PNNL-24331

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Simplified Processing Method for Meter Data Analysis

November 2015

KM Fowler
AH Colotelo
JL Downs
KD Ham
JW Henderson
SA Montgomery
SA Parker
CR Vernon

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

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 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov

This document was printed on recycled paper.
(8/2010)

PNNL-24331

Simplified Processing Method for Meter Data Analysis

KM Fowler
AH Colotelo
JL Downs
KD Ham
JW Henderson
SA Montgomery
SA Parker
CR Vernon

November 2015

Prepared for
the U.S. Department of Army
under U.S. Department of Energy
Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Contents

Figures

Tables

1.0 Introduction

This report documents a simplified data processing method that can be used to identify and resolve suspect and missing metered electricity data so it can be used to analyze building energy use. Energy managers need data analysis and management tools that can be consistently applied to metered energy data for energy data evaluation and reporting. The Meter Data Management System (MDMS) is the Army's enterprise system for energy and water data management. MDMS does not currently offer users exports of processed metered energy data, which means users must apply their own methods when analyzing the data. A standard data processing method that could be used directly by energy managers and/or one that is consistently applied within MDMS could offer energy managers a mechanism to quickly assess the energy use of multiple buildings, and minimize the need for energy managers to manually process the energy data.

For this document, the phrase 'data processing' is the method used to adjust, clean up, or fix suspect building interval meter data so that it is usable for typical utility data analysis techniques.

The data processing method defined in this report was developed by Pacific Northwest National Laboratory (PNNL) for the Deputy Assistant Secretary for the Army (DASA). The method was derived with electricity data and may be useful to energy managers that frequently analyze metered energy data. *The data processing method is designed for annual or monthly consumption analysis or to create daily and monthly load profiles of the interval consumption data.*

First-hand knowledge of a building's operations may result in the need for additional data processing by an analyst familiar with the building. However, data processing consistency can be improved and staff time limitations can be minimized by using a standard method to address more commonly found suspect and missing metered data.

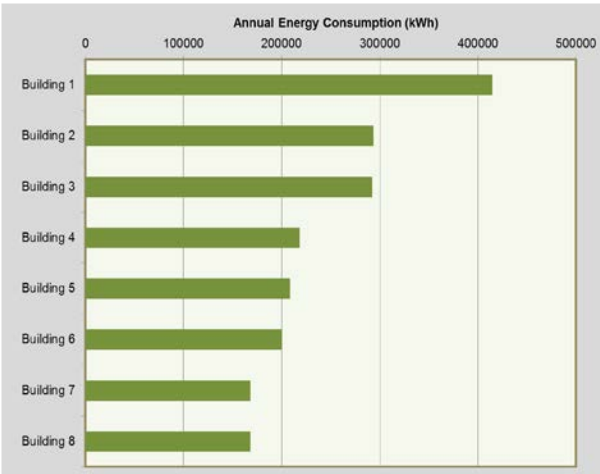
1.1 Rationale for Processing Metered Energy Data

Occasional data quality issues are inevitable during meter interval data collection. There are many different components in a metering system ranging from the metering device to communications to the data acquisition system (system components described in more detail in Appendix A: Metering System Architecture) and suspect or missing data can occur at any component within the metering system. The most common reasons for suspect or missing data include:

- Loss of power or service to the building, meter, or data communication device
- Meter, sensor, or communication errors because of broken equipment
- Meter, sensor, or communication software malfunctions
- Meter, sensor, or communication system replacement or reset/restart

Due to the inevitability of data quality issues, raw metered data must be processed to be useful for analysis (ERCOT 2010a). Data processing includes a data quality check and addressing data quality issues when they are relevant to the planned data analysis. The risk of using unprocessed data for analysis is that suspect or missing data may skew consumption estimates and result in an incorrect assessment of a building's energy needs.

Total annual energy consumption and energy use intensity (EUI) are commonly used to evaluate the annual consumption for typical and similar buildings (Figure 1-1). To evaluate energy use and prioritize actions to improve energy efficiency, the calculated annual values must represent the full energy use for the year. If data are erroneous or missing, a building's energy use may be under- or overrepresented, implying that the building is operating more or less efficiently than it actually is. Underestimating energy use may also lead to discrepancies with energy costs, or fuel usage if electricity is supplied from a generator.



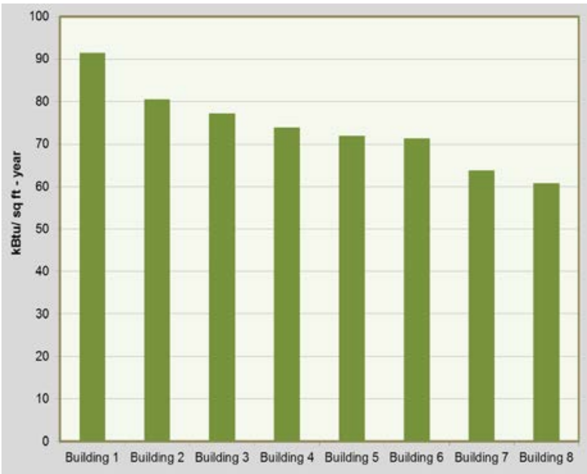


Figure 1-1. Total annual energy consumption and energy use intensity for eight similar buildings using processed data.

Figure 1-2 displays the unprocessed annual energy consumption and EUI for the same buildings shown in Figure 1-1. The annual energy consumption and EUI calculated with unprocessed data was extremely high for Building 1 and extremely low for Building 8. Interpretation of the unprocessed data may have led to an understanding that Building 1 was highly inefficient and should be investigated further, requiring additional resources. Conversely, Building 8 appears to be very efficient based on the unprocessed data. Data processing allowed for the suspect and missing data to be addressed enabling a more meaningful comparison of the buildings.

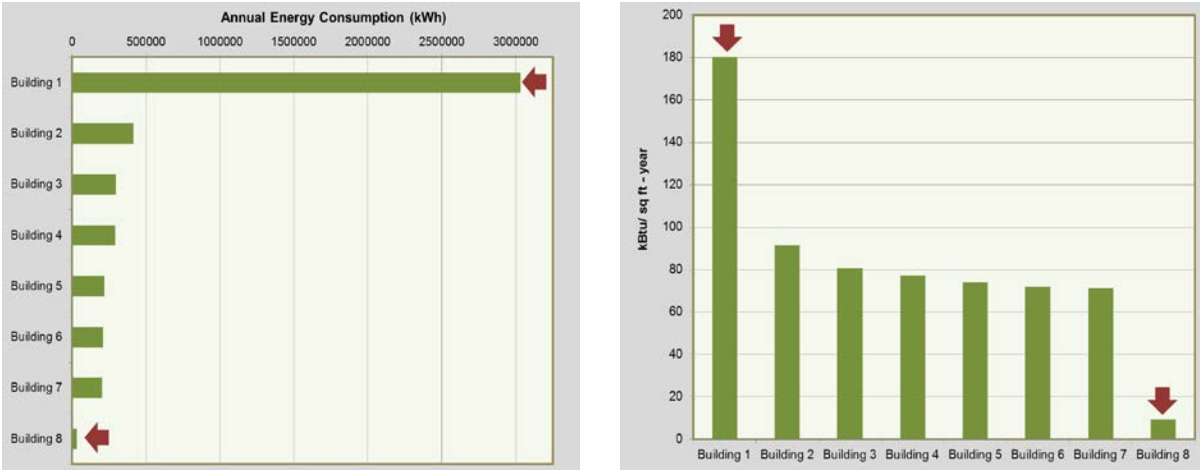


Figure 1-2. Total annual energy consumption and energy use intensity for eight similar buildings using unprocessed data.

