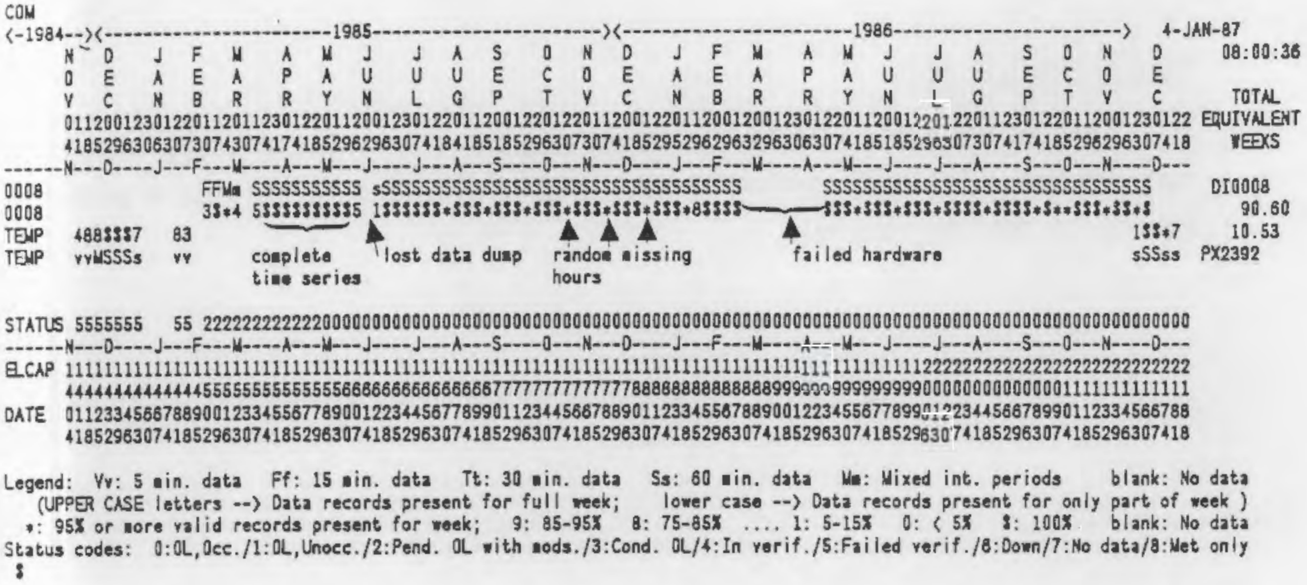
3. extended periods of missing data ranging from 200 hours to several months - Bad modems or other hardware failures have been the principal causes of these more extensive blocks of missing data.

The discussion of missing values refers only to entire records or time periods. Current ELCAP data collection protocol requires that readings from all active channels must be part of the end use or the entire record is archived. Normally, there are not cases where, for a particular collection period, there are values for seven channels and no values for six channels. However, there are changes in the number of sensors that are installed at the site.

To get an idea of which missing data values need to be of concern, the logger status information provided by the ELCAP data management tool EASE can be used. EASE displays a graphical representation of the amount of data available for a particular logger over a particular length of time. Figure 3 is an example of the type of information provided through the EASE status option.

The next matter of concern is how the extraction tool will treat missing data values. Extraction simply produces a data set that is a complete time series, with each row representing a measurement at some unique point in time. An extraction might consist of a 5-minute, hourly, monthly, or even yearly data accumulation.

The simplest extraction is just a record-by-record copy from the original database. If the data of interest has a lower time resolution (i.e., a 5-minute resolution instead of 60-minute resolution) than that used to collect the data, it is necessary to combine two or more data records to produce the requested data record.



```
COM
<-1984--><-----------------------1985-----------------------><----------------------1986----------------------->   4-JAN-87
       N  D    J   F   M    A   M   J    J   A   S    O   N   D    J   F   M    A   M   J    J   A   S    O   N   D       08:00:36
       O  E    A   E   A    P   A   U    U   U   E    C   O   E    A   E   A    P   A   U    U   U   E    C   O   E
       V  C    N   B   R    R   Y   N    L   G   P    T   V   C    N   B   R    R   Y   N    L   G   P    T   V   C       TOTAL
       01120012301220112011230122011200123012201120012201220112001220112001200123012201120012201220112301220112001230122 EQUIVALENT
       41852963063073074307417418529629630741841851852963073074185295296296329630630741851852963073074174185296296307418  WEEXS
------N---D----J---F---M----A---M---J----J---A---S----O---N---D----J---F---M----A---M---J----J---A----S---O---N----D---
0008                FFMm SSSSSSSSSSS sSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS         SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS    DI0008
0008                3$*4 5$$$$$$$$$$5 1$$$$$$*$$$*$$$*$$$*$$$*$$$*$$$*8$$$$$----------$$$*$$$*$$$*$$$$*$$$$*$**$$$*$$*$      90.60
TEMP   488$$$7   83                                                                                       1$$*7     10.53
TEMP   vvMSSSs   vv       complete     lost data dump    random missing        failed hardware              sSSss   PX2392
                          time series                    hours

STATUS 5555555   55 22222222222200000000000000000000000000000000000000000000000000000000000000000000000000000000000000
------N---D----J---F---M----A---M---J----J---A---S----O---N---D----J---F---M----A---M---J----J---A----S---O---N----D---
ELCAP 1111111111111111111111111111111111111111111111111111111111111111111111111111111111122222222222222222222222222222
      4444444444444555555555555555666666666666666677777777777777888888888888888999999999999999000000000000000001111111111111
DATE  0112334566788900123345567789001223445677899011234456678890112334556788900122345567789901223445667899011233456678
      4185296307418529630741852963074185296307418529630741852963074185296307418529630741852963074185296307418529630741 8

Legend:  Vv: 5 min. data   Ff: 15 min. data   Tt: 30 min. data   Ss: 60 min. data   Mm: Mixed int. periods     blank: No data
   (UPPER CASE letters --> Data records present for full week;    lower case --> Data records present for only part of week )
  *: 95% or more valid records present for week;   9: 85-95%   8: 75-85%  .... 1: 5-15%   0: < 5%   $: 100%   blank: No data
Status codes:  0:OL,Occ./1:OL,Unocc./2:Pend. OL with mods./3:Cond. OL/4:In verif./5:Failed verif./6:Down/7:No data/8:Met only
 $
```

