AN ASSESSMENT OF THREE SMART METER DISAGGREGATION PRODUCTS IN THE PG&E TERRITORY

Submitted to

EVALUATION MEASUREMENT & VERIFICATION
Pacific Gas and Electric Company
245 Market Street
San Francisco, CA 94111

Submitted by

SBW CONSULTING, INC.
2820 Northup Way, Suite 230
Bellevue, WA 98004

In association with

NEW HORIZON TECHNOLOGIES
ECOMPONENTS TECHNOLOGY, INC.

Final Report

CALMAC Study ID: PGE0386.01

January 25, 2017

sbw

ENERGY + WATER + EFFICIENCY

ACKNOWLEDGEMENTS

The authors want to thank the homeowners who participated in this study. Their patience and assistance during numerous visits by our team to their homes (mapping every electric circuit and outlet, installation of gas and electric measurement devices, maintenance and removal of the measurement system) made this work possible.

We also could not have accomplished this study without the determination of PG&E to get real data, the cooperation of the disaggregators and Silver Spring Networks, the inventiveness of the trades people, and the special coding and processing by our data handlers, Check-it and eComponents Technology. We are grateful for all the work that the team invested on this project.

ERRATA NOTE

To fully address comments received after stakeholder review of the draft report in August 2016, SBW Consulting, Inc. reanalyzed all data and prepared a revised report with additional details. This current report reflects SBW's complete reanalysis of the data and results. As expected, while a few details (and associated graphs) changed, the overall conclusions presented in the original report were unaffected.

Table of Contents

TABLES

FIGURES

EXECUTIVE SUMMARY

Using smart meters, many utilities can now collect, from each of their customers, measurements of electrical use for intervals as short as one minute. Pacific Gas and Electric (PG&E) has deployed such meters to their residential customers in recent years. The data available from these meters has many possible applications, one of which is to provide customers with information that may help them operate their homes in a more energy efficient manner. In particular, PG&E wanted to determine whether the one-minute interval measurements of total electrical energy supplied to a home could be accurately disaggregated into end uses, without relying on any other information about each home, e.g., demographics, appliance and equipment inventory, or building characteristics. If this proved possible, PG&E would be able to routinely provide their customers with information about how much they were spending on end uses such as refrigeration or space cooling. PG&E sponsored the test described in this report to determine whether companies (referred to in this report as vendors) that offer smart meter data analysis services (also known as *disaggregation*) could provide accurate end use estimates if they were given solely smart metering data collected from PG&E residential customers.

Background

The California Public Utilities Commission (CPUC) authorized PG&E to conduct Project #1.18, Demonstrate SmartMeter™- Enabled Data Analytics to Provide Customers with Appliance-Level Energy Use Information, under the EPIC (Electric Program Investment Charge) program. The test documented in this report supports the second objective of Project #1.18:

> *Assess and compare the current analytical capability and accuracy of energy disaggregation software*

In addition, this test provides insights on the accuracy of energy disaggregation techniques that may be useful in other applications of smart meter data, such as the improvement of tools used in estimating savings from various residential energy efficiency measures.

Methodology

This test required the collection and processing of many direct measurements of electric end use in six homes from December 2014 thru April of 2015. These data were used in determining the accuracy of the vendor's disaggregation products. Our data collection efforts proceeded in three stages. First, we collaborated with PG&E in selecting the six test homes. Next, we developed measurement plans for each home. These measurement plans covered both electric and gas end uses, although only the electric uses were relevant to the vendor test. In the final stage, we implemented the measurement plan by installing and operating the required circuit and plug load power monitoring equipment.

All collected data was tested to determine whether it fell within expected ranges and for other indications of error conditions such as repeated identical values. Check-sum comparisons, e.g., total energy feeding a panel compared to the sum of the energy to each of its breakers, were particularly useful in diagnosing problems with the measurement system. We also routinely looked for measurement points that failed to function. Problems were resolved by working with homeowners to adjust the network of plug load monitors or by sending our staff to the home. We also performed tests of the PG&E smart meter data and compared that data to our end use measurements to identify what portion of the data collected should be used in assessing the accuracy of the vendor estimates of end use.

PG&E sought firms who had market-ready web based products that were capable of disaggregating total one-minute electricity usage into separate end uses for residential customers. PG&E selected three vendors for this test and compensated them for their services. Each vendor utilized proprietary algorithms to develop estimates for a set of standardized end-uses at the hourly level for each of their assigned homes. Two vendors were asked to prepare these estimates for 87 homes; the third was asked to estimate end uses for 85 homes. The test sites were included among the data sent to each vendor, but the vendors were blinded as to which homes were the actual test sites.

In our analysis, we defined accuracy as the ratio of vendor estimate to measured end use, expressed as a percentage. This meant that if the vendor estimate for an end use was the same as the measured value the accuracy would be 100%. If they under-estimated by half the amount of measured energy use, then the accuracy would be 50%.

Findings

Our overall finding is shown in Figure ES-1. Each panel in the figure has the same vertical scale so that the size of end uses can be compared. The scale is blank so that we do not reveal the measured end uses. This secures the data so that it can be used for tests of other products in the future. The panels are sorted by the size of the measured end use from left to right and top to bottom. The numbers above each bar are the % of the measured use, which for the vendors (Blue = A, Red = B and Green = C), indicates their accuracy in estimating the end use.

It is important to note that the figure combines measurements for all six test homes and across all five months of the test. In general, we find that the accuracy worsens when we examine it at the month level or for specific sites.

Overall, the green vendor comes closest to estimating the measured end uses. Green nearly matched (111%) the Spa_Pool Pump end use (9% of total use). However, they misestimated by more than a factor of two other end uses that account for significant shares of total use, such as Other (22% of total use), Other Pump (6% of total use) and Space Heater (6% of total use).

The Blue and Red vendors misestimated by more than a factor of two for a larger number of significant end uses. For Blue, those end uses are Lighting (24% of total use), Other, Refrigerator_Freezer (11% of total use), Other Pump, Space Heater and Cooker (4% of total use). For Red, those end uses are Lighting, Other, Spa_Pool Pump, Electric Vehicle (8% of total use), Other Pump and Cooker.

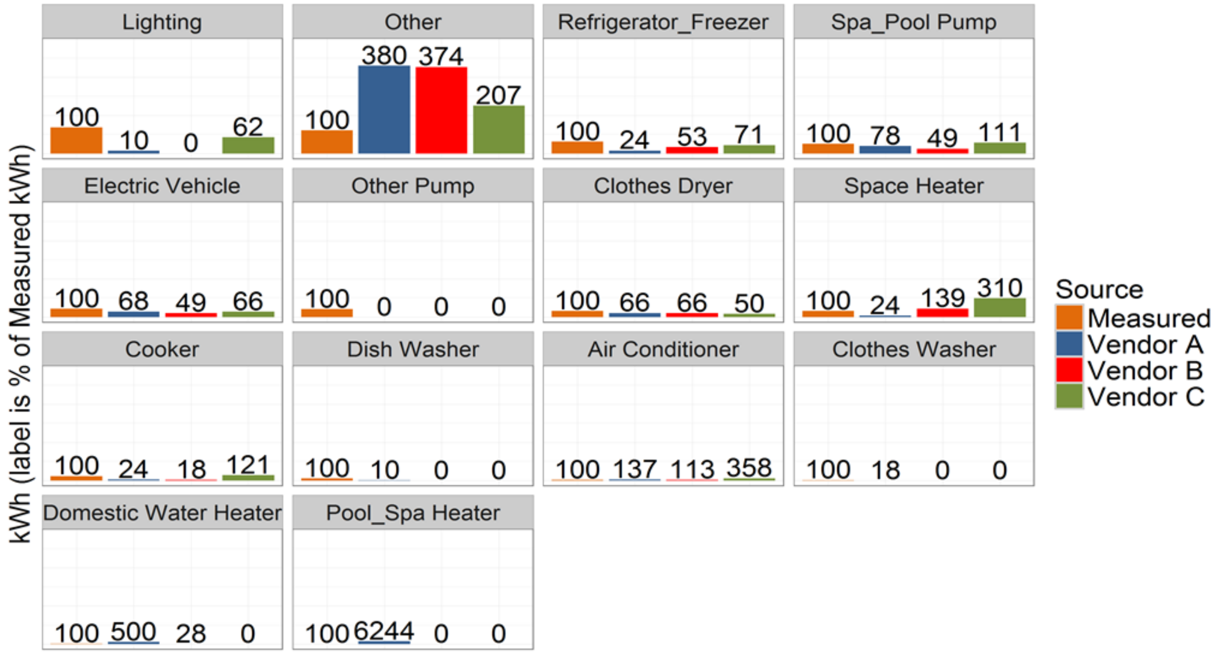

Figure ES-1: Percent of Measured by End Use (All Test Sites - December 2014 thru April 2015) - All Panels Use the Same Scale

We conducted an experiment to see whether sharing some of the measured end use data with the vendors would improve their estimates. After the vendors provided their initial estimates for the entire study period, they were given 1 week of the measured data. With the exception of this 1 week, the measured data were kept confidential and were not (and will not be) released. Due to the significant resources expended to acquire the measurement data, PG&E decided to keep these data confidential in anticipation that it would be utilized in future disaggregation algorithm tests. The vendors were given the measured one-minute channel level data for the 6 test sites for March 1-7 and supplementary data they could use to map the channels to the assigned end use and process the data properly. In addition, the vendors were supplied basic information about the households including the number of people and the appliance information (type and count). They were then asked to resubmit their end use estimates for March. The additional data was not universally helpful for the vendors; however, it did lead to substantial improvements for most of the large end uses. The impact of the additional data was not consistent across end uses.

Conclusion

Based on the data obtained and analyzed under this effort, our results suggest that the tested products available at the onset of this study would not be able to consistently and accurately disaggregate residential home energy usage given the same constraints as the test environment established in this project. While our findings indicate that vendor estimates were close to actual measured values in certain limited instances, these instances represented end uses that constituted a small fraction of overall home energy use. Focusing on only the largest energy end uses, our results indicate that PG&E would have to accept a greater than ±30% error tolerance in disaggregation estimates. The findings from this test indicate that product advancements are needed to improve accuracy.

1. INTRODUCTION

Using smart meters, many utilities can now collect, from each of their customers, measurements of electrical use for intervals as short as one minute. Pacific Gas and Electric (PG&E) has deployed such meters to their residential customers in recent years. The data available from these meters has many possible applications. One of which is to provide customers with information that may help them operate their homes more efficiently. In particular, PG&E wanted to determine whether the one-minute interval measurements of total electrical energy supplied to a home could be accurately disaggregated into end uses, without relying on any other information about each home, e.g., demographics, appliance and equipment inventory, or building characteristics. If this proved possible, PG&E would be able to routinely provide their customers with information about how much they were spending on end uses such as refrigeration or space cooling. PG&E sponsored the test described in this report to determine whether companies (referred to in this report as vendors) that offer smart meter data analysis services (also known as *disaggregation*) could provide accurate end use estimates if they were given solely smart metering data collected from PG&E residential customers.

1.1. Background

The California Public Utilities Commission (CPUC) authorized PG&E to conduct Project #1.18, Demonstrate SmartMeter™- Enabled Data Analytics to Provide Customers with Appliance-Level Energy Use Information, under the EPIC (Electric Program Investment Charge) program. That project had the following objectives:

1. Demonstrate appliance-level itemization of monthly bill charges for residential customers;
2. Assess and compare the analytical capability and accuracy of energy disaggregation software; and
3. Understand customer perception of the end-use cost presentations and the value of the disaggregated data.

The test documented in this report supports the second objective of Project #1.18. In addition, this test provides insights on the accuracy of energy disaggregation techniques that may be useful in other applications of smart meter data, such as the improvement of tools used in estimating savings from various residential energy efficiency measures.

Disaggregation of total energy use into its constituent end use parts is not a new problem. The earliest work in this field dates back to the 1980s. However, the widespread adoption of gas and electric billing meters that record total use at the hourly or sub-hourly level has increased interest in disaggregation and opened a potential large national market for disaggregation vendors and their products. Unfortunately, even though many disaggregation products are offered, they are based on proprietary algorithms, which until recently had not been subjected to independent testing. Some organizations have been conducting independent testing including PG&E, SDG&E and Pecan Street and the conclusions from that work are compared in Section 3.3 of this report.

Much of the prior disaggregation research and product development involved methods that rely on special hardware for high time-resolution measurements (>1,000 measurements per second) whole house electric use. This test for PG&E makes a unique contribution by relying on measurement of whole house electric use that can be provided (when appropriate firmware updates are applied) by the smart meters that PG&E has installed for all end users. The vendors selected for this test were asked to estimate standardized end uses from one-minute interval measurements of whole house electric use. No other study to date has performed an independent test of this nature.

1.2. Project Objectives

The primary objective of this test was to determine the accuracy of end use estimates provided by three vendors of whole house disaggregation products. Specifically, PG&E was interested in the accuracy each vendor could achieve when estimating standardized end uses solely based on one-minute total use measurements made with the standard PG&E residential electric meter.

Accuracy was determined for a series of homes at the monthly level for each standardized end use. The vendor estimates were compared to the direct measurements of these standardized end uses achieved through the installation and operation of circuit- and plug-level power measurement equipment.

A secondary objective was to determine how much the accuracy of the vendor estimates would be improved if they were provided some of the end use direct measurement data and then allowed to re-estimate end use for a portion of the test period.

1.3. Overview of the Test

The test was a blind experiment. PG&E meters provided one-minute total energy use measurements for 166 homes. Direct end use measurements were also obtained from six of these homes (referred to in this paper as test sites). Approximately 60 devices or circuits were monitored in each of the test sites. Three vendors were selected to test their ability to accurately estimate end uses for the test sites. The meters serving each of the homes participating in this study were modified so that they could continue to record normally for billing purposes, but could also provide measurements of electric energy use for each minute.