Issues with data quality and completeness can also affect the understanding of daily consumption patterns. Figure 1-3 displays a daily load profile with unprocessed (left) and processed data (right; note the significant difference in scale for energy units). The unprocessed data has extreme positive values with corresponding negative values. Using the data processing method outlined in this document, the pattern of energy consumption can be observed based on the day of the week, with notable differences between weekdays and weekends. Figures 1-2 and 1-3 demonstrate why processing of metered energy data is needed before data analysis.

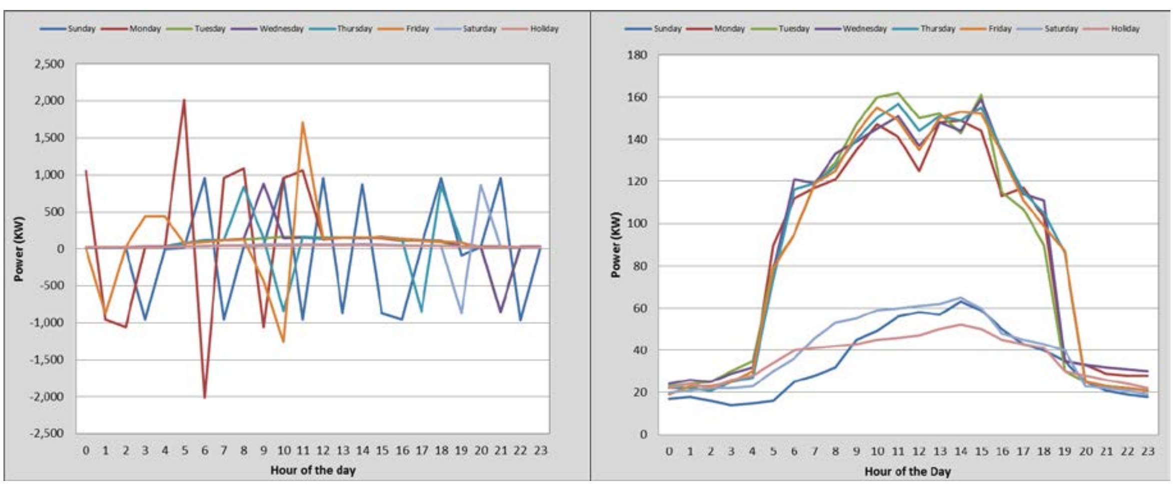


Figure 1-3. Daily load profiles using unprocessed (left) and processed (right) data.

A literature review was conducted to identify the best method for processing metered energy data. The majority of literature discussing suspect and missing metered energy data and data processing methods are published by either utilities or providers of metered data management systems. Utilities use meter data validation rules to ensure meters are operating properly, which minimizes suspect data, thus also minimizing the need for data processing (CPUC 1999a and

1999b; Moore 2008). Data validation, estimation, and editing (VEE) can be used to identify problematic data and address data quality issues in a manner that meets the needs of the data user (Moore 2008).

1.2 Report Organization

This report describes common types of suspect data in the context of typical utility data analysis techniques and provides a data processing method that could be applicable to MDMS datasets. Section 2 describes the first step of the data processing method, which is to identify suspect and missing data. Section 3 contains the data gap filling process that is applied to the data that have been identified and flagged as suspect or missing. A summary of the data processing method and how it could be used within the Army are discussed in Section 4. Appendix A explains the major components of metering systems. Appendix B discusses how utilities manage suspect interval data.

2.0 Identifying Suspect Data

'Suspect data' are metered data that do not conform with what would be expected. Suspect data may include missing data, negative values, or extremely high values.

The first step in data processing is to be sure the dataset being examined is in the expected format, that is, the fields are showing the type of data you would expect, for example, dates in the date field are in the expected style. Once the data format is confirmed, it is essential to determine how you want to use the data, as the type of processing depends on what questions the data will be used to address or what summary will be generated using the data. As mentioned above, the data processing method outlined in this report is designed to quickly get large quantitities of electricity interval data usable for annual or monthly consumption analysis or to create daily and monthly load profiles of the interval consumption data. There are other approaches to checking and fixing data errors or gaps, not discussed in this document, that vary in complexity. The more you know about a building's operations the more you can tailor the processing method. Following this data processing method, an analyst may identify other aspects of the dataset that need to be addressed or modifications which require further examination of the raw data. Whenever possible it is recommended that an experienced energy data analyst examine the dataset and further adjust the estimates as deemed appropriate based on known operations (IMEM 2010).

2.1 Considerations for Data Processing

Before a metered dataset is used for analysis, it should be examined to determine if data processing is needed. The quantity of metered data Army energy managers need to review and the time constraints they operate under makes an automated data processing method highly valuable. An automated method also offers a consistent approach to data processing. This simplified method does not address all possible suspect data scenarios, so it is possible that data processing method could adjust a dataset incorrectly. Ideally an analyst familiar with the dataset will examine the data before and after processing to determine whether the processing method worked as designed.

An interval data value is considered null when no data are present. This can occur when a timestamp is present and no values are recorded, or when an expected timestamp is not present.

Although building-level electricity, natural gas, and water meter interval data were examined during this study, the data processing method was developed using the electricity data, because of the size of that dataset. Despite the study's emphasis on electricity, aspects of the method are adaptable to water and other energy sources. However, if the method is going to be used for other types of interval data an experienced analyst should be used to adapt the method for the specific utility. One example of the differences to be aware of is that zeros or null values are more likely to be valid data for natural gas and water than they would be for electricity.

2.2 Electricity Interval Meter Data Types

Generally, the data stored in the Army's MDMS are collected using totalizing meters, which record cumulative energy consumption totals (kWh). Alternative meter types include interval meters which record consumption (kWh) and/or power demand (kW) over a specific time interval, for example, every 15 minutes. The method outlined in this report has been tested using MDMS data from totalizing meters, site-provided energy data from totalizing meters, and interval data received directly from site data loggers. It is suggested that data processing be conducted using the interval consumption rates and these can easily be calculated for data from totalizing meters.

Table 2.1 shows an example of data from MDMS that was collected using a totalizing meter including a column with the interval consumption calculated by subtracting the previous meter reading from the current meter reading. The interval consumption can then be used for typical data analyses such as annual energy consumption, annual EUI, monthly EUI, and daily load profiles.

Table 2-1 Example Data File from MDMS with raw usage reading from totalizing meter and calculated consumption values.