FIGURE 3. EASE Status Option - Types of Missing Data Found in the ELCAP Collection

ELCAP engineering data can be extracted two ways with respect to metering levels. The two methods of extraction capture data either at the channel level or end-use level. The channel-level data are the most disaggregated form available. Within each data record, a data point corresponds to a specific sensor channel within the data logger itself. The channels often monitor a single phase of a device or group of similar devices. On the other hand, end-use-level data are derived by summing the readings from all channels that monitor devices with a similar end use.

As an example of the channel versus end-use level of extraction, suppose that a building has four, three-phase heat pumps, each dedicated to one-fourth of the building's floor area. The measurement plan may show that channels 25, 26, and 27 monitor A, B, and C phases of the southeast and southwest heat pumps, respectively; and channels 28, 29, and 30 monitor the A, B, and C phases of the northeast and northwest heat pumps, respectively. The extraction of channel-level data would allow comparisons of loading between the A, B, and C phases of the pairs of heat pumps. Extraction at the end-use level would result in a high-voltage alternating current (ac) HVAC value equal to the sum of the readings on channels 25 through 30.

## METEOROLOGICAL DATA

The meteorological data collected as part of the ELCAP study cannot be sum checked. Instead, reasonableness checks are used to provide a primary source of quality control.

## CHARACTERISTICS DATA

This section discusses the characteristics data used to control the extraction of the engineering data. Another type of characteristics data, which will not be addressed in any detail in this document, is the characteristics data used to describe the physical and economical characteristics of the buildings and their occupants. However, the physical and economical characteristics data are subject to two types of errors: data entry and incorrect information from the survey instrument itself.

### CONTROL DATA TYPES

Four types of data are in this collection. The first type of data describes what each energy channel is measuring. This includes the identification of the channel, the parameters used to convert the raw data from digital counts to engineering units, and a field stating whether or not the channel has been verified. The second type of data is information about which individual channels should be added together to produce a specific end use. The third relation contains the sum-check equations used to verify the data. The final relation contains information about the verification status of the logger and the time resolution of the data.

These data are used to control the processing and extraction of the engineering data. It is possible that incorrect results may be generated at extraction time, not because the underlying engineering data are flawed, but because the characteristics data used to control the engineering data extraction are incorrect. Discussed below are some of the problems with the control data and potential ways in which to identify the problems.

### DATA TIME-STAMPING CHARACTERISTICS

Originally, all control information was contained in non-time-stamped files that were keyed to particular loggers. However, it soon became obvious that the information would change over time as building owners modified their buildings, instrumentation errors were corrected, or the sensor complements at the building were changed. Therefore, the information was moved into the more flexible form of a relational database. Now all pertinent parameter sets are available at all times making it possible to analyze data available for a logger even with two or more changes in the metering system at the site.

If the time stamps on the control data are incorrect, one may extract data that are incorrect or be denied data access altogether. This is especially problematic when end-use assignments change. For instance, suppose that the air-conditioning end use should be c25 + c26 from May 1 until X date, but should be c30 + c31 for any data before May 1. If the control information were not entered for the earlier parameter set, and the data are extracted for an entire calendar year, incorrect values will be returned for the dates from January through May.

## DATA REVIEW

Analysts should understand the nature of the data they are working with. The redundant ELCAP measuring protocol has made the automated data quality checks possible for the bulk of the energy data.

The ELCAP data consists of engineering and characteristics data. The engineering data are made up of energy channel data and meteorological data. The characteristics data makes the interpretation and checking of the engineering data possible. In order to interpret and check the engineering data, the ELCAP analyst needs to

- understand the developmental state of each type of data used
- understand the potential problems of each data type
- report any anomalies found.

For example, a characteristics database data entry error could give an analyst wind-speed data when wood-stove data was requested. Because data has not been historically sum checked, an installation problem or hardware failure may be apparent in current data. The data review plan for any research project is just as important as the analysis plan. The extent to which the data review is required will be determined by the sensitivity of the particular analysis-to-data aberrations.

The initial verification process--bringing sites on line--has provided an amazing track record of predicting future data quality. During the Residential Standards Demonstration Program (RSDP) and Residential Base (RESBASE) thermal analyses, very few sites were removed from data quality degradation analyses once they were brought on-line. Few installation errors or hardware failures have been found to date that were not identified in the initial 1 to 1-1/2 week inspection comprising the initial verification process.

ELCAP data quality is quite high, even when considering its present stage of development. This, however, does not free the analyst from the responsibility of understanding the data and reviewing the data that a professional reputation may rest on. The budget for data analysis should always include the budget for data review.

## 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

S. G. Hauser
Battelle Portland Office
500 NE Multnomah, Suite 650
Portland, OR 97232

No. of
Copies

W. M. Warwick
Battelle Portland Office
500 NE Multnomah, Suite 650
Portland, OR 97232

### ONSITE

13 Pacific Northwest Laboratory

R. G. Pratt
W. F. Sandusky (5)
G. M. Stokes
Publishing Coordination
Technical Report Files (5)