Homes participating in the study were assigned an arbitrary identifier that masked their identity. The vendors did not know which homes were the test sites, nor were the vendors provided any information about the characteristics of any of the homes. Data was provided on 87 homes to two of the vendors and on 85 homes to the third. There was substantial overlap in the homes assigned to the three vendors, so end uses for some homes were estimated by more than one vendor. Each vendor was asked to provide estimates for a standardized set of electrical end uses at the hourly level for each of their assigned homes over a five-month period from December 2014 to April 2015. The accuracy of the vendor estimates was determined by comparing them to the direct end use measurements for the test sites.

1.4. Report Organization

The balance of this report is organized into three chapters and one Appendix.

- **Chapter 2 – Methodology**. This chapter describes our data collection for the test including how homes were selected, measurement plans established, measurement systems installed and operated throughout the test period. In this chapter, we also describe our data processing procedures, including development of virtual measurements, end use standardization, data aggregation, quality control and the limitations of the direct end use measurements. Finally, this chapter describes how vendor estimates of end use were obtained and compared to the direct measurement of end use.
- **Chapter 3 – Results**. This describes the characteristics of the test homes and the energy use. It also presents the accuracy of the vendor estimates across all test homes and end uses and for selected end uses by month and home. The final section of this chapter compares the design and conclusion of this test to other similar research.
- **Chapter 4 – Conclusions**. In this chapter, we present our conclusions regarding the accuracy of the tested vendor products, lessons learned in the conduct of this test.
- **Chapter 5 – Recommendations.** This chapter provides recommendations for how the data collected could be used in future tests or other applications.
- **Appendix A – Comments and Responses**. This chapter contains comments submitted by reviewers of the draft version of this report and our responses to each of these comments.

2. Methodology

This test required the collection and processing of many direct measurements of electric end use in six homes. These data were used in determining the accuracy of the vendors' disaggregation products. With the exception of one week of measurement data, these data were kept confidential and were not (and will not be) released. Due to the significant resources expended to acquire the measurement data, PG&E decided to keep these data confidential in anticipation that it would be utilized in future disaggregation algorithm tests. This chapter describes how we collected and processed that data and carried out the comparison.

2.1. Data Collection

Our data collection efforts proceeded in three stages. First, we collaborated with PG&E in selecting the six test homes. Next, we developed measurement plans for each home. These measurement plans covered both electric and gas end uses, although only the electric uses were relevant to the vendor test[1]. In the final stage, we implemented the measurement plan by installing and operating the required circuit and plug load power monitoring equipment.

2.1.1. Test Site Selection

PG&E requested volunteers for this test from its employees and contractors. Owners of 37 homes expressed interest in the study. To be selected, the home had to be a primary residence and be served by a PG&E electric meter. Gas end uses also had to be served by a PG&E gas meter. Each homeowner provided information about the home's location and the characteristics of the structure, occupants and energy using equipment. All but nine of these homes were eliminated for the following reasons:

1. **Solar electric**. Some homes were eliminated because they had solar electric generation and were net metered[2]. For such a home the meter readings would not equal to the sum of the end uses within the home and the variation of use over time would in part be due to the variation in output from the solar electric system.
2. **Combined space and water heating systems**. It was not feasible to separately measure heating and cooling energy for these systems.
3. **Remodeling plans**. PG&E wanted to collect a full year of both gas and electric end use data from the test sites. Measurements would be substantially more complicated if the homeowners completed remodeling projects during this period.

[1] Vendors did not offer disaggregation services for gas use. Even if they had, it would not have been possible to conduct a test that was comparable to the electric test. At the time of this test, PG&E's gas meter resolution was one therm (100,000 BTUs or the equivalent of 29 kWh), which meant that one-minute readings of use were not useful.

[2] One home decided to install solar after we installed our measurement system. The homeowner agreed to not turn the solar equipment on until after our end use measurement period.

4. **Antiquated wiring**. A number of homes had outlets that lacked a ground wire. This along with other information such as photographs of electrical panels was used to determine whether the wiring in the home was antiquated and not suitable for this test.

Device/circuit inventories were completed for nine homes. This included a complete inventory of all electric and gas powered devices served by the PG&E meters. Electric devices were included in the inventory if they were permanently connected to a circuit, i.e., hard-wired, or the homeowner reported that the device was always plugged into the same outlet. We ignored devices that were occasionally plugged into an outlet or were often plugged into different outlets. All electric circuits in each home were traced so that each device could be associated with a specific circuit. These inventories were compiled, analyzed and used by PG&E in selecting six test sites. The selected test sites have the following characteristics:

1. **Climate zones**. Two of the test sites were in or near San Francisco and the other four were in or near Stockton, California. Stockton is hotter in the summer season. It has approximately five times as many cooling degree days as San Francisco.
2. **Floor Area**. The test sites were of similar size, with floor area ranging from 2,100 to 3,000 square feet.
3. **Year Built**. Five out of six of the test sites were more than 15 years old. One was built in 2007.
4. **Occupants**. Five out of six test sites had two adult occupants. Of these, only two had occupants that were children (one each). One home had five adult occupants and no children.
5. **Heating and Cooling**. All test sites had gas space heating equipment, either central or wall units. Four of the test sites had central air conditioning.
6. **Cooking**. Two of the test sites only had electric cooking equipment. The other four had a mix of gas stove tops and electric ovens. None of the test sites had a gas oven.
7. **Refrigeration**. Four test sites had one refrigerator. One of them had two and one home had four.
8. **Water Heating**. All test sites heated water with a single gas fired water heater.
9. **Clothes Washing and Drying**. All test sites had a single electric clothes washer, five test sites had a single electric clothes dryer, and one home had a gas clothes dryer.
10. **Pools and Spas**. Two of the test sites had a pump for a swimming pool or spa.
11. **Electric Vehicle Charging**. Two test sites had an electric vehicle charger for the entire study while one home added a charger during the study.
12. **Television**. Three test sites had one television and three had two televisions.

2.1.2. End Use Standardization

During the design of this study, we searched for but were unable to find any national or international standards for classifying energy using equipment by end use. However, clear definitions were needed to guide the vendor estimates of end use and our direct measurements of end uses in the test sites. Each of the vendors had their own list of end uses that they normally estimate and their own definitions of what equipment comprises those end uses. However, PG&E wanted to have one standardized set of end uses so that the accuracy of one vendor could be compared directly to that of another. Standardization would also be necessary if PG&E were to routinely offer disaggregated monthly bills to its residential customers.

We worked with PG&E staff and the vendors to establish a standardized list of end uses and definitions of what equipment comprised each. Drafts of these definitions were provided to the vendors for comment. In addition, we continued to refine the definition as we completed the measurement planning process for selected test homes. As we found unanticipated energy use equipment we modified the definitions and provided the updated definitions to the vendors for comment. The definitions were finalized by the time the measurement system installation was complete at the test homes, which gave the vendors a couple of months to make any necessary final adjustments to their disaggregation routines. Table 1 shows the final descriptions of each end use provided to the vendors.

Table 1: Definition of Electric End Uses Provided to the Vendors

End Use	Description of end use provided to vendors
Space Heater	Plug-in heater, baseboard, wall (with and without fan), fan (in gas unit heater), heat lamp, radiant panel, boiler, furnace, auxiliary fans and pumps
Air Conditioner (AC)	Portable AC, package terminal AC, window AC, evaporative cooler, ceiling fans, central AC, evaporative cooler, attic fan, auxiliary fans and pumps
Domestic Water Heater	Tank (resistance or heat pump), tank less, fans on condensing hot water heater, recirculation pumps
Pool_Spa Heater	Pool / spa heater
Lighting	Fixtures hard-wired, fixtures plug-in, controls, bathroom mirror defogger
Refrigerator_Freezer	Refrigerator, freezer, combined refrigerator / freezer, drink / wine cooler
Cooker	Microwave, stove, oven, convection oven, exhaust fans with integrated lighting and controls, devices that are for heating food, e.g. toaster, toaster oven
Clothes Dryer	Clothes dryer
Clothes Washer	Clothes washer
Dish Washer	Dish washer
Electric Vehicle	Electric vehicle
Spa_Pool Pump	Spa / pool pump
Other Pump	Other pump

End Use	Description of end use provided to vendors
Other	Built-in and mobile vacuums, cooking and food preparation appliances not listed under Cooker, remote controls, telephones, chargers for consumer electronics, standalone exhaust fans, microwaves not used for cooking, garage door opener including integrated light, mirror defogger, humidifier, clock radio, surveillance cameras, tuner, Blue-ray / DVD / VCR, receiver, amplifier, powered speakers, radios and stereos, television, set-top box / DVR, game console, computer / accessory, other devices not elsewhere listed

As is clear from the table, some end uses contain a number of specific devices. Though not formally documented, it was mentioned by some of the vendors that they were not able to identify all the devices listed for each end use. That inability would be expected to reduce the accuracy of their end use estimates.

PG&E anticipated other applications of the data collected from this test site. For example, both the gas and electric measurements might be used in improving tools used to model homes and develop or evaluate energy efficiency upgrades. The data might also be used to understand the load characteristics of specific equipment that might be the subject of efficiency incentives. To meet these other possible applications, we divided some of the standardized end uses into more detailed end uses:

- Space Heater was divided into Room Space Heater and Central Space heater
- Air conditioner was divided into Room Air Conditioner and Central Air Conditioner
- Other was divided into Audio/Visual System, Television, Set-Top Box/DVR, Game Console, Computer/Accessory and Unclassified.

2.1.3. Measurement Planning

We analyzed the circuit and device inventory, along with the physical layout of electric panels and devices in each home to determine the least cost and most reliable plan for installing power measurement equipment and the associated equipment for wireless communication within the house and with our remote data collection systems. The plan included intentional redundancy. For example, we measured (where feasible) the main feed to each electric breaker panel and we measured each of the circuits controlled by those breakers. This allows for a comparison of the panel total use to the sum of the use on each of the breakers. We were able to meter the main feed for four of the six homes. At one of the remaining two homes we had to sum the energy from one pair of subpanel total meters and 15 circuit level meters to calculate the whole house value. At the last house, we had several subpanel totals and several circuit meters and a plug load meter for a subpanel where we could not get a measurement.

Conventional true power measurement devices were used with split-core current transformers (CTs) around one leg of single phase circuits or each leg of two phase circuits (such as those serving clothes dryers). There was one case where one leg of a two-phase device malfunctioned, and so we doubled the single-phase value to get the total. Corresponding potential transformers (PTs) were also connected to obtain the voltage for each circuit.

Figure 1 shows a portion of the measurement plan for one of the homes. This illustrates how we connected each device in a home to a panel and circuit and gathered the information needed to determine whether it could be monitored at the circuit level, by a plug load monitor or estimated by subtracting plug load measurements from circuit-level measurements.

Fuel	Device	Breaker	Hard wired?	Outlet #	Final Sensor Type	Disagg End Use
Electric	Wii	6T	NO	50	Plug	Other
Electric	Wi-fi router	6T	NO	48	Plug	Other
Electric	Washer	1T	YES	0	Plug	Clothes washer
Electric	projection TV	5B	NO	46	Plug	Other
Electric	Treadmill	6T	NO	48	Plug	Other
Electric	Tool battery charger	1T	NO	36	Subtraction	Other
Electric	Toaster oven	23	NO	25	Plug	Cooker
Electric	Tankless gas water heater	7	NO	1	FromCT	Domestic Water Heater
Electric	Sump pump	1B	NO	12	FromCT	Other Pump
Electric	Sump pump	10	NO	16	Plug	Other Pump
Electric	Sump pump	9B	NO	63	Plug	Other Pump
Electric	Sub-woofer	6T	NO	47	Plug	Other
Electric	Sub-woofer	6T	NO	50	Plug	Other
Electric	Steam bath generator	9T	YES	0	FromCT	Other
Electric	Sprinkler controller	7T	NO	13	FromCT	Other
Electric	Shredder	12	NO	37	Plug	Other
Electric	Seat massager	3T	NO	44	Plug	Other
Electric	Routers	6T	NO	62	Plug	Other
Electric	Router	11	NO	70	Plug	Other
Electric	Refrigerator	21	NO	29	FromCT	Refrigerator/Freezer
Electric	Rangehood fan/lt	13	YES	0	FromCT	Cooker
Electric	Projector	6T	NO	56	Plug	Other
Electric	Printer/fax/scanner	5B	NO	47	Subtraction	Other
Electric	Power Strip: S3	12	NO	42	Plug	Other

Figure 1: Example Lines from a Home Measurement Plan

2.1.4. Installation

Monitoring of the test homes was accomplished through collaboration between the PG&E metering department and our team. PG&E upgraded firmware for each home's smart meter so that they could obtain whole house energy use at one-minute intervals. These data were obtained throughout the test period and were stored in a database accessible to our team and to the vendors. The data were used by the vendors as input to their disaggregation products. The data were used by our team in confirming that our end use measurements added up to the total for each house.

Our team was responsible for all measurements on the customer-side of the smart meter. Figure 2 shows how we typically installed power measurement components for an electrical breaker panel. In this case, a separate electrical enclosure was mounted below the breaker panel. Sensor leads pass between it and the breaker panel, where the CTs and PTs have been installed. The new enclosure below contains the necessary multi-channel power recording devices, along with the wireless transceiver that communicate one-minute recordings of electrical use to a central controller.