Timestamp	Raw Usage Reading (kWh)	Calculated Consumption (kWh)
4/25/2011 9:00	5769537	19
4/25/2011 9:15	5769591	54
4/25/2011 9:30	5769647	56
4/25/2011 9:45	5769703	56
4/25/2011 10:00	5769760	57
4/25/2011 10:15	5769838	78
4/25/2011 10:30	5769916	78

2.3 Identifying Suspect and Missing Data

To develop this method, 23 million records of consumption data from more than 1,000 building-level energy meters were reviewed and analyzed. The analyzed data ranged from multiple years of 15-minute interval data for some buildings, and months of data provided directly from data loggers for other buildings. The data came from Army bases located throughout the world. Suspect or missing data were identified and then categorized by type and frequency for each building. Common types of suspect data were summarized.

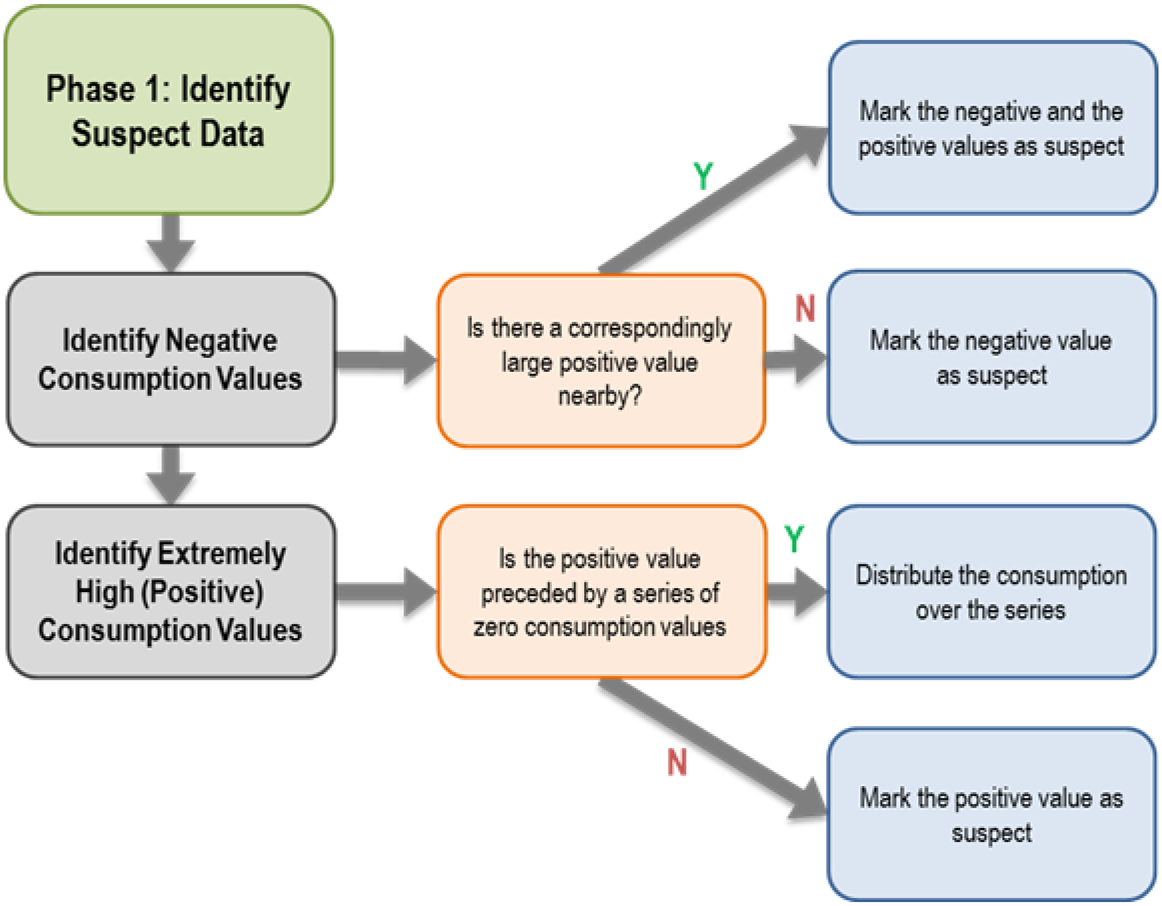


Figure 2-1. Flow diagram outlining the process used to identify suspect data.

There are numerous reasons for suspect or missing data – loss of power, equipment failures, human error, and weather-related failures cover the most common reasons. Table 2.2 describes how the suspect and missing data were categorized by type and frequency for each building and common types of suspect data. Once the suspect or missing data have been categorized they can be addressed.

Table 2-2 Possible Reasons for Suspect and Missing Data in Metered Energy Data

Types of Suspect Data	Possible Reasons for Suspect Data
Negative consumption with offsetting positive consumption	Meter, sensor, or communication errors because of broken equipment or software malfunctions
	Loss of power or service to the building, meter, or data communication device
Negative consumption with no offsetting positive consumption	Meter, sensor, or communication errors because of broken equipment, software malfunctions, equipment replacement, or reset/restart
Large positive consumption values with no offsetting negative consumption	Meter, sensor, or communication system replacement or reset/restart
“0” consumption	Meter, sensor, or communication errors because of broken equipment or software malfunctions
	No electricity use during that time period
Missing timestamps or data gaps	Loss of power or service to the building, meter, or data communication device

An example of the frequency of suspect or missing data from a building meter is presented inTable 2.3. Note that consumption values of zero are not necessarily erroneous values, they are just flagged as suspect given it is unusual for a building to have no energy draw. The most frequent suspect data type for each building was missing timestamps or data gaps.

Table 2-3 Example of Suspect and Missing Data Example Building Dataset

	Timestamps Encountered	Negative Consumption Values	"High" Positive Consumption Values	"0" Consumption Values	Missing Timestamps
n	133587	118	0	8312	9130
%	100.0%	0.1%	0.0%	6.2%	6.4%

As stated previously, every metered dataset is unique. One of the variations can be the type of suspect data, as noted above. Another variation can be on the duration or frequency of the suspect data. Table 2.4 provides an example of how the duration of the data gaps varied for three example buildings.

Table 2-4 Example Summary of the Frequency of Missing Timestamps by Data Gap Duration

Buildings	Timestamps Encountered	Percentage of Timestamps Missed by Data Gap Length <2 hrs	2-4 hrs	4-12 hrs	12-24 hrs	1 day - 1 week	1 week - 4 weeks	>1 month
Building A	1238855	0.15%	0.45%	3.34%	0.21%	2.96%	8.93%	14.49%
Building B	4685798	0.01%	0.00%	0.00%	0.00%	2.50%	0.02%	0.03%
Building C	8912531	0.58%	0.13%	0.14%	0.28%	0.55%	0.13%	5.65%

An example of consumption data with missing timestamps where it is difficult to determine whether there was a meter or communication issue is displayed in Table 2-4. Some of the buildings noted specific issues with their metering systems or communication systems, which can account for the longer periods of data gaps.

Table 2-5 Example of Missing Timestamp Data.

(Timestamps before and after the five-hour fata gap are highlighted)

Timestamp	Raw Usage Reading (kWh)	Calculated Consumption (kWh)
11/1/2012 10:00	90707.59	3.81
11/1/2012 11:00	90711.44	3.84
11/1/2012 16:00	90714.25	2.81
11/1/2012 17:00	90715.55	1.30
11/1/2012 18:00	90716.91	1.36

The proportion of suspect or missing data will vary depending on the building. It is recommended that the analyst summarize the occurrence of these data, before further analysis. Whether or not a user needs to process a dataset by fixing suspect data and filling data gaps depends on how the data are going to be used. **The data processing method presented in this report was designed to process data for annual or monthly consumption analysis or to create daily and monthly load profiles of the interval consumption data.** If an analyst wants to know exactly what the metered values are, no data processing should occur. Similarly, if a detailed understanding of the operations using building automation system data along with metered consumption is the intent, the methods outlined here are not likely to provide the level of detail needed.