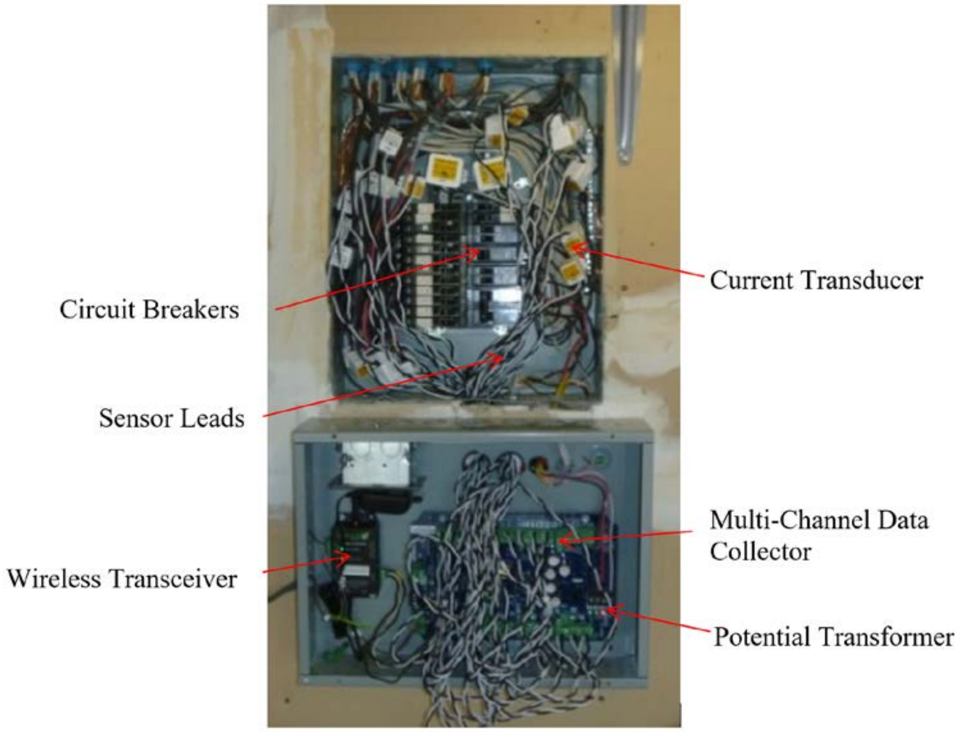

Figure 2: Typical Installation of Power Measurement for an Electrical Breaker Panel

The circuit and device inventory allowed us to identify the end use of each electric circuit. In many cases a circuit served more than one end use. A plan was developed for disaggregating each of these circuits. In some cases, we were able to install wireless plug load power measurement on devices or power strips to disaggregate the energy use into the target end uses as shown in Figure 3. In cases involving circuits that serve hard-wired end uses, we used the plug load power measurement to measure all of the non-hard-wired devices on the circuit. The hard-wired devices were the difference between power measured at the breaker and the sum of non-hard-wired plug load measurements. This “virtual” channel technique was also used to derive energy use for some non-hard-wired circuits.

Figure 3: Wireless Plug Load Power Measurement

All measurement points within each home were wirelessly connected to a multi-protocol controller (Figure 4) that communicated with our primary remote storage via a cellular connection. Plug load measurements reported via z-wave protocols, while circuit measurements were reported over a Zigbee network. Only one controller was installed in each home, so it was critical that it be placed in a location where it would communicate with all of the z-wave and Zigbee sensors throughout the home. The controller device polled each power measurement once a minute and recorded the energy used since the last polling. These recordings were pushed to our remote database once each minute. The gateway had the ability to store many days of data to protect against data loss from interruptions in cellular communication.

Figure 4: Wireless Multiprotocol Hub

It was not possible to completely disaggregate some circuits. Any circuit that served more than two hard-wired end uses was classified based on the end use which had the largest rated power draw. These could have been separately measured but it would have required re-wiring a portion of the home, which was not feasible for this test. In some cases, more than one end use was present in a single device, e.g., an outdoor fountain or a bathroom exhaust fan that had integrated lighting. These devices would have to be disassembled in order to separately measure the end uses, which was also beyond the scope of this test.

Installation of the measurement systems for the six homes was accomplished over a three-week period. Our crews, consisting of two technicians and an electrician booked approximately 2,800 labor hours in the process of installing 326 monitoring points, the associated data acquisition equipment, and inspection by local electrical inspectors to ensure that the work met local electrical codes. Table 2 shows the number and type of physical monitoring points that support the end use measurements across all six homes. These include monitoring panel mains and other points used for quality control tests.

Table 2: Physical Monitoring Points

End Use	Plug Load	Circuit	Total
Air Conditioner	2	15	17
Clothes Dryer	0	11	11
Clothes Washer	3	4	7
Cooker	8	22	30
Dish Washer	3	4	7
Domestic Water Heater	1	1	2
Electric Vehicle	2	4	6
Lighting	6	60	66
Other	124	24	148
Other Pump	1	2	3
Pool_Spa Heater	0	1	1
Refrigerator_Freezer	6	7	13
Spa_Pool Pump	0	4	4
Space Heater	3	8	11
Grand Total	**159**	**167**	**326**

Table 3 shows the points derived by subtraction where physical monitoring was not practical. A common application for this technique were circuits that powered many lighting fixtures but also powered non-lighting devices. We used plug load monitors for the non-lighting devices and substracted that use from the total use of the circuit, which was monitoried with conventional CTs/PTs at the breaker panel.

Table 3: Use of Subtraction in Measuring End Uses

End Use	Plug Load	Circuit	Subtraction	Total
Air Conditioner	2	15	1	18

End Use	Plug Load	Circuit	Subtraction	Total
Clothes Dryer	0	11	0	11
Clothes Washer	3	4	0	7
Cooker	8	22	5	35
Dish Washer	3	4	0	7
Domestic Water Heater	1	1	0	2
Electric Vehicle	2	4	2	8
Lighting	6	60	114	180
Other	124	24	32	180
Other Pump	1	2	9	12
Pool_Spa Heater	0	1	0	1
Refrigerator_Freezer	6	7	0	13
Spa_Pool Pump	0	4	1	5
Space Heater	3	8	1	12
Grand Total	**159**	**167**	**165**	**491**

2.1.5. Startup and Maintenance

During the early period of data collection (October-November 2014) test sites were revisited by our team to replace some malfunctioning equipment, and resolve various other data quality problems. We continued to stay in touch with each of the homeowners at least once a month following these site visits. The main purpose was to detect any changes in the home that would affect our measurements, such as moving a device from one outlet to another or plugging in a new appliance. No changes occurred during the year of measurement that required modification to the power measurements in breaker panels. However, we did work with the homeowners to adjust for changes in plug loads, by installing additional plug load power measurement for new devices, and adjusting the end use energy equations if plug load devices were moved.

2.1.6. Limitation of Measurement System

The process of implementing end use measurement systems at each of the six homes revealed certain limitations, namely:

- **End use contamination**. Some measured devices, such as bathroom ventilation units, which contained both a fan and light fixture, were of mixed end use. We had to assign these devices to an end use based on an estimate of which end use accounted for the largest share of the device's use over the test period. Other examples were pumps that served both space and water heating or fans that served both space cooling and space heating.

- **Very small or temporary loads.** We did not attempt to measure the energy use of power strips. In addition, and probably more importantly, we only measured devices that were always plugged into a wall outlet or had a hard-wired connection to a circuit. Devices like vacuum cleaners stored in a closet were not directly measured. However, it is possible their energy use was captured at the circuit level and included in our measured end uses.
- **Measurement resolution.** The resolution of our electrical measurements was .001 kWh (1 watt-hour). Some devices used less than this amount in a minute. For such devices, the use would be reported after sufficient time has passed to accumulate at least .001 kWh, making the intervening one-minute intervals appear to have no energy use.

2.2. Data Processing

This section describes how the data flowed from the measurement points in each home (PG&E smart meter, circuit, or plug load power monitors) to the final edited form used in assessing the accuracy of vendor estimates of end use.

2.2.1. Data Flow

Figure 5 shows the path that the data followed from the measurement points in each home to the data repository where we performed analysis and quality testing. There are two separate paths followed. The first is the path followed from power sensors installed by our team, which included smart switches for plug loads and current transformers for circuits. The smart switches have built-in true power meters and communicate their power measurements to the multi-protocol controller that we installed in each home, via a Z-wave wireless network.

The current transformers (along with potential transformers) were located in breaker panels and connected via low voltage leads to a series of Dent PowerScout power meters. Each PowerScout communicated using the Modbus protocol with a Multi-IO converter which subsequently communicated via the wireless Zigbee network to the multii-protocol controller. This Zigbee network allowed multiple PowerScout meters to communicate in homes that had breaker panels located in different rooms and avoided running wires from the breaker panels to the single controller that collected data from both plug loads and breaker panels.

The controller communicated over the public cellular network with Check-It Solutions, which supplied the controller and operated a cloud-based data acquisition service. Power measurements were acquired from all monitoring points once each minute and stored in the Check-It Solutions database. Each day the collected data was transferred via FTP to a SQL data repository operated by eComponents Technology (eCT) where a continuous time series of the un-edited one minute power measurements along with various monitoring status information was assembled.

The other data path shown in the figure is for the one-minute power measurements from the PG&E smart meter. These meters are connected to the Silver Spring Networks (SSN) and were accessed via an applications programming interface (API). As shown, both our team and the disaggregation vendors had direct access to the data via this API. The vendors also used this API to report back their estimates of hourly end uses for each home. Our team used the API to access these vendor estimates and to acquire the one-minute PG&E smart meter power measurements.

Our team brought all of the data together for analysis and quality testing. Data was regularly copied from the eCT SQL repository and from the SSN network.

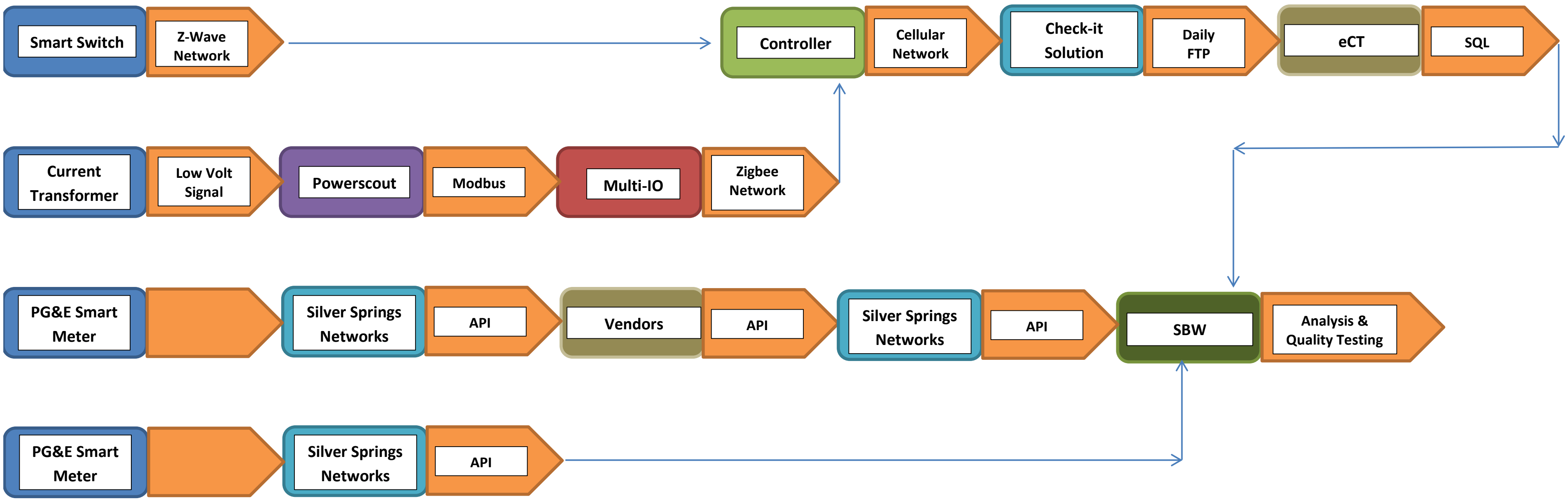

Figure 5: Data Flow

2.2.2. Data Quality Review and Editing

All collected data was tested to determine whether it fell within expected ranges and for other indications of error conditions such as repeated identical values. Check-sum comparisons, e.g., total energy feeding a panel compared to the sum of the energy to each of its breakers, were particularly useful in diagnosing problems with the measurement system. We also routinely looked for measurement points that failed to function. Problems were resolved by working with homeowners to adjust the network of plug load monitors or by sending our staff to the home. We also performed tests of the PG&E smart meter data and compared that data to our end use measurements to identify what portion of the data collected should be used in assessing the accuracy of the vendor estimates of end use.

Figure 6 illustrates our process of data quality testing and editing that led to a final dataset used in the assessment of accuracy. Three paths are shown, one for the PG&E smart meter data, another for end use monitoring data collected by our team, and a final path for vendor end use estimates. These paths interact as decisions made about the quality of measurements from one source affect which measurements from other source are retained in the final data set that is used for assessing vendor accuracy.

We acquired the PG&E smart meter data from the SSN API repeatedly during the test period (December 2014 through April 2015). We provided feedback to PG&E concerning the quality of this data, which resulted in some data being reprocessed and posted again via the API. Even with this feedback and reprocessing, there were still some hours which did not have a complete set of one-minute measurements for the PG&E smart meters. These hours were dropped from the accuracy assessment. Table 4 shows how these hours were distributed over the months for each site. In total, 11% of all hours were dropped.