For the analysis of metered interval data directly from data loggers, rather than the data available from MDMS, additional information from the meter may be useful in determining whether the data are valid. For example, if the voltage readings from the meter are too high, too low, or zero, the amperage is too high or zero, or the power factor is unreasonable or too low, it may indicate that the consumption values are not reliable and should be marked as suspect. The combination of readings indicating invalid data would vary among meter or data logger types.

For this data processing method we have established that if a dataset for an individual building has more than 25% of the data flagged as processed or estimated, consider whether the data are valid or sufficient to address the specific analysis objectives. This is a judgement call and the key is to document the assumptions, estimation methods, and limitations of the dataset.

2.4 Preparing Suspect and Missing Data for Data Gap Filling

This section describes the steps necessary to prepare the datasets for processing after identifying suspect or missing data. These rules identify areas where data gaps, flagging, or interpolation will occur.

2.4.1 Negative Consumption with Offsetting Positive Consumption

If negative consumption data with offsetting positive consumption data are identified in the dataset, the processing method is to take the sum of the large positive value and negative value:

1. If the sum equals zero or is negative, replace the large positive and negative values with null. Null indicates the value is suspect and should be addressed later in the process. If there are zero value readings between the erroneous values, flag those values as suspect data, creating a data gap.

2. If the sum is positive and the affected duration is 1 hour or less divide the sum by the number of time interval increments between the values and replace zero values with that averaged value (linear interpolation between the points; Table 2.6). Flag these values for additional quality check following any automated processing.

3. If the sum is positive and the affected duration is greater than 1hour, flag this value for additional quality check following any automated processing.

Table 2-6 Example of Original Dataset with Negative Consumption and Offsetting Positive Consumption (left) and the Dataset Prepared for the Gap Filling Process (right).

Timestamp	Raw Usage Reading (kWh)	Calculated Consumption (kWh)
11/4/2012 5:00	3169329	11.50
11/4/2012 5:15	3169340	11.25
11/4/2012 5:30	3169352	-45.25
11/4/2012 5:45	3169409	57.00
11/4/2012 6:00	3169363	-45.75
11/4/2012 6:15	3169420	57.25
11/4/2012 6:30	3169374	-46.00
11/4/2012 6:45	3169432	58.00

Timestamp	Raw Usage Reading (kWh)	Calculated Consumption (kWh)	Data Source
11/4/2012 5:00	3169329	11.50	Collected
11/4/2012 5:15	3169340	11.25	Collected
11/4/2012 5:30	3169352	null	*Suspect*
11/4/2012 5:45	3169409	null	*Suspect*
11/4/2012 6:00	3169363	null	*Suspect*
11/4/2012 6:15	3169420	null	*Suspect*
11/4/2012 6:30	3169374	null	*Suspect*
11/4/2012 6:45	3169432	null	*Suspect*

2.4.2 Negative Consumption with No Offsetting Positive Consumption

If negative consumption data with no offsetting positive consumption data exist in the dataset, the processing method is to replace the value with null and flag it as suspect data.

2.4.3 Large Positive Consumption with No Offsetting Negative Consumption

Large positive consumption values can be defined as consumption values greater than the 75th percentile or values 10 times higher than the mean interval consumption value for the entire dataset. Different datasets may need different boundaries depending on the quality and consistency of the data. The intent is to only eliminate values that the analyst is confident are not valid consumption values.

If large positive consumption data with no offsetting negative consumption data exists, the data processing method is to flag the value as suspect.

1. If the large positive consumption value is preceded by missing data, the consumption can be distributed evenly across those timestamps. Flag these values for additional quality check following any automated processing (Table 2.7). If the distributed consumption value is still 10 times greater than the mean value, replace the value with null. Flag these values for additional quality check following any automated processing.

2. If the large positive consumption value is not preceded by missing data, replace the value with null.

Table 2-7 Example of Original Dataset with a Large Positive Consumption with No Offsetting Negative Consumption (left) and the Dataset Prepared for Data Gap Filling Process (right).

Timestamp	Raw Usage Reading (kWh)	Calculated Consumption (kWh)
10/26/2012 10:00	90707.59	3.81
10/26/2012 11:00	90711.44	3.84
10/26/2012 16:00	90740.55	29.11
10/26/2012 17:00	90741.85	1.30
10/26/2012 18:00	90743.21	1.36

Timestamp	Raw Usage Reading (kWh)	Calculated Consumption (kWh)	Data Source
10/26/2012 10:00	90707.59	3.81	Collected
10/26/2012 11:00	90711.44	3.84	Collected
10/26/2012 12:00	n/a	null	*Suspect*
10/26/2012 13:00	n/a	null	*Suspect*
10/26/2012 14:00	n/a	null	*Suspect*
10/26/2012 15:00	n/a	null	*Suspect*
10/26/2012 16:00	90740.55	null	*Suspect*
10/26/2012 17:00	90741.85	1.30	Collected
10/26/2012 18:00	90743.21	1.36	Collected

Once suspect data have been flagged the data gap filling process can be applied.

3.0 Data Gap Filling Process

The data gap filling process described in this section assumes the processed electricity data will be used for the types of analysis described in Section 1.0 (i.e., annual and monthly consumption, daily profiles) and that the estimated data will be flagged. Any data point that is estimated using a data processing method must be marked as estimated and should not be used to develop future estimates (CPUC 1999b). Data processing changes the measured values and thus influences the analysis. Metered datasets frequently have at least minor errors, so managing the suspect data is a commonly accepted analytical process. The methods provided in this report offer a minimal level of processing focused on changing the values that make multiple building comparisons challenging because buildings with missing or erroneous data would appear to perform better or worse than buildings with a complete dataset. The methods were developed after examining utility industry data processing techniques and adapting them to the likely analysis needs of Army installation energy managers (Aclara Technologies LLC 2008; AEMO 2014; AUC 2014; CPUC 1999b; Elhub, 2014; ERCOT 2010b; IESO 2012a and 2012b; IMEM 2013; PG&E 2009, Rossel 2013).

'Estimation' is the process of substituting suspect and null values with values generated from known data (IESO, 2012a).

Data gaps are the null values created from processing the suspect data (Section 2.0) and for the other missing data within the dataset. The objective of this step in the data processing is to fill as many data gaps as are reasonable and appropriate for data analysis. Figure 3-1 outlines the general flow of gap filling, based on the type of data analysis being conducted.

The phrase 'data gap' refers to a missing timestamp or timestamps where no metered values are present, also referred to as a 'null' value. A metered value of 'zero' is not necessarily an error. It is possible that no energy was used over a time interval.

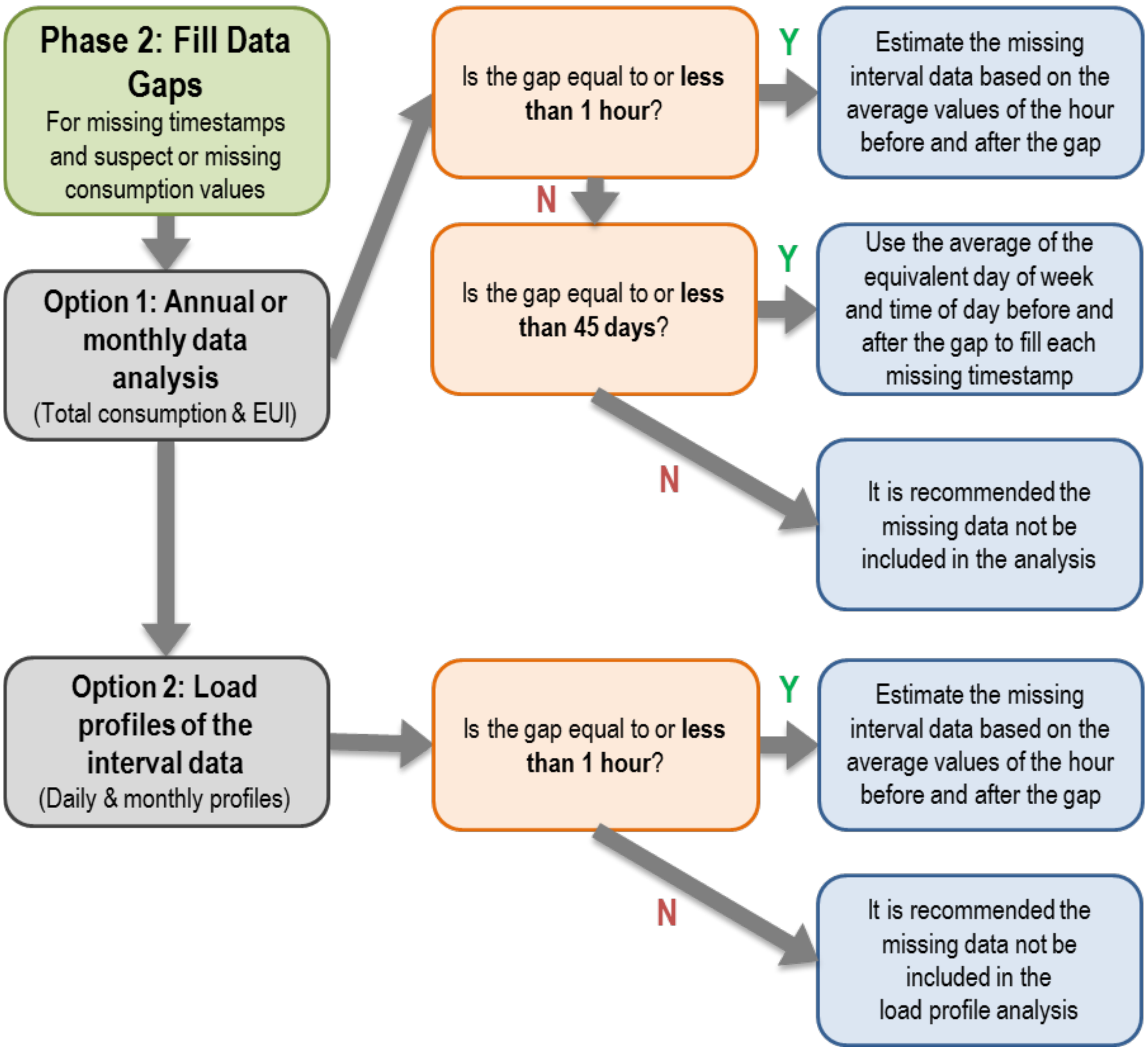


Figure 3-1 Flow diagram outlining the process used to fill data gaps for annual or monthly data analysis and daily or monthly load profiles.

For annual and monthly data analysis (Option 1), the gap filling procedure is dependent on the length of the data gap. For daily or monthly load profiles (Option 2), it is recommended that only data gaps less than or equal to one hour be filled. The gap filling process is defined below for data gaps that are: a) less than 1 hour, b) one hour to one week, or c) greater than one week. It is recommended that data gaps no greater than 45 days be estimated to avoid misrepresentation of energy consumption. The length of 45 days is a judgement call based on the worst case of those 45 days being half of a season. In that case, missing data greater than 45 days would result in more than half of a season being estimated. Ideally the data gaps would be much smaller or non-existent.

3.1 Data Gaps Less than or Equal to One Hour

If the data gap is less than or equal to one hour one of two approaches could be used. The simplest approach is to linearly interpolate between the missing data points. The second approach is to take the average collected data (blue text) for the hour preceding the gap and the hour following the gap. The average value is used to fill gaps and record is marked as estimated (red text). The second approach is shown in Table 3-1.

Table 3-1 Example data gap filling procedure for data gaps less than one hour. Estimated values are depicted by red text.

Timestamp	Consumption (kWh)	Data Source
5/2/2012 12:45	1.923	Collected
5/2/2012 13:00	1.876	Collected
5/2/2012 13:15	1.952	Collected
5/2/2012 13:30	1.906	Collected
5/2/2012 13:45	2.037	Estimated
5/2/2012 14:00	2.037	Estimated
5/2/2012 14:15	2.273	Collected
5/2/2012 14:30	2.171	Collected
5/2/2012 14:45	2.101	Collected
5/2/2012 15:00	2.095	Collected

3.2 Data Gaps Greater than One Hour and Less than or Equal to One Week

If the data gap is greater than one hour and less than one week, take the average of the collected value for the timestamp for that time of day and day of the week during the week before the gap and that day of the week following the gap, that is, seven days before and seven days after. If the values for both the week before and the week after are not available, use the one that is available. Table 3-2 and Table 3-3 provide simplified examples of missing timestamps for approximately eight and a half hours (34 timestamps, not individually shown) where measured data are available before and after, and where measured data are only available before. In Table 3-2 the missing timestamp data gap was filled by averaging the consumption for that time of day and day of the week during the week before and the week after the data gap.

Table 3-2 Data gap filling example for data gaps greater than one hour and less than or equal to one week with before and after data.

Timestamp	Consumption (kWh)	Data Source	Notes
4/10/2012 15:30	1.969	Collected	
4/17/2012 15:30	1.962	Estimated	34 timestamps missing in a row
4/24/2012 15:30	1955	Collected	

In Table 3-3 the missing timestamp data gap was filled using the collected value from the previous timestamp only because there were no collected values available after the missing timestamp. If the previous week's timestamp is not available, the analyst may look as far as three weeks from the timestamp in question. If any of the days involved in this process are holidays then the similar timestamps need to be other holidays or a Sunday.

Table 3-3 Data gap filling example for gaps greater one hour and less than or equal to one week without before and after data.