Table 4: Hours with Incomplete PG&E Smart Meter Measurements

Site	Dec	Jan	Feb	Mar	Apr	Total	% of Entire Period
Site 1	12	1	195	5	1	214	7%
Site 2	10	0	193	89	0	292	9%
Site 3	26	2	197	282	129	636	19%
Site 4	22	5	206	6	1	240	7%
Site 5	8	0	195	240	0	443	14%
Site 6	10	1	193	141	0	345	11%
Total	**88**	**9**	**1,179**	**763**	**131**	**2,170**	**11%**

The second path in Figure 6 shows how we handled the data obtained from our end use monitoring. The first editing of the data occurred at the Check-It Solutions database. When a device failed to report every minute, the value that was reported was averaged over the timeframe that the device was not reporting. A status flag was set to indicate that the data had been averaged. The second editing of this data occurred when we extracted measurements from the eCT SQL database. Soon after we started collecting monitoring data, during the initial setup work, we noticed that some of the smart switches (plug load power monitors) would record impossibly high values for short periods of time. This problem was investigated with the manufacturer and with Check-It Solutions, but no explanation was ever developed. We decided to set these values to zero as they generally occurred in period of zero power usage for the plug load devices they were monitoring. All plug load values over 1,000,000 Watt-Seconds (equal to 16.7 kW, which is far beyond the capacity of any circuit in a home) for a single minute were set to zero.

The next steps of data quality testing shown in the figure were actually performed iteratively. We conducted many rounds of testing in which various levels of checksums were computed at the circuit, breaker panel and whole house level to identify configuration errors (usually a scaling factor). Other edits affected how an end use was defined which could involve addition and subtraction of a number of separate monitoring points, or multiplication in the case where we monitored on one phase of a balanced two phase circuit. Some monitoring points were of mixed end use, and we had to make a judgement about which end use accounted for the largest share of use, which determined its end use assignment. For example, pumps and fans that served both heating and cooling loads were assigned to the Heater end use because the test period was dominated by winter months (December through April). In addition, we used data from each home's measurement plan to establish high-value limits for specific plug load devices. These limits allowed us to detect and set to zero values that were less than 1,000,000 watt-seconds but still above the amount of energy that could be consumed by a device in specific period of time.

Some end uses were defined by equations that involved subtraction. For example, a portion of the lighting end use might be derived from a circuit by subtracting non-lighting plug load measurements from that circuit's total use. At times, these subtractions would result in negative values. These negative values were set to zero as part of determining the home's lighting end use. Subtraction was performed at the hourly level. This editing was relatively infrequent, affecting less than 0.5% of the hourly records.

In addition, we observed intermittent outages that involved the wireless networks of smart switches (plug loads) and the controller in each home. These resulted in the energy use being accumulated, for some devices and periods of time, over longer than a one-minute interval and then averaged over a series of one-minute intervals. This was an infrequent problem and had little effect once the data were aggregated to the one-hour level (vendors also estimated end use at the hourly level) .

In the final stage of our quality testing, we aggregated the end use measurements to the daily level and summed all end uses to estimate whole house electrical use by day for each site. Similarly, the PG&E smart meter data was aggregated to this same level. We compared these two measurements of whole house use and excluded from the data that was used for the accuracy assessment any days in which the difference was greater than +/-5%, which we considered to be the likely limit of measurement error for our system of measurement. Some of the days that fell outside of the thresholds occurred because of our work to correct installation errors at each site during the early months of the test. Others may be due to the use of temporarily connected loads that we intentionally ignored in our measurement planning process.

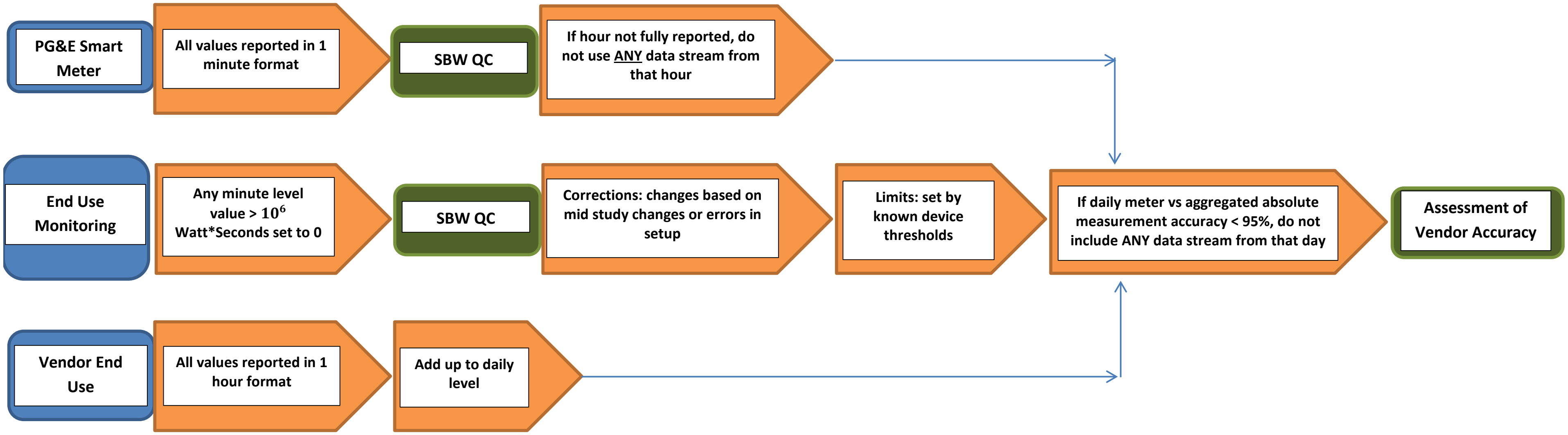

Figure 6: Quality Review and and Editing

Table 5 shows the number of days retained for the accuracy assessment by home and month. Some homes did better than others. The worst was Site 1 for which we retained only 50% of the days. We retained more than 75% of days for all other sites. The best sites were Site 2 and Site 5 for which we retained 98% or more of days. Across all sites, we retained 81% of all days of data for use in the accuracy comparison.

The electrical use for any day retained could be +/-5% of the PG&E smart meter measurement. So, for some days the sum of our end use measurements was high and for some days it was low, but within this 5% bound. As shown in Table 5, the average difference varied across homes, with a low of 95% and high of 102%. It is important to note that across all homes and days, the total end use measured energy is 100% of the PG&E smart meter energy during the test period.

Table 5: Days Retained by Site for Accuracy Test

Site/ Month	Days Kept	Total Days	% of kWh Kept	% of Days Kept	Measured % of Meter kWh
Site 1	**65**	**130**	**57%**	**50%**	**95%**
Dec	1	31	10%	3%	84%
Jan	8	31	34%	26%	90%
Feb	18	21	91%	86%	101%
Mar	25	31	85%	81%	101%
Apr	13	16	78%	81%	100%
Site 2	**126**	**127**	**99%**	**99%**	**100%**
Dec	31	31	100%	100%	100%
Jan	31	31	100%	100%	99%
Feb	20	21	95%	95%	100%
Mar	28	28	100%	100%	101%
Apr	16	16	100%	100%	99%
Site 3	**90**	**113**	**81%**	**80%**	**99%**
Dec	12	31	47%	39%	98%
Jan	27	31	87%	87%	98%
Feb	20	20	100%	100%	100%
Mar	20	20	100%	100%	102%
Apr	11	11	100%	100%	100%
Site 4	**108**	**130**	**88%**	**83%**	**99%**
Dec	23	31	81%	74%	104%
Jan	30	31	98%	97%	100%
Feb	18	21	93%	86%	97%
Mar	25	31	84%	81%	96%

Site/ Month	Days Kept	Total Days	% of kWh Kept	% of Days Kept	Measured % of Meter kWh
Apr	12	16	85%	75%	97%
Site 5	**117**	**120**	**98%**	**98%**	**98%**
Dec	30	31	97%	97%	98%
Jan	30	31	97%	97%	98%
Feb	19	20	95%	95%	98%
Mar	22	22	100%	100%	98%
Apr	16	16	100%	100%	97%
Site 6	**95**	**125**	**76%**	**76%**	**102%**
Dec	13	31	33%	42%	112%
Jan	29	31	94%	94%	99%
Feb	20	21	96%	95%	97%
Mar	25	26	97%	96%	98%
Apr	8	16	52%	50%	105%
Grand Total	**601**	**745**	**83%**	**81%**	**100%**

The third path in Figure 6 is for the end use estimates developed by the vendors. We acquired these shortly after the end of each month for the preceding month. We excluded hours in which the PG&E smart meter had missing minutes. We also excluded days where the sum of our end use measurement did not match the PG&E smart meter within +/- 5%.

The final data set included all three types of data: PG&E smart meter, our end use measurements, and vendor end use estimates. It was aggregated to the daily end use level by home and only contained data that we felt was of sufficient quality to support the assessment of vendor accuracy.

2.3. Vendor End Use Estimation

PG&E sought firms who had market-ready web based products that were capable of disaggregating total one-minute electricity usage into separate end uses for residential customers. PG&E selected three vendors for this test and compensated them for their services. As shown previously in Figure 5, all of the PG&E smart meter data was made available to the vendors. Each vendor utilized proprietary algorithms to develop estimates for each of the standardized end-uses (described previously in Section 2.1.2) at the hourly level for each of their assigned homes over a five-month period from December 2014 to April 2015. Two vendors were asked to prepare these estimates for 87 homes; the third was asked to estimate end uses for 85 homes. The homes provided to each vendor overlapped, so that more than one vendor estimated end uses for some of the homes. The test sites were included among the homes sent to each vendor, but the vendors were not told which homes had end use measurement systems. The vendors delivered their hourly estimates for each home's end uses a few days after the end of each month during the test period.

The vendors had direct access to the PG&E smart meter data for all the homes they were assigned. Our team did not edit or limit access to that data in any way, so they had access to all the one-minute data recorded by those meters. They were able to use all of that data in estimating end uses, even though only a portion of the days in the test period were used for comparing our measured end uses to the vendor end use estimates, as described in Section 2.2.2.

Initially, the data reported by the vendors did not add up to the PG&E smart meter measurement of use for each home. We informed the vendors that the sum of their end uses needed to match the use measured by the PG&E smart meter at the hourly level for each home. We believe that they met this request by placing all of the difference in the Other end use.

An experiment was conducted to see whether sharing some of the measured end uses with the vendors would improve their estimates of end use. After the vendors provided their initial estimates for the entire study period, they were given a portion of the measured data. Each vendor was given the measured one-minute channel-level data for the six test sites for the week of March 1-7, 2015, as well as supplementary data they could use to map the channels to the assigned end use and process the data properly. In addition, the vendors were supplied basic information about the households, including the number of people and the appliance information (type and count). They were then asked to resubmit their end use estimates for the month of March 2015.

2.4. Accuracy Testing

The algorithms used by the vendors in estimating hourly end use are proprietary. Neither PG&E or our team have any information about the specification of these algorithms. The vendors were asked to provide results from prior testing of the accuracy of their respective algorithms, especially tests conducted by independent third parties. None of the vendors provided our team with any prior test results.

We explored a number of methods for characterizing the accuracy of the vendor end use estimates. An accuracy metric had previously been developed by the Electric Power Research Institute (EPRI), which has the following functional form.

$$EPRI\_Accuracy = 1 - \frac{1}{\bar{x}}\sqrt{\frac{1}{n_{obs}}\sum_{i=1}^{n_{obs}}(x_i - \hat{x}_i)^2}$$

where:

x_i = measured values

$\bar{x}$ = the mean of x_i

$\hat{x}_i$ = vendor predictions

n_{obs} = number of observations

This metric was used in a similar study conducted by San Diego Gas and Electric (SDG&E), which is further discussed in Section 3.4.1. The authors of the latter study chose to remove measure values near zero when they computed this metric. They did this because they felt that false positives and false negatives had a significant impact on the calculated accuracy value. It is not clear to us why this would be justified. Consequently, we computed this metric without removing any values as we examined each month of data and the vendor estimate of end use. A perfect EPRI accuracy value is unity (1). Most of our results were negative values (or very small positive values) which represented a very poor score that did not seem helpful for comparison purposes. It was also difficult to interpret the value of the metric, as there are no established standards for what value of this metric would constitute a good match between vendor estimates and measured end uses.

Due to the limitations of the EPRI formulation, we chose a metric that directly compares the vendor end use estimate to the measured value. That is, we assessed accuracy using the ratio of vendor estimate to measured end use, expressed as a percentage. If the vendor estimated the same use for an end use as was measured for any hour, day, month or the whole test period, the accuracy would be 100%. If they estimated half the amount of energy use, then the accuracy would be 50%.

In the context of a disaggregated PG&E monthly bill sent to a home, we are still confronted with the question of what accuracy percentage is sufficient. This metric does provide advantages in that it is easily interpreted and has a simple physical explanation. However, the energy used in any period of time differs greatly between one end use and another. It is important that when comparing the accuracy across end uses, that the magnitude of the end use energy consumption is considered. As the results presented in the next section make clear, we have accomplished this by always plotting the estimated and measured end use on the same scale in figures that compare the accuracy for more than one end use.

3. RESULTS

This chapter presents the results of the test and discusses how these results compare to other similar recent studies. To put these results in context, we start by describing the energy use of the test homes.

3.1. Energy End Use in the Test Homes

The end use shares[3] of total use measured in the six test sites is shown in Figure 7. Four end uses (Light, Electric vehicle, Refrigerator_Freezer, and Spa_Pool Pump) account for 74% of total use. Combined, Room and Central space heating account for 5%, about the same as Clothes Dryer. Air Conditioners (Room and Central) account for only 1%. We expect that this would have been much larger if the test period had extended into the summer months. The largest share of use is for Other. This end use comprises many specific devices, some of which account for a larger share of total use than some of the other end uses. For example, Audio_Visual system, Computers_Accessory, and Set-top box_DVR, all use more than Clothes Washers. After Other, the largest end uses are Lighting, Refrigerator_Freezer and Electric Vehicle. Half the test sites had electric vehicle chargers, and the figure underscores how important this end use can be if it is present in a home. Use for the Spa_Pool pump was nearly as large, even though only two test sites had this end use, so it is also important when present.

[3] These shares reflect all of our measurements for the sites after data editing described in Section 2.2.2. This includes days of measurements that did not meet our criteria for comparison to vendor estimates, i.e., where the sum of our end use measurements fell within 5% of the PG&E smart meter values.

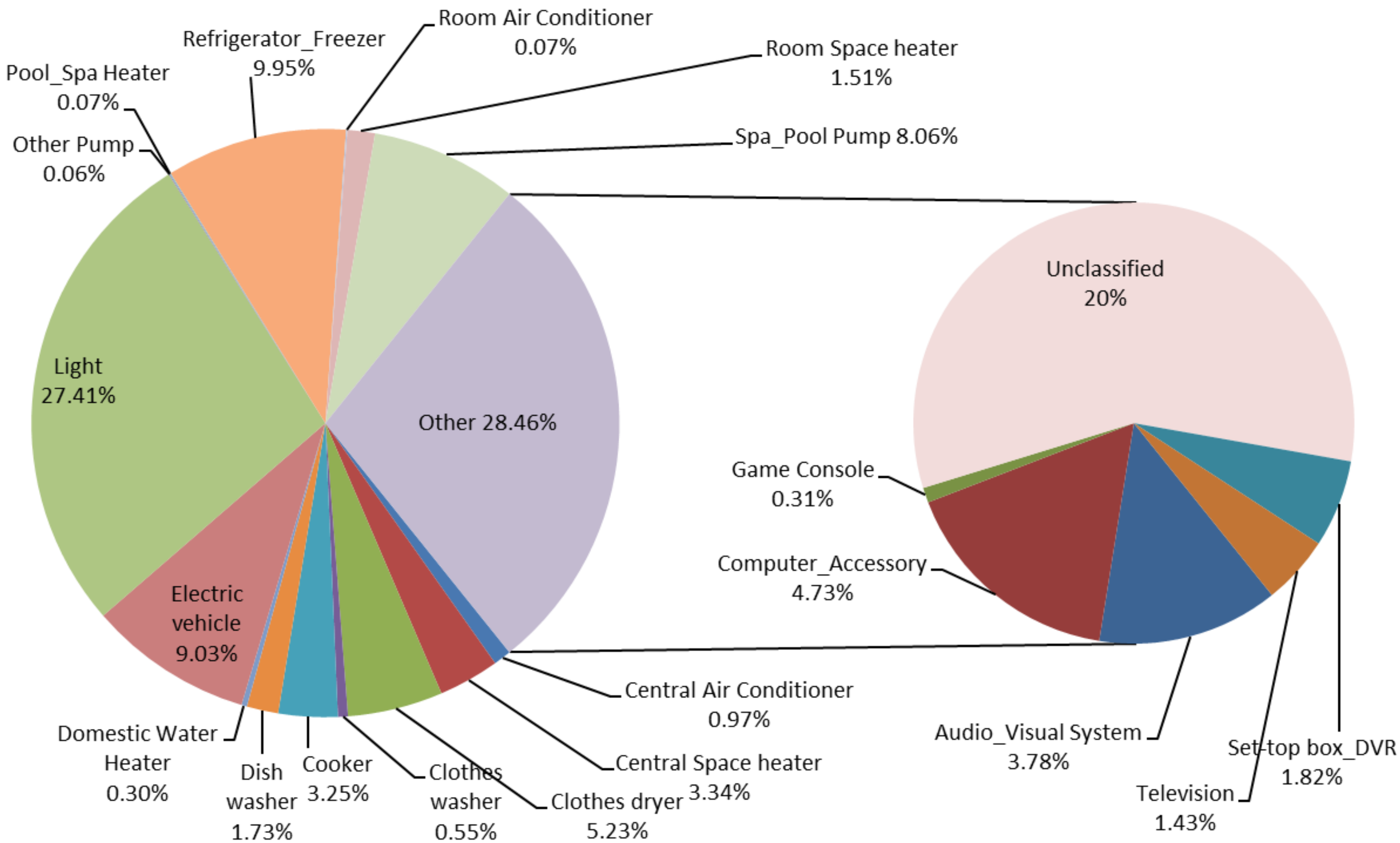

Figure 7: End Use Shares (All Sites December 2014 thru April 2015)

The Other end use is clearly important at 28% of total use across all six test sites. Unfortunately, as shown in Figure 7, it comprises many types of devices. The largest portion of the Other end use is associated with devices that did not fall into any of our pre-defined device categories (shown as Unclassified within Other in the figure), accounting for 20% of total use.

Figure 8 provides more detail of what is happening within the Unclassified category. Due to limitations of our measurement system, not every device was individually monitored, so even within this detailed level there is some grouping of devices. In this figure, we group some devices together in the Electronics, Kitchen Appliances, and Miscellaneous categories.

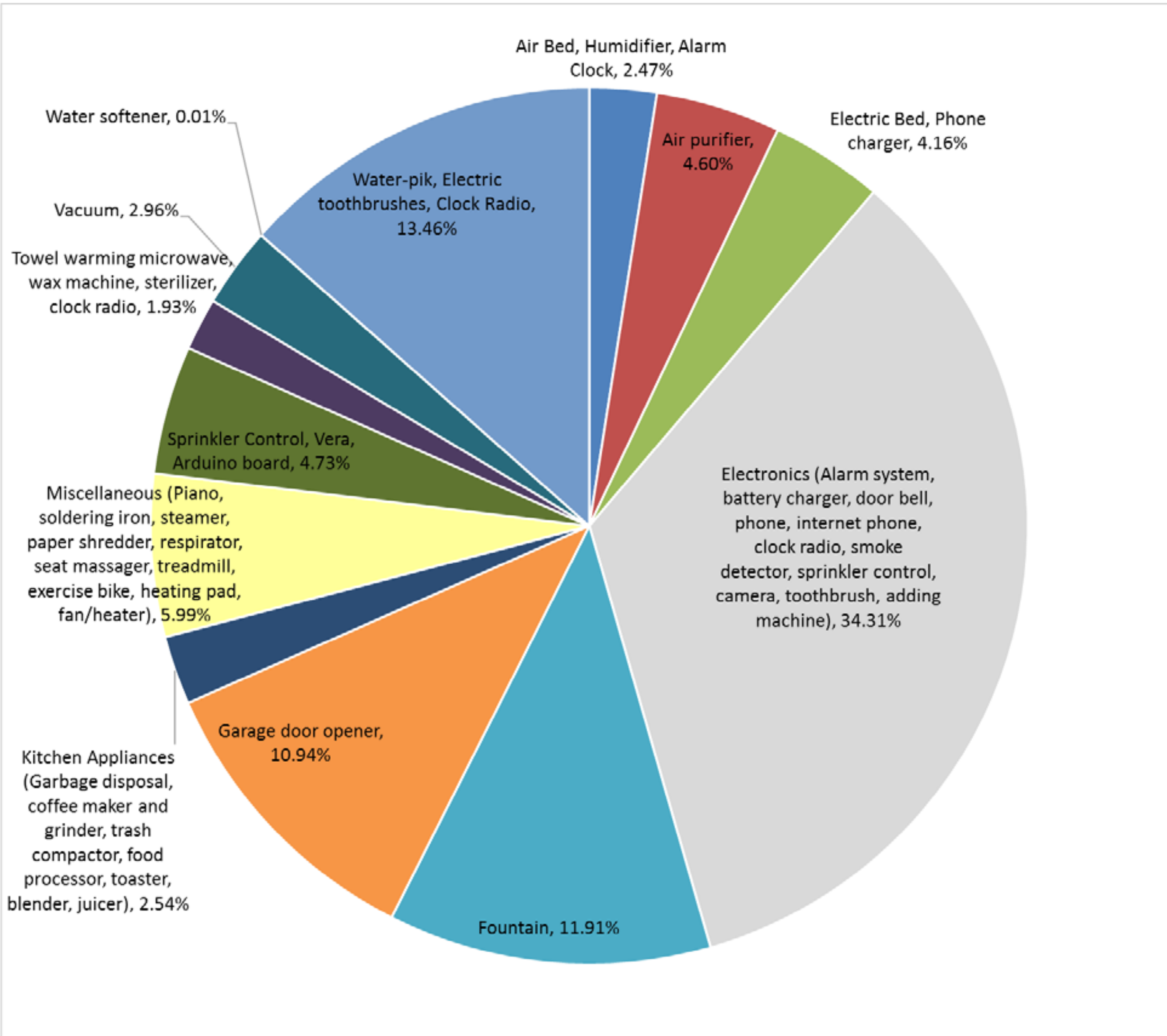

Figure 8: Unclassified Share of Disaggregation Other

Other multiple device categories were created either by circuit-level monitoring or by measuring multiple devices with a single plug load monitor (usually via a power strip).

The slice labeled Unidentified represents energy use on circuits where nothing was permanently plugged into any of the outlets powered by the circuit. We had no information about what devices are responsible for this use

Table 6 provides a complete listing of all the devices (and grouped devices) sorted by percent of energy use within the Unclassified group.

Table 6: Devices in Unclassified Group

Category/Device	% Of Unclassified End Use Energy	
Electronics	**34.31%**	
Cordless phone, adding machine		0.92%
Phone		0.13%
Sprinkler Control		0.83%
Clock radio		2.81%
Phone and Clock-radio		12.94%
Surveillance camera		0.04%
Home security power		1.05%
Alarm system		< 0.005%
Chargers		6.76%
Tool battery charger		2.71%
Web Phone Charger		< 0.005%
Battery charger		3.52%
Doorbell Chime		1.51%
Toothbrush UV		1%
Internet Web Phone		< 0.005%
Security Camera		0%
Water-pik, Electric toothbrushes, Clock Radio	**13.46%**	
Fountain	**11.91%**	
Garage door opener	**10.94%**	
Miscellaneous	**5.99%**	
Respirator		0.02%
Fan/heater and exercise bike		1.03%
Heating pad		0.82%
Treadmill		0.08%
Seat massager		< 0.005%
Soldering Iron		2.86%
Steamer		0.34%
Paper shredder		0.16%
Fan		0.14%
Piano		1%
Sprinkler Control, Vera, Arduino board	**4.73%**	
Air purifier	**4.60%**	
Electric Bed, Phone charger	**4.16%**	

Vacuum	**2.96%**	
Central vacuum		2.38%
Roomba		0.59%
Kitchen Appliances	**2.54%**	
Coffee maker		2.37%
Coffee maker and Phone		0.11%
Garbage Disposal		0.05%
Trash Compacter		< 0.005%
Air Bed, Humidifier, Alarm Clock	**2.47%**	
Humidifier		1.20%
Air Bed, Humidifier, Alarm Clock		1.27%
Towel warming microwave, wax machine, sterilizer, clock radio	**1.93%**	
Water softener	**0.01%**	
Grand Total	**100.00%**	

Our measurements also reveal the diurnal variation for each end use, as shown in Figure 9. In this figure, the thickness of a color band is the total energy used by the test sites for that end use and hour. As shown in the figure, some end uses have very predictable and uniform energy use throughout the day, such as Refrigerator_Freezer. Most end uses vary substantially, such as Clothes Dryers or Spa_Pool Pump. Even though only two test sites have spa or pool pumps they account for a large share of peak use in the early morning hours.

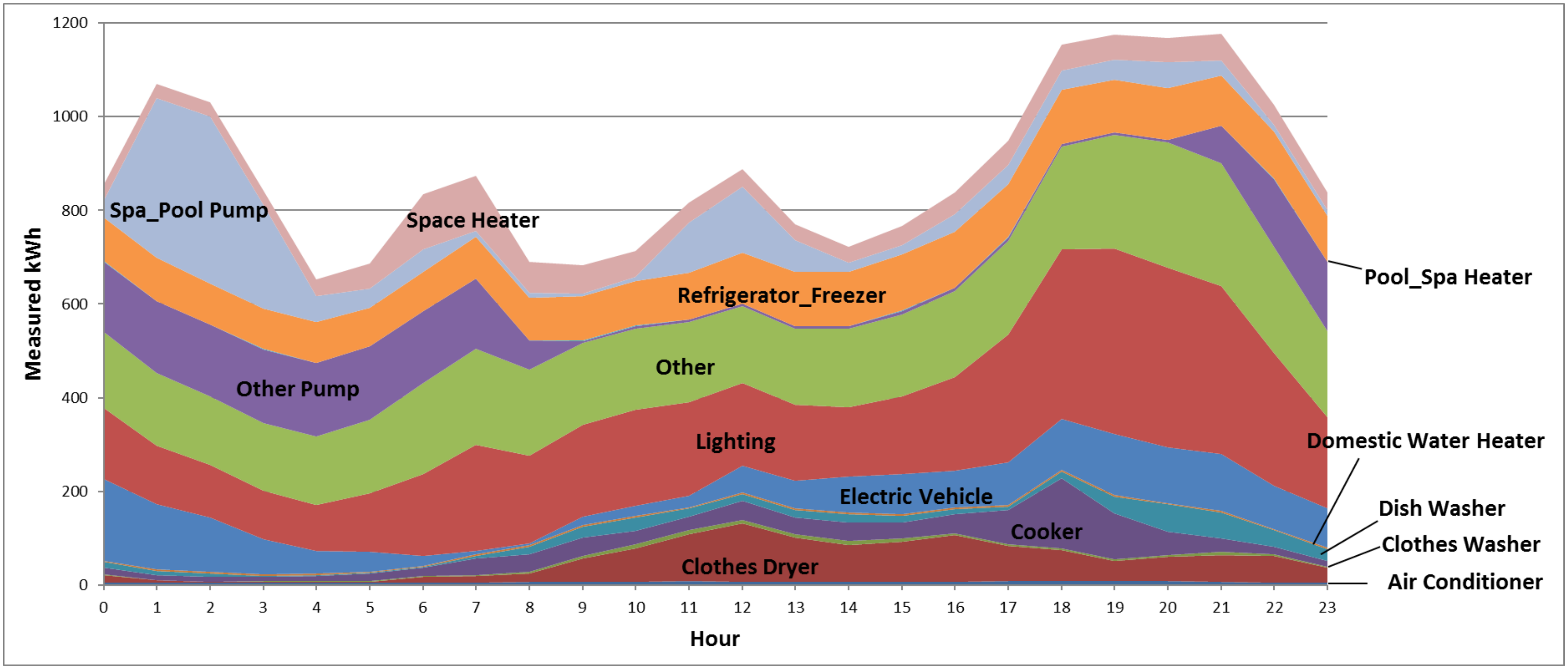

Figure 9: Average Hourly End Use (All Sites December 2014 thru April 2015)

3.2. Accuracy of Vendor Estimates

We compared our end use measurements to the estimated end uses provided by the vendor. This section describes the results of that comparison. We compared two rounds of vendor estimates. In the first round, the vendors estimated end use energy without have any information about the test homes. In the second round, the vendors were asked to re-estimate end use energy for a month, after receiving a week of our end use measurements and information describing the test homes.