Timestamp	Consumption (kWh)	Data Source	Notes
4/10/2012 15:30	1.969	Collected	
4/17/2012 15:30	1.969	Estimated	34 timestamps missing in a row
4/24/2012 15:30	null	*Suspect*	

3.3 Data Gaps Greater than One Week and Less than or Equal to 45 Days

If the data gap is greater than one week, electricity use data for the week(s) immediately before and the week(s) immediately following the gap to estimate missing values. If both weeks are available, calculate the average of the collected value for the timestamp from the previous week at the same time of day and same day of the week and the collected value for the same timestamp for the following same time of day and same day of week, that is, if the timestamp is on a Monday, average the values from the last Monday that had collected data and the next closest Monday in the future that had collected data. For gaps greater than one week, the day would be filled with the same set of numbers for each of the time intervals. If timestamps are missing from one of these weeks, preceding or following, use the available measured value from a single week. Table 3-4 provides an example of how two weeks of missing data could be filled using the data available before and after the missing timestamps.

Table 3-4 Data gap filling example for data gaps greater than one week and less than or equal to 45 days.

Timestamp	Consumption (kWh)	Data Source	Day of Week
3/9/2015 0:00	0.0307	Collected	Monday
3/10/2015 0:00	0.032	Collected	Tuesday
3/11/2015 0:00	0.628	Collected	Wednesday
3/12/2015 0:00	0.0333	Collected	Thursday
3/13/2015 0:00	0.596	Collected	Friday
3/14/2015 0:00	0.0333	Collected	Saturday
3/15/2015 0:00	0.58	Collected	Sunday
3/16/2015 0:00	0.9569	Estimated	Monday
3/17/2015 0:00	0.6395	Estimated	Tuesday
3/18/2015 0:00	1.267	Estimated	Wednesday
3/19/2015 0:00	0.572	Estimated	Thursday
3/20/2015 0:00	0.907	Estimated	Friday
3/21/2015 0:00	0.972	Estimated	Saturday
3/22/2015 0:00	1.157	Estimated	Sunday
3/23/2015 0:00	0.9569	Estimated	Monday
3/24/2015 0:00	0.6395	Estimated	Tuesday
3/25/2015 0:00	1.267	Estimated	Wednesday
3/26/2015 0:00	0.572	Estimated	Thursday
3/27/2015 0:00	0.907	Estimated	Friday
3/28/2015 0:00	0.972	Estimated	Saturday
3/29/2015 0:00	1.157	Estimated	Sunday
3/30/2015 0:00	1.883	Collected	Monday
3/31/2015 0:00	1.247	Collected	Tuesday
4/1/2015 0:00	1.905	Collected	Wednesday
4/2/2015 0:00	1.111	Collected	Thursday
4/3/2015 0:00	1.218	Collected	Friday
4/4/2015 0:00	1.911	Collected	Saturday
4/5/2015 0:00	1.733	Collected	Sunday

Data gap filling is best performed by someone who is familiar with the building(s) and with building energy analysis. Familiarity with the buildings allows for the analyst to understand whether the data losses are truly a data loss or something happening in the building. Familiarity with the building energy analysis allows the analyst to judge whether the estimations are appropriate for the building. In lieu of an analyst familiar with the building(s), this simplified processing method can be applied to ensure the buildings have a similar quantity of data when they are being examined.

4.0 Summary

Ideally energy data analysis is performed by someone with adequate time, first-hand knowledge of a building's operations, and expertise in energy data analysis. Army energy managers are time constrained and have large quantities of metered data they are expected to manage and analyze regularly. They are in need of data analysis and management tools that will make it easier to compare building performance to identify which buildings have the potential to reduce energy use.

The data processing method defined in this report was derived with electricity data and was intended for annual or monthly consumption analysis or to create daily and monthly load profiles of interval consumption data. The two-phased process, outlined in Figure 4-1, first identifies and flags suspect data and then addresses the suspect data with a gap filling process.

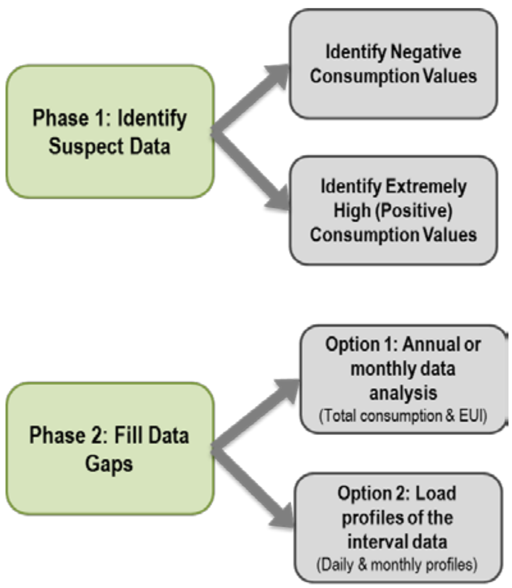


Figure 4-1 Data Processing Method Phases

The data processing method described in this document could be used directly by Army energy managers so that they have a consistent method to apply to their data analysis. As the Army's enterprise system, MDMS could potentially embed this method into its data management system, which would offer energy managers a mechanism to quickly assess the energy use of multiple buildings, and minimize the need for energy managers to manually process the energy data. A standard data processing method incorporated into MDMS with the user option of downloading processed or raw data would be beneficial to users. Viewing analysis reports with either the processed or raw data could also be a helpful option within MDMS. Having the option to use the processed or raw data would allow energy managers to use their first-hand knowledge of building operations, when appropriate, while simplifying the data analysis process in most cases.

5.0 References

Aclara Technologies LLC. 2008. Meter Data Management: The Key to Unlocking the Benefits of Advanced Metering. Aclara Software, Inc., Hazelwood, Missouri.

Australian Energy Market Operator (AEMO). 2014. Metrology Procedure: Part B: Metering Data Validation, Substitution and Estmation Procedure for Metering Types 107. MT_MA1680. Metering Department, AEMO, Australia.

Alberta Utilities Commission (AUC). 2014. AUC Rule 021: Settlement System Code Rules. AUC, Alberta, Canada.

California Public Utilities Commission (CPUC). 1999a. Direct Access Standards for Metering and Meter Data (DASMMD) in California. Decision 98-12-080. CPUC, San Francisco, California.

CPUC. 1999b. Standards for Validating, Editing, and Estimating Monthly and Interval Data (Attachment). CPUC, San Francisco, California.

Elhub. 2014. Standard for Validation, Estimation and Editing (VEE) of AMS metering values. Norwegian electricity market project: Elhub, Oslo, Norway.

Electric Reliability Council of Texas (ERCOT). 2010a. ERCOT Protocols, Section 10: Metering. ERCOT, Austin, Texas.

ERCOT. 2010b. ERCOT Protocols, Section 11: Data Acquisition and Aggregation. ERCOT, Austin, Texas.

Independent Electricity System Operator (IESO). 2012a. Market Manual 5: Settlements, Part 5.2: Metering Data Processing, Issue 25. IESO, Toronto, Ontario, Canada.

IESO. 2012b. Revenue Metering: Marketplace Training. IESO, Toronto, Ontario, Canada.

Interim Mindanao Electricity Market (IMEM). 2013. Metering Standards and Procedures, Issue 1.0. IMEM, Phillipines.

Moore, Sharelynn. 2008. Key Features of Meter Data Management Systems. Itron White Paper. Itron Inc., Liberty Lake, Washington.

Pacific Gas & Electric Company (PG&E). 2009. The Electric ESP Handbook: A Guide to Pacific Gas & Electric Company's Direct Access Procedures for Electric Energy Service Providers. PG&E, San Francisco, California.