3.2.1. Initial Vendor Estimates

We determined the accuracy of the vendors' products by comparing the vendors' estimates of electric end uses to our direct measurements in the six test sites. To ensure the vendor comparisons were based on the quality of the disaggregation and not the quality of the data, we excluded from the comparison data that did not meet our quality criteria, as described in Section 2.2.2. Our accuracy metric is "Percent of Measured." We examined this metric for all test homes combined, including data for the entire test period. In addition, we examined this metric by month for all homes combined, and by home and month. We have used a consistent graphical design to display these results. The results for all homes and the entire period are shown using this design in Figure 11.

PG&E may use the data obtained in this test to evaluate the accuracy of other vendor products. Therefore, so as not to reveal information that would compromise future evaluations, the kWh scale is not shown in this or other similar figures. However, each panel uses the same scale and thus the height of each bar can be compared across panels. The number that appears above each bar is the percent of the measured end use. For example, in the Electric Vehicle panel, the Blue vendor's estimated use was 70% of the measured use when we summed estimates across all homes for the test period.

Some vendors did not report estimates for all of the standardized end use requested by PG&E:

- Vendor A (Blue) – Other Pump.
- Vendor C (Green) – Clothes Washer, Dish Washer, Domestic Water Heater, Other Pump, and Pool_Spa Heater.

To allow a fair comparison of the three vendors, we display a 0 for all end uses that vendors failed to report.

As discussed in Section 2.2.2, we only used a portion of our end use measurements in assessing vendor accuracy. We only included days for which the sum of our measured end use for a home matched the PG&E smart meter measurement of total use within 5%. The bars in Figure 10 indicate the portion of total use (all homes for the entire test period) associated with each of the standardized end uses that meet our matching criteria. As shown, Lighting and Other account for about half of all use. Use related to space conditioning (heating and cooling and the related pump and fan loads) are less than 10%. This is not surprising as all of homes used gas as the primary heat source and, although the test includes homes in the warmer Stockton area, the test period was the cooler December thru April period. Use for electric vehicle charging is substantial, even though only two of the homes had chargers during the whole test period. If all homes had chargers for the whole period, this use could be nearly the same as lighting. Use for domestic water heating is also small, as all of the homes used gas for heating water and the electric use is associated with auxiliary water pumping loads. Use for spa and pool pumping was substantial, almost as large as for refrigeration and freezing, even though only two homes had such pumps. Use per pump is similar to the use per electric vehicle charger.

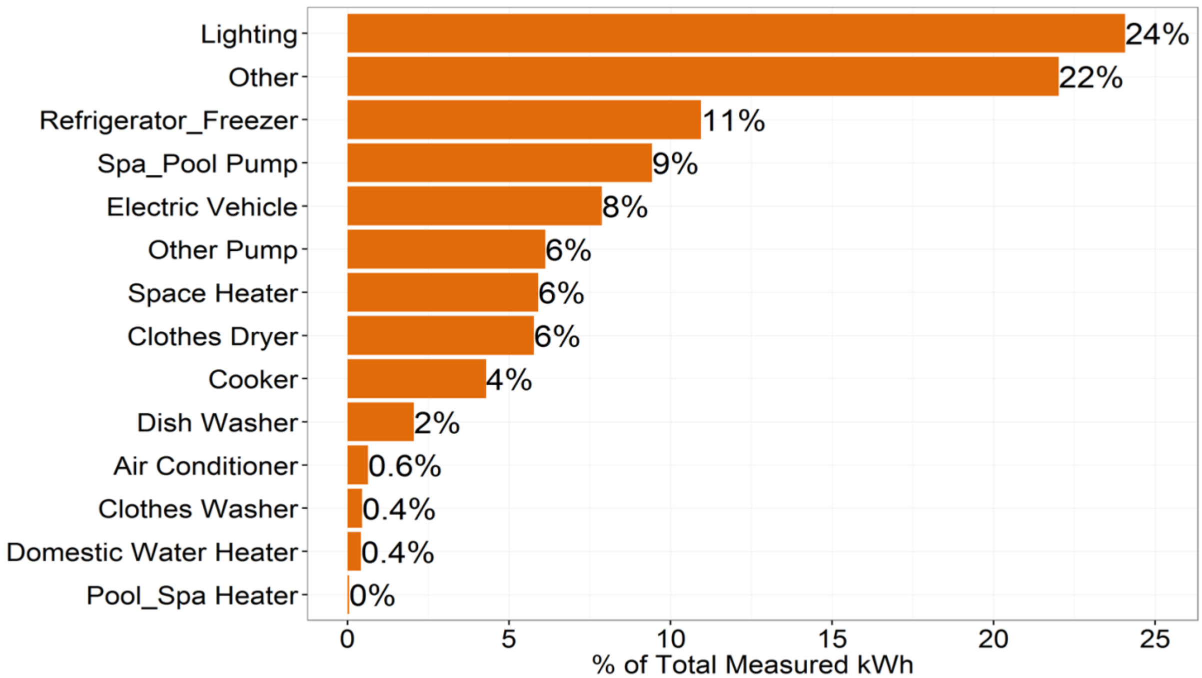

Figure 10: Measured End Use included in Assessment of Vendor Accuracy (All Test Sites - December 2014 thru April 2015)

3.2.1.1. Overall Accuracy

Overall, the green vendor comes closest to estimating the measured end uses. Green nearly matched (111%) the Spa_Pool Pump end use. However, they misestimated by more than a factor of two other end uses that account for significant shares of total use, such as Other, Other Pump and Space Heater. They also misestimated Air Conditioner by more than a factor of two, but Air Conditioner accounted for only about 1% of the total use in the test period.

The Blue and Red vendors misestimated by more than a factor of two for a larger number of significant end uses. For Blue, those end uses are Lighting, Other, Refrigerator_Freezer, Other Pump, Space Heater and Cooker. For Red, those end uses are Lighting, Other, Spa_Pool Pump, Electric Vehicle, Other Pump and Cooker. Red is within 39% of the measured Space Heater, even though none of the test homes used electricity for their primary heat source. Our measured Space Heater end use was associated with the electric fans in the heating equipment that distribute the heat throughout the home. Blue and Red both estimated nearly 70% of the Clothes Dryer end use. Blue was within 22% of the measured Spa_Pool Pump end use. Their estimate for the Pool_Spa Heater end use was off by a large percentage, but this was a very small end use.

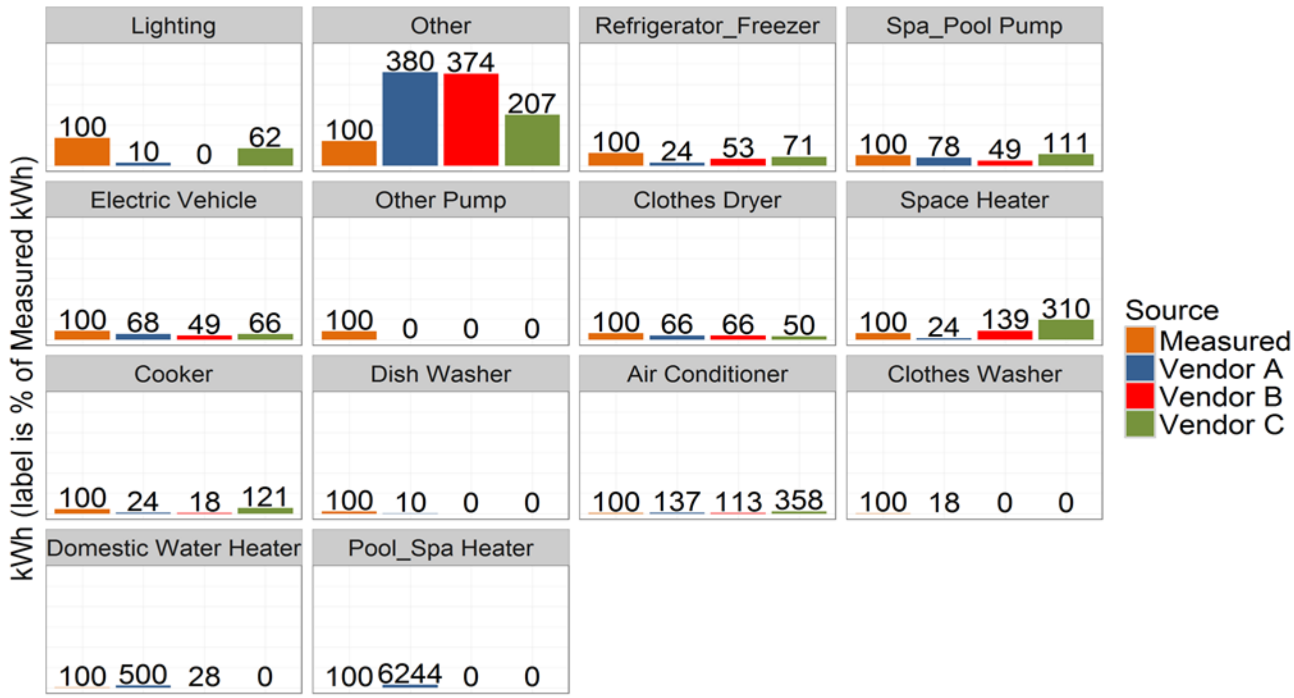

Figure 11: Percent of Measured by End Use (All Test Sites - December 2014 thru April 2015) - All Panels Use the Same Scale

Although the test period was in the winter and early spring, there was some cooling and the Red vendor estimated 113% of the measured Air Conditioner use and 139% of space heater use in this period, which was the closest of all vendors for the HVAC end uses.

All vendors had similar levels of accuracy for Electric Vehicle and Clothes Dryer. None of the vendors estimated any use for Other Pump.

3.2.1.2. Accuracy by Month

We also examined how accuracy varied across the months of the test period. As a general pattern, all vendors were less accurate in estimating end use for individual months, even when we averaged their accuracy over all test homes. This was true even for the outdoor temperature sensitive loads Space Heater and Air Conditioner. Measured lighting use varied substantially over the test period, even though we would not expect it to be a function of outdoor temperature. For the two vendors who reported Lighting estimates, their accuracy was not related to the size of the measured use.

As Figure 12 shows, the Green vendor estimates are between 46% and 84% of the measured Lighting end use in each of the months in the test period. However, the Green vendor's estimate for Space Heater (Figure 13) are high by more than a factor of two in four of five months. They estimate substantial Air Conditioner (Figure 14) use in December and January, when there is little (December) or zero (January) measured use. In March and April, they estimate approximately zero, when there is substantial measured Air Conditioner use.

The Blue vendor's lighting estimate is low by a factor of four to 33 across the months. Their Space Heater estimate is low by more than a factor of two across all months. They correctly estimated Air Conditioning use as zero in January and their estimate in December of zero is nearly correct. However, they also estimated zero for February when there is some Air Conditioning use. Their estimate for March and April, which have the largest Air Conditioning use are high by, respectively, 98% and 129%

The Red vendor estimated zero Lighting use in all months. Their estimate for Space Heater was high, but within 39% for December, January, and March. Their estimate was high by a larger margin in February and April. Their estimate of Air Conditioning use was close for December, but was high or low by more than 50% in all other months.