Rossel, R. 2013. Meter Data Management and Repository - VEE Standard for the Ontario Smart Metering System, Issue 4.3. IESO_STD_0078. Independent Electricity System Operator (IESO), Ontario, Canada.

Appendix A: Metering System Architecture

Metering system architecture is the conceptual model that defines the structure of the metering system; it describes and represents the structure of the metering system components; how the components are organized; and how they interact. There are many different types of metering equipment, each with a host of options and variations. The purpose of this section is to discuss the overall architecture of a comprehensive metering system from the base sensors to the user interface. Understanding the metering system architecture will help in the overall planning process, especially when planning for the development of enterprise-level systems. Figure A illustrates the generalized structure of an enterprise-level (agency-wide) metering system. Some equipment may combine multiple elements, and some systems can exist without certain components.

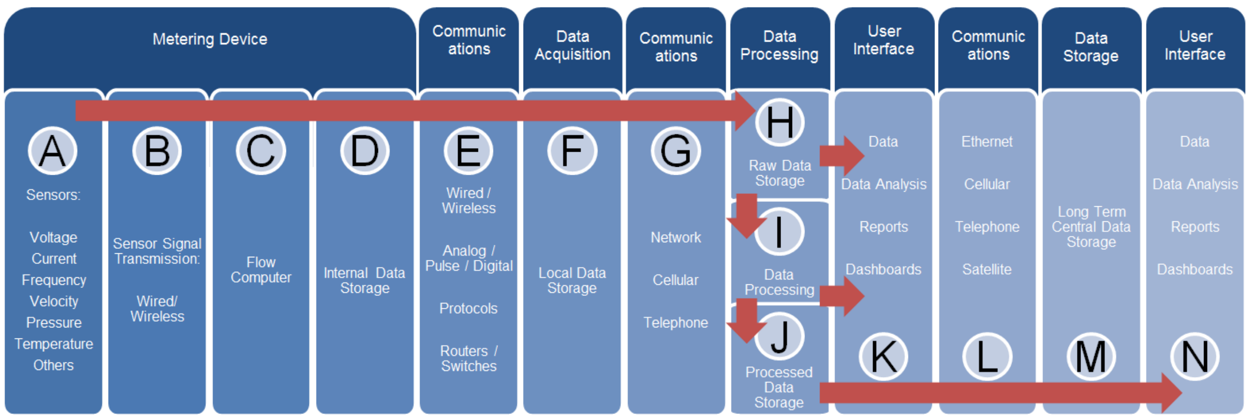


Figure A.1 Metering System Architecture

Metering Device

The metering device consists of four basic elements: sensors (A), signal transmission (B), flow computer (C), and internal data storage (D). Every meter starts with sensor technology. Electric meters require sensors for voltage and current. Other meters rely on sensors for frequency, pressure, velocity, temperature or other parameters. The sensors then must transfer a signal (B) to a processor or computer (C). The sensor signal (B) may be wired or wireless, depending on the device. The processor, or flow computer (C), calculates the values of interest for example, kilowatts per hour, pounds, and gallons per minute. The meter may also have some type of internal memory (D). In the case of an electric smart meter, the internal memory may be capable of storing months of high-frequency interval energy data in addition to a host of additional sensor data. In the case of a basic steam meter, internal memory may be limited to dynamic short-term storage of sensor data being processed by the flow computer. Some meters, such as your basic water meter, may have no internal storage memory.

Communications

In this metering system component, data are communicated (E) from the metering device to a data acquisition system. In practice, this component is actually two elements. As illustrated in Figure A, the meter transmits the data and the data acquisition system (F) receives the data. Originally, this process involved someone manually reading the meter, taking notes, and later writing or typing the noted meter readings as the data moved into the next component. While manually connecting and downloading metered data is still in practice today, the state of the technology is moving more towards remote access and automation. The communications process may be limited to one-way communications, for example, automatic meter reading [AMR] systems, or may be capable of two-way communications, for example, advanced metering infrastructure [AMI] systems. The communications process may be wired or wireless; the signal type may be analog, pulse, or digital; the data may be communicated using any number of protocols; and data may follow any number of pathways to get to the next component. Although automated communications and remote access reduce labor costs and transcribing errors, this component can still be the weak link in the overall metering system. This communications element is one of the most critical in the entire system architecture. Many metering system data losses can be attributed to a failure in this component.

Data Acquisition

For facilities with multiple meters, the data acquisition system (F) scans, reads, and records the meter data. If the meter's flow computer did not apply a timestamp to the data package, the data acquisition system should apply a timestamp as the meter data are recorded. If multiple systems are responsible for applying the timestamp to the metered data, it is important that the time clocks be synchronized with a process for checking time clock accuracy and re-synchronizing periodically. The data acquisition system is frequently an interim component between the meter(s) and the central data storage (H). But for small and local systems, the data acquisition system may include the local central data storage function. For facilities that use a building automation system (BAS) to read and record metered data, the BAS may serve as the data acquisition system. A central energy information system (EIS) computer may serve this purpose. For larger metering systems, the data acquisition system serves as an intermediate step; collecting data from nearby meters and transmitting the data onto the next element. Because it takes time to scan, read, transfer, and record metered data, this component can be a weak link in the process. Trying to transfer too much data in too little time or connecting too many meters to a single data acquisition system can result in skipped scans and lost data. For very large metering systems, multiple data acquisition systems may be warranted.

Communications

In this step, data are transmitted from the data acquisition system (F) on to a central storage component (H). In this communication process (G), the data are likely transmitted in digital form but can use any number of protocols and pathways. The data transmission can still be wired or wireless, using Ethernet, cellular, hard-wired telephone, or even satellite communications. Poor communications in this step does not necessarily mean permanently lost data if the data can be retransmitted from the data acquisition system or internal meter data storage.

Data Storage and Processing

Figure A illustrates the data storage and processing component in three elements. Data from the data acquisition system (F) is communicated (G) initially to raw data storage (H). Raw metered data should always be stored and backed up. Raw data, however, may have data gaps or data quality issues. Raw data may benefit from data pre-processing (I) or cleaning before being used for analysis. The utility industry is accustomed to the fact that some portion of the interval data will be missing, redundant or incorrect. Validating, estimation, and editing (VEE) is the process used by utilities to identify problematic data, fill gaps, remove redundancy, and utilize other checks to address data quality issues.[1] After the data are processed, the modified data file should be stored (J) and backed up.

Site-Level User Interface

The site-level user interface (K) is the component where the data are available for data analysis and data can be visualized using charts or dashboards. The user interface can access a copy of the pre-processed data (N) and may even have direct access to the metering device's internal data storage (D) and the flow computer (C). In this illustration, the user interface is for a local, site-level, facility user. In the case of an enterprise-level metering system, three more components may exist above the site-level system.

External Communications

Data from the facility data storage can be further communicated (L) to an enterprise level. In this communication process, a copy of the raw data (H), the pre-processed data (J), or both may be communicated to the enterprise level. The transmission process may use Ethernet, cellular, hard-wired telephone, or even satellite communications. Poor communications in this step does not necessarily mean permanently lost data because data transfer can always be validated and retransmitted if necessary.