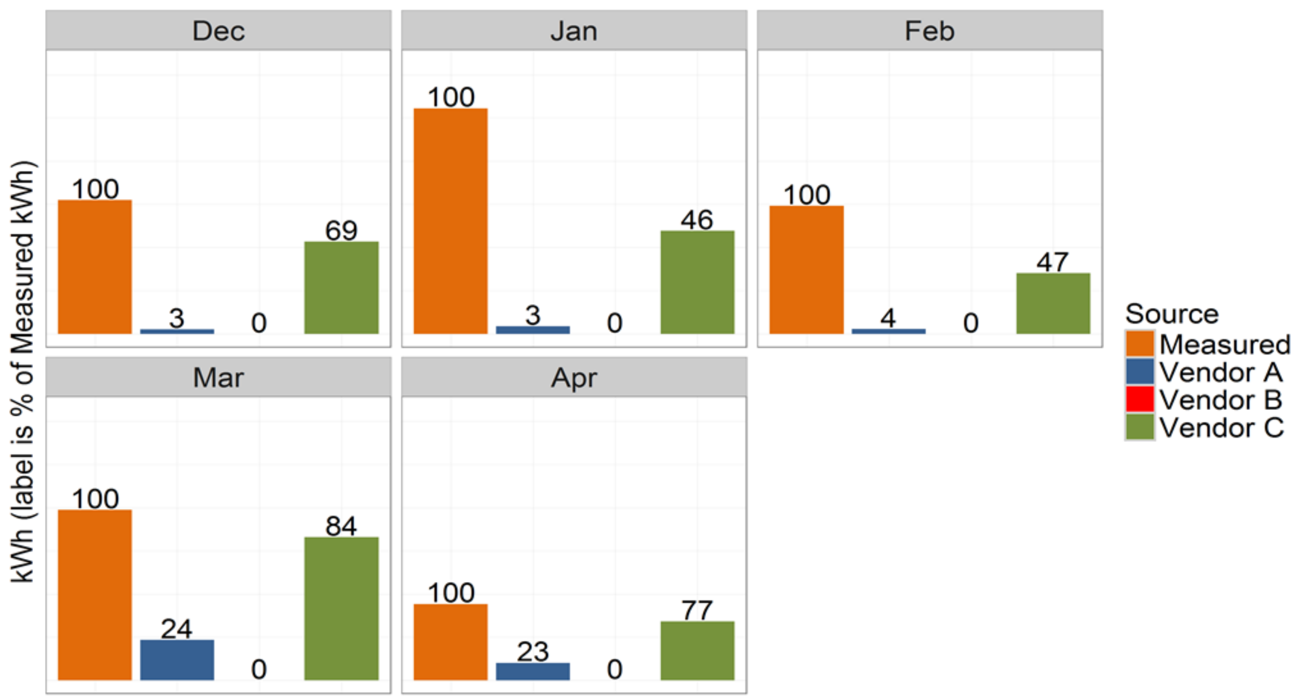

Figure 12: Percent of Measured for Lighting by Month (All Sites) - All Panels Use the Same Scale

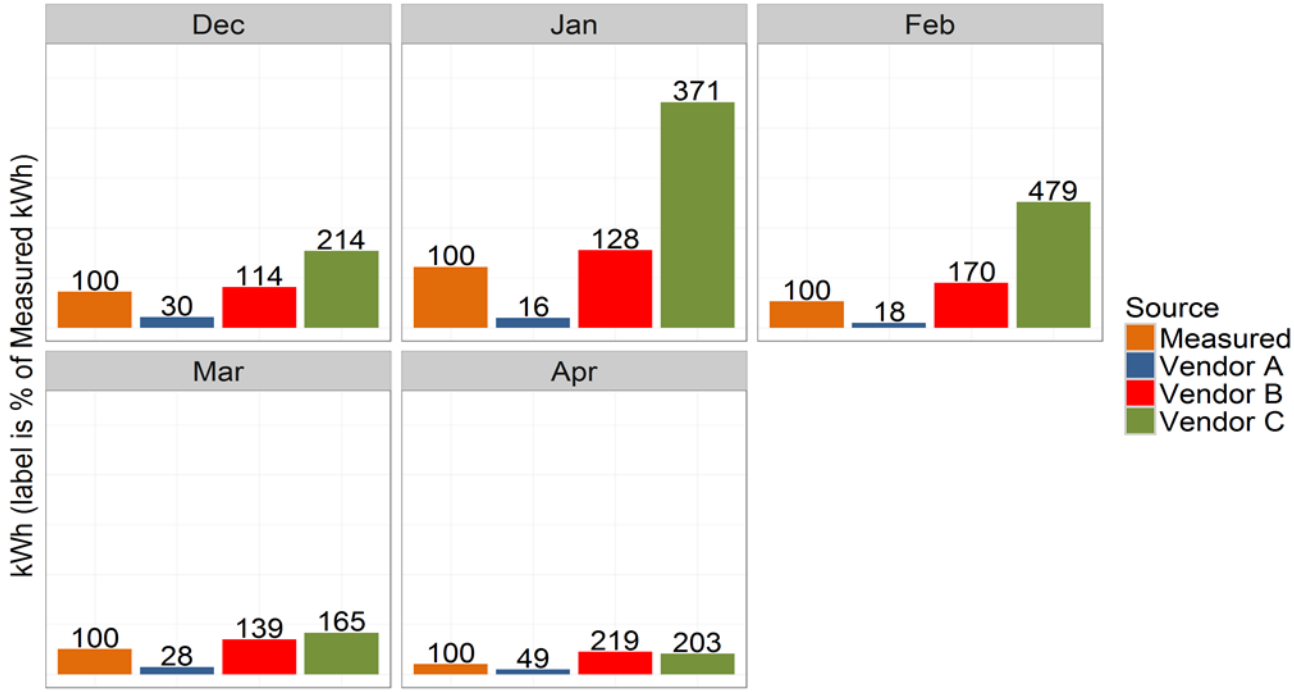

Figure 13: Percent of Measured for Space Heater by Month (All Sites) - All Panels Use the Same Scale

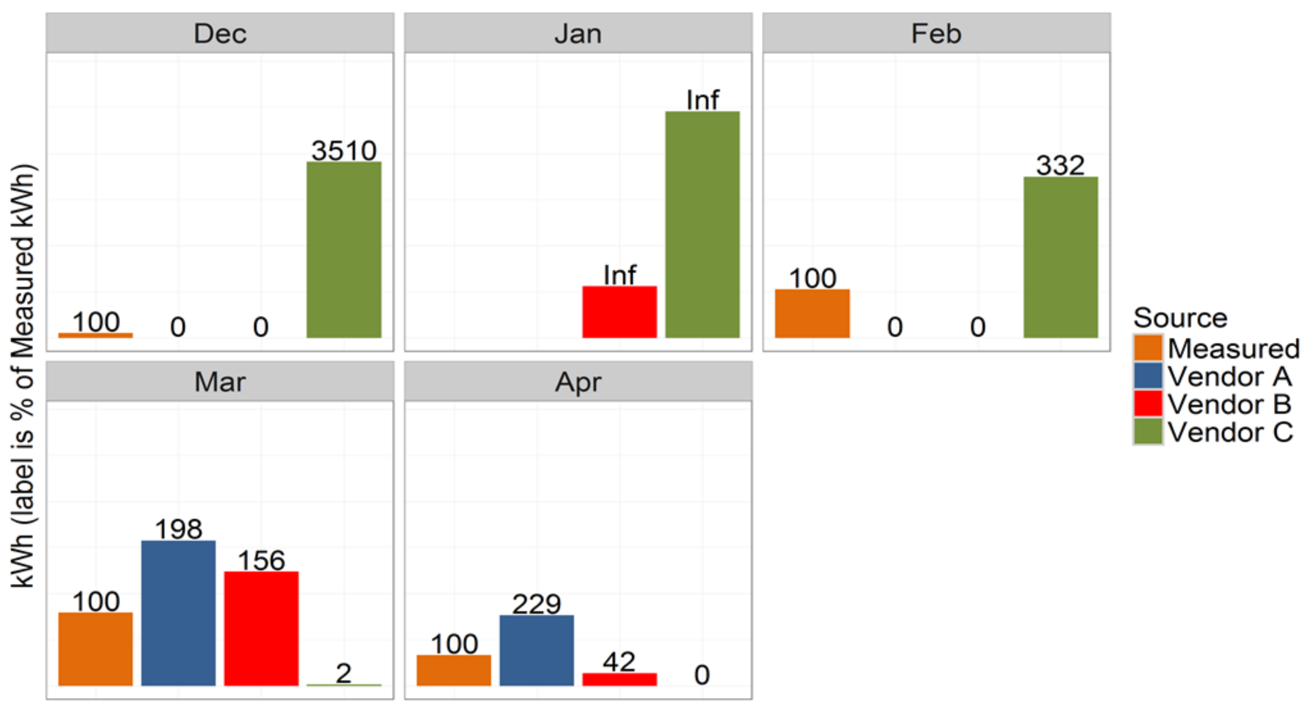

Figure 14: Percent of Measured for Air Conditioner by Month (All Sites) - All Panels Use the Same Scale

3.2.1.3. Accuracy by Site

We also examined how accuracy varied across sites. For each site, we summed end uses across all months in the test period. When we looked at accuracy by site, we could also assess how well vendors did in detecting end uses, as some end uses, such as Spa_Pool Pumps and Electric Vehicle, are only present in certain homes.

Lighting is substantial at all sites, but varies significantly between them. As shown in Figure 15, the Green vendor does the best job of estimating lighting, although their estimate is high for two sites and low for four others. In two cases, the estimates are low by more than a factor of two. As noted earlier, the Red vendor did not estimate Lighting for any site. The blue vendor's estimate is low by a factor of at least four on all sites

As with Lighting, the Other end use (Figure 16) is substantial for all sites, but highly variable across the sites. All vendor estimates are high by more than a factor or two, except for Green's estimate for two of the sites and Red's estimate for one site. Red and Blue are high by more than a factor of five for two of the sites.

Spa or pool pumps were present in two homes, as shown in Figure 17. These end uses are large consumers of electricity. Blue detected this end use in all homes. Red detected this end use only in the two homes where it was present. Green detected this end use in those two homes, but also estimated substantial use in a third home that did not have this end use. Four of the six estimates of this end use, in homes where it was present, were within 40% of the measured use.

Like Spa_Pool Pump, the Electric Vehicle end use (Figure 18) is large when present. Two of the homes had chargers for electric vehicles at the beginning of the test and one installed a charger part way through the test. Only Blue detected all three chargers. Red and Green only detected this end use at one home. Red was within 20% of the measured use for one home and Blue was within 15% of the measured use for another home.

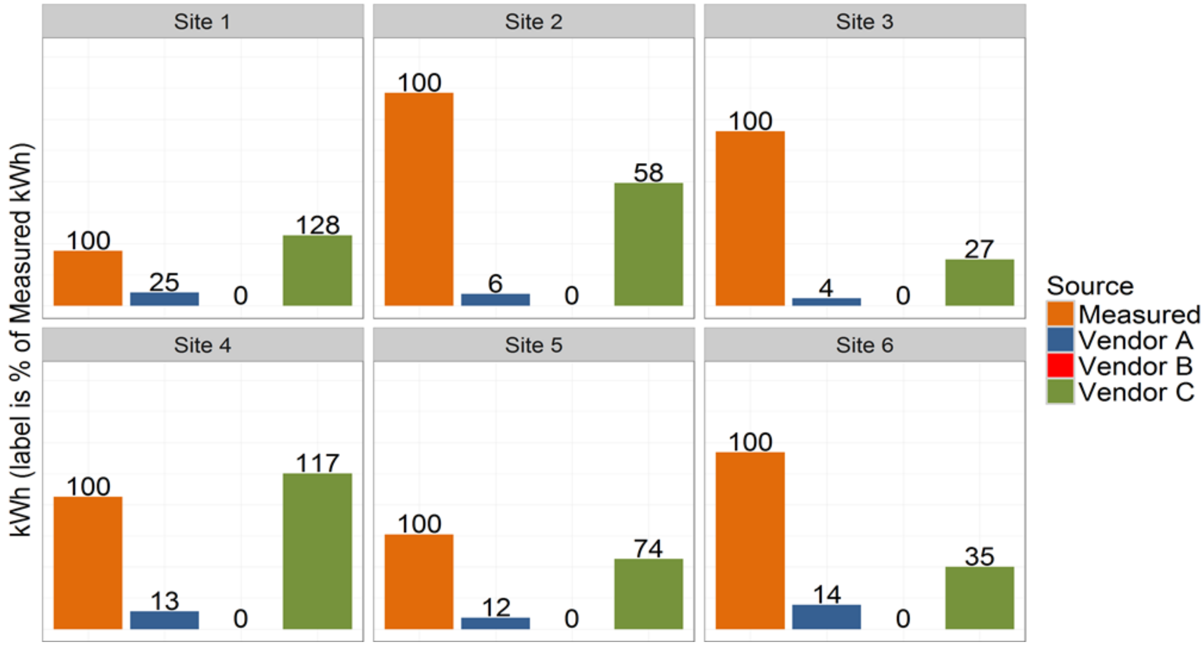

Figure 15: Percent of Measured for Lighting by Site (December 2014 thru April 2015) - All Panels Use the Same Scale

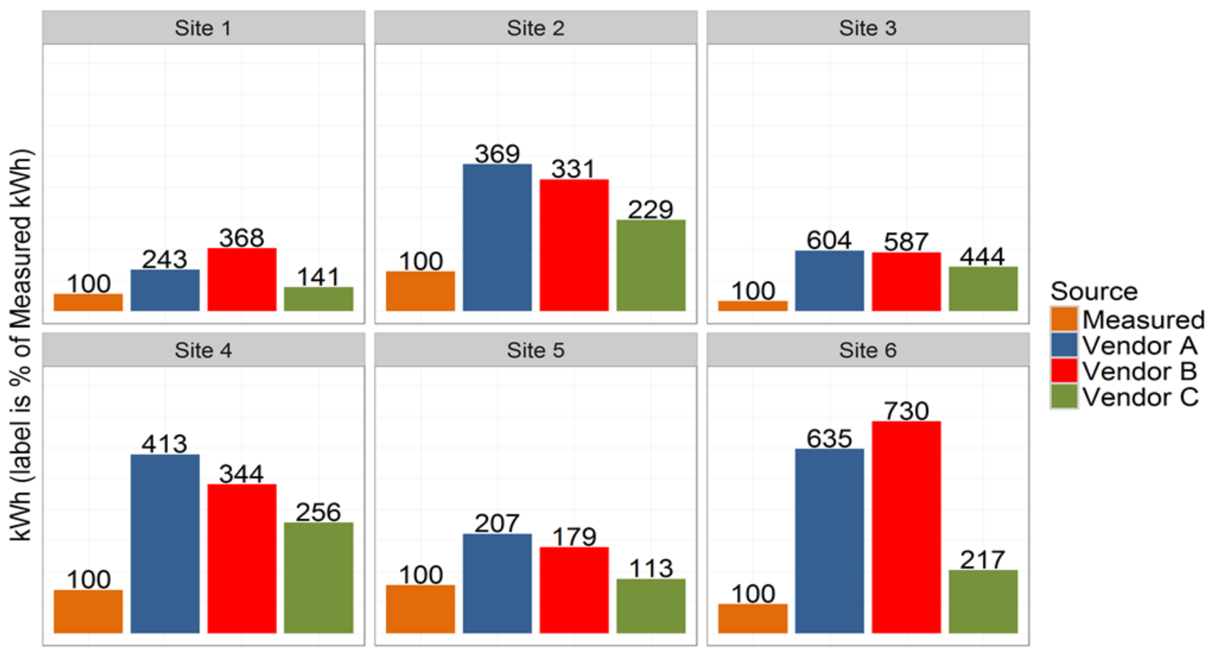

Figure 16: Percent of Measured for Other by Site (December 2014 thru April 2015) - All Panels Use the Same Scale