[1] Moore, S. 2008. Key Features of Metered Data Management Systems. Itron White Paper. Itron, Inc. Liberty Lake, WA. June 2008.
https://www.itron.com/na/PublishedContent/Key%20MDM%20Features%20Whitepaper_FINAL.pdf.

Enterprise-Level Data Storage

The enterprise, or agency-level, data storage (M) component is similar to the site-level data storage and processing components. In fact, the data processing and data back-up functions could exist at this level rather than at the site level.

Enterprise-Level User Interface

The enterprise-level user interface (N) component is where the enterprise-level data analysis occurs. Similar to the site level user interface (K), some systems could be designed to provide direct access to lower level data acquisition systems or even meters. However, more frequently they are used for higher-level data analysis, benchmarking, and other comparative analysis.

Appendix B: Lessons Learned from Utility Interval Data Collection Process[1]

When meter interval data has gaps or errors there can be a need for the data to be processed so that it is usable for data analysis. A common question is how utilities manage data losses. While the evolution to a smart grid and advanced metering infrastructure has put even more burden on the data collection process, electric utilities have been collecting time interval energy use data on commercial and industrial customers for decades.

Data Collection Process

Advanced meters are frequently equipped with internal memory for storing metered energy data; whereas meters equipped with advanced metering devices tend to use a data acquisition system to capture and store the metered energy data. Utilities tend to have storage memory as close to the meter as possible, which generally means internal meter memory. In this way, meter data loss generally only occurs when there is a failure in the meter. One of the advantages of having meters with internal memory is that the data remains in the meter until the memory is over written by newer data. Utilities tend to use meters that are capable of storing 30-days of interval data internally.

The Weak Link—Communications

An informal survey of utility representatives revealed that interval data loss is a common occurrence but is manageable. The most frequent cause of data loss is a failure in the communication system between the meter and the central utility data collection computers. Having a robust communication system is the greatest way to minimize data loss. Meters equipped with several days of internal memory allow the utility to reconnect with the meter and repeat the data download process, should any data be dropped in the data collection communication process. A local data acquisition system connected to a meter with an advanced-metering device allows the same opportunity to reconnect and repeat the data download process. For utilities that use automated metering reading (AMR) devices with radio frequency communications between the meter and mobile data collection devices located in utility service

[1] References:

- Rule 021, update 2.4. Alberta Utility Commission. Calgary, Alberta, Canada. January 2014. www.auc.ab.ca
- Market Manual 5: Settlements; Part 5.2: Metering Data Processing, Issue 25. Independent Electricity System Operator (IESO). MDP_PRO_0032. Toronto, Ontario, Canada. September 2012. www.ieso.ca
- Direct Access Standards for Metering and Meter Data (DASMMD); Attachment VEE—Standards for Validating, Editing, and Estimating Monthly and Interval Data; California Interval Data VEE Rules, Revision 2.0. California Public Utility Commission. March 1999.
- Mater Data Management and Repository (MDM/R)—VEE Standard fro the Ontario Smart Metering System, Issue 4.3. Independent Electricity System Operator (IESO). IESO_STD_0078. Toronto, Ontario, Canada. January 2013. www.ieso.ca

vehicles, this means the service vehicles must drive near the meters multiple times the ensure all the data are collected. For utilities that have the ability to remotely connect with meters, such as through telephone lines or network connections, this means the data collection system is likely to connect and download data several times over the billing period to ensure all data that can be collected from the meter is collected and stored. How frequently? One utility reported they connect with customer meters as frequently as every 15 minutes. For a utility with hundreds of thousands to millions of customers, this requires a lot of communication time and computing power. Utilities will have large data centers dedicated to continuously communicating with meters, downloading data, and checking system status.

Validation, Estimation, and Editing

Another lesson learned from utilities is that interval energy data needs to be checked and processed before use. Utilities use a processing method called validation, estimation, and editing (VEE). The need for sound VEE principles evolved from the fact that collecting metered energy data is not a perfect process, utility billing, or settlement practices, need to be sound, and the utility industry is highly regulated. While software that performs the VEE process is available from a number of vendors, much of the VEE process is regulated by the local utility commissions or similar government entity. Utilities seeking to enhance the customer experience, minimize customer complaints, and improve overall meter data management performance, can also customize or enhance the VEE process to fit their specific needs.

Data that has not gone through the validation process is considered raw data. If the data is provided for informational purposes only, such as for a dashboard or other local display, validation is not required and the raw data may be used directly. For most other purposes, particularly if the data is to be used for billing purposes, also known as the settlement process, the data must go through the VEE process—this is generally a requirement of the local utility regulatory agency. Data validation checks are designed to identify things that can go wrong at the meter or local data acquisition system and cause the data collected to not reflect the actual consumption.

In many systems, data goes through much of the validation process as it is being collected by the utility's data acquisition system. The validation process consists of a series of tests, including clock synchronization check, meter identification check, missing interval check, duplicate time interval check, zero check, negative value check, static value check, pulse overflow check, spike check, high/low usage check, and sum check. Customized tests can also be added to the validation step by the utility to be performed on select meter types or customer class categories. As noted, much of the validation process is performed as the data is being downloaded. These validation checks are performed not only to validate the energy consumption data being downloaded, but also to identify if the meter requires physical inspection, maintenance, or recalibration. Failing some of the validation checks may result in a meter being flagged for immediate maintenance or replacement and data may be labeled suspect until the meter issue can be resolved. The sooner problems are identified, the sooner they can be reconciled.

Failure of any of the validation checks will require that the data be verified. Verified data refers to data that failed at least one of the required validation checks but was determined to represent actual usage. Valid data refers to data that passed all the validation checks or has been verified. Once data has passed the validation process, the interval data is stored as valid data.

At this stage, the data may still have gaps, or missing interval data. As noted before, data gaps frequently are caused by drops in the communication system. Reconnecting to the meter and repeating the download process is a first line of defense against data loss. Utilities may even travel to the meter and physically connect and download data to avoid data loss. Utilities try to minimize lost data because it can negatively affect revenue or result in billing disputes. Some data loss, for one reason or another, is always going to occur. To fill data gaps the VEE software uses an estimation process. While different vendor software may use different algorithms to fill data gaps, there are standards that are followed in the estimation process. In the California Interval Data VEE Rules, if the data gap is 2 hours or less, point-to-point linear interpolation is used to estimate the missing data, assuming a power failure is not involved. If the data gap is more than 2 contiguous hours, the average of "like" reference days is used to estimate the missing data. These rules are generally applied to hourly interval data gaps. For 15-minute interval data, the estimation process is generally still performed hourly with the 15-minute interval data assumed to be a flat load throughout the hour. Different regulatory jurisdictions, however, may have slightly different rules. The key point is that estimated data is used to fill the gaps in the metered data. The estimated data is flagged, so data analysis tools can be triggered to perform analytical functions with or without the estimated data.

Beyond the estimation function, the data may also be edited. Editing is a feature used by utilities for specific reasons and under careful guidance and controls. The editing function can be used to replace data that may have failed some of the validation checks or allows data to be overridden as a result of special or unusual events.

Once through the VEE process, the interval energy data are stored by the meter data management system where the processed data can be accessed by the customer information system (for bill generation) and other data analysis.

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by **Battelle** *Since 1965*

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

www.pnnl.gov