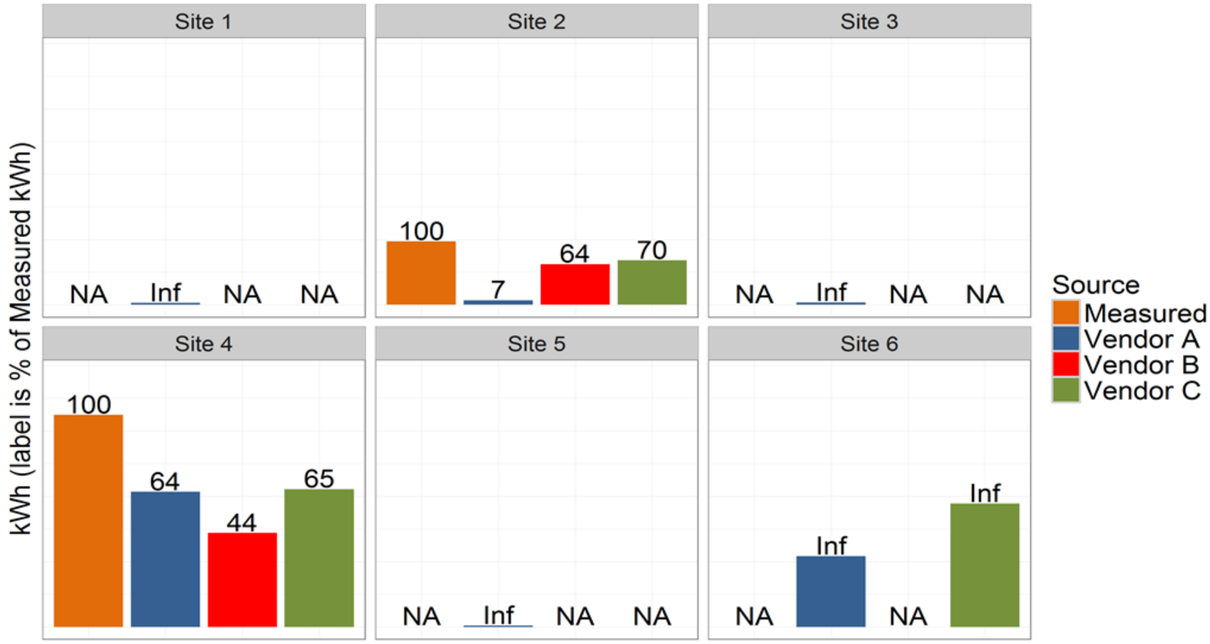

Figure 17: Percent of Measured for Spa_Pool Pump by Site (December 2014 thru April 2015) - All Panels Use the Same Scale

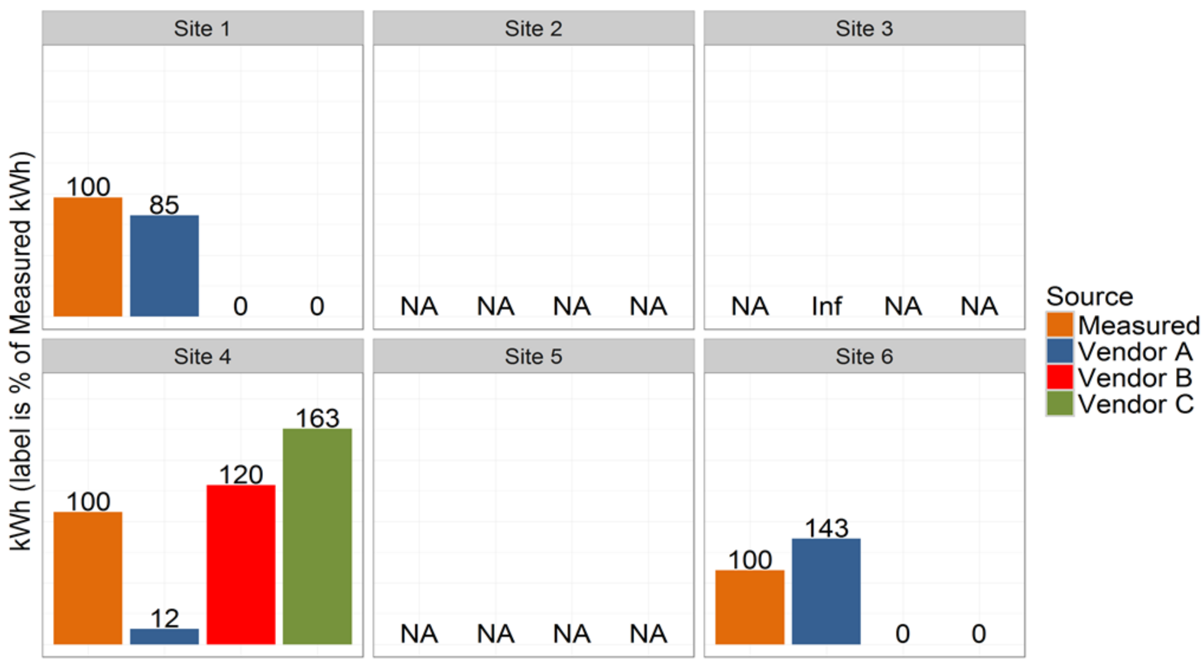

Figure 18: Percent of Measured for Electric Vehicle by Site (December 2014 thru April 2015) - All Panels Use the Same Scale

3.2.2. Impact of Sharing Additional Data

We conducted an experiment to see whether sharing some of the measured end uses with the vendors would improve their estimates of end use. After the Vendors provided their initial estimates for the entire study period, they were given some of the measured data. In order to preserve the confidentiality of the collected data for use in testing future disaggregation algorithms, the vendors were only provided with one week of actual measurement data. The remainder of the data was not (and will not be) released. The vendors were given the measured one-minute channel level data for the 6 test sites for March 1-7 and supplementary data they could use to map the channels to the assigned end use and process the data properly. In addition, the vendors were supplied basic information about the households including the number of people and the appliance information (type and count). They were then asked to resubmit their end use estimates for March.

The additional data was not universally helpful for the vendors. However, it did lead to substantial improvements for most of the large end uses. Figure 19 shows the impact combined across all vendors. The impact of the additional data was not consistent across end uses. For example, the vendor estimates improved for the Refrigerator_Freezer end use but worsened significantly for Space Heater. Among the substantial end uses, accuracy was most improved for Spa_Pool Pump and Electric Vehicle. The improvement in Electric Vehicle was mainly due to the vendors improving their detection of this end use as demonstrated in Figure 20 for Vendor B.

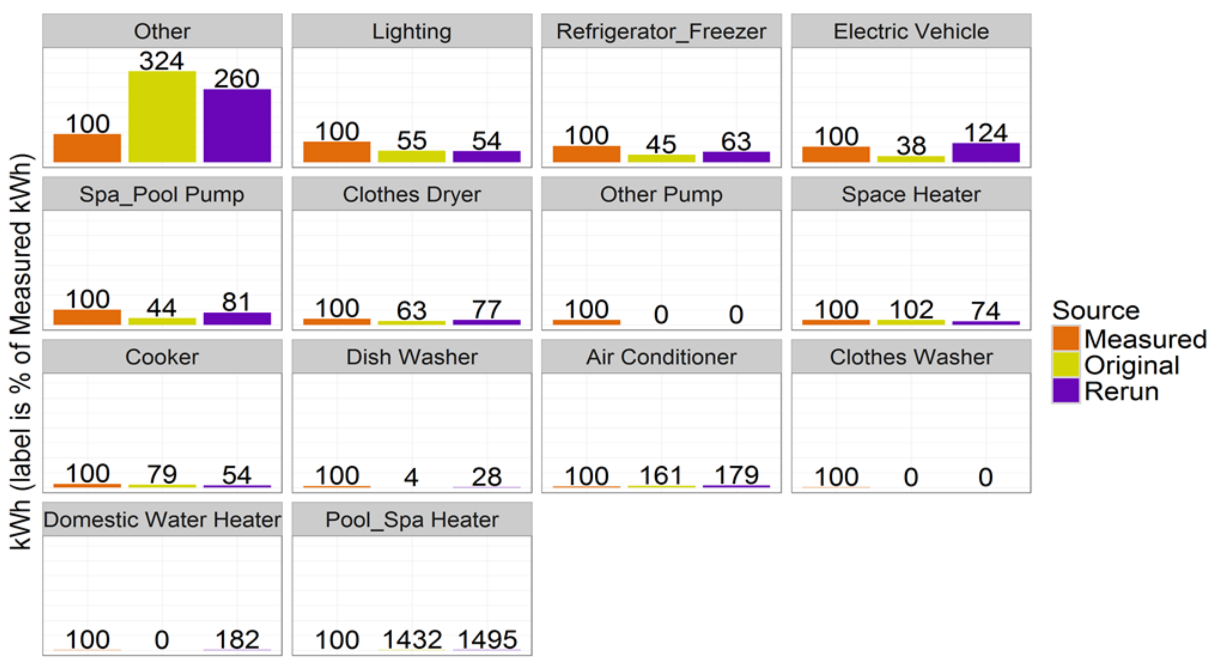

Figure 19: Changes in Percent of Measured Energy by End Use (March 2015) - All Panels Use the Same Scale

Figure 20: Vendor B - Changes in Percent of Measured Energy for Electric Vehicle End Use by Site (March 2015) - All Panels Use the Same Scale

3.3. Limitations of this Test

There are three major limitations of this test that should be kept in mind when considering the accuracy results.

1. **Sample Size**. As discussed in section 2.1.1, a diverse collection of six homes was selected for this test. However, an intentionally selected (non-random) sample of six homes is not statistically representative of the single-family home population served by PG&E. Thus, the accuracy findings cannot be taken as typical of what would be achieved if all such customers were tested. However, the objective of the EPIC project was to determine whether it was possible to accurately estimate end uses for individual homes and not on average across many homes. The six test sites provide a strong indication of the whether or not this objective can be achieved.
2. **Period**. Resources for this test only allowed it to be sustained for five months, starting in December 2014. Even for homes in the Stockton area, this is the coolest portion of the year. Air conditioning is not frequently required during this time of year. The accuracy for the Air Conditioner end use might be different if the test had been conducted for an entire year. However, there was still significant cooling energy used by some homes, and the test provides some indication of how well the vendors could estimate that use.
3. **Mixed End Uses**. Substantial effort was expended in reviewing 37 candidate test sites to minimize mixed end uses and confounding factors such as net metering due to solar electric equipment. We also selected homeowners who were willing to be involved throughout the test period in helping to maintain accurate end use measurements. However, most homes

have some devices that consume energy in more than one end use, even if it is only a single bathroom ventilation system that includes a light fixture. We had to assign these mixed devices to a single end use. These mixed end use devices introduce little error, given their rated capacity and likely use pattern.

3.4. Comparisons with Other Research

Considerable research has been conducted over the last 30 years on the topic of end use disaggregation. Much of this research focuses on whether disaggregation is possible using special meters that record high time-resolution whole house electric use. The meters used often record use more than 1,000 times per second. Recently, some vendors, such as the ones that participated in this study, have developed disaggregation products that can be used with data that is practical to collect with PG&E's smart meters (one-minute intervals). We conducted a brief examination of three studies relevant to PG&E's smart metering. The purpose of this examination was to determine whether they have reached similar conclusions regarding the accuracy of disaggregation products.

We found that two of these studies provided relevant results. Although there are substantial differences in the study designs, we find they are largely consistent with the results from PG&E's test.

3.4.1. SDG&E Study

In August of 2014 a study[4] of disaggregation vendor performance sponsored by the San Diego Gas and Electric Company (SDG&E) was released. The report, produced by NegaWatt Consulting, was titled "Residential Disaggregation" and it focused on verifying the accuracy of non-intrusive load monitoring. We will refer to it as the NILM study. On the surface the NILM study appears similar to this study. However, there are significant differences that make it difficult to compare the results of the two studies.

The NILM study included four disaggregation vendors and 11 SDG&E employee homes. The vendors were allowed to choose which appliances (or appliance groups) they would estimate from the whole building metering. The study installed gateways to the SDG&E smart meter to gather and record 10-second power measurements, which were provided to the vendors. A summarized version of these 10-second measurements was also prepared (at one-minute and 15-minute intervals) and provided to the vendors. In addition, hourly data from the SDG&E Green Button Connect program was provided to the vendors. Power meters were installed for selected circuits where only an appliance was on the circuit. In the case of refrigerators, which could not be isolated on a circuit, the circuit data was processed to remove the energy that appeared to be from other devices on the circuit. Values near zero were eliminated from the data because they "significantly decrease accuracy."

[4] *Residential Disaggregation,* Emerging Technologies Program, Project ID ET13SDG1031, San Diego Gas And Electric Company, 8/22/2014.

The vendors were also supplied with the zip code of the homes, and in a second stage of the research, home appliance survey data. There was a third stage where vendors were provided with plots showing the results of their disaggregation and given the opportunity to improve (only one vendor chose to participate in the third stage).

The vendors could use any or all of the four data sets provided. For the first stage, three vendors used the 10-second data and three vendors used the Green Button (GB) data. However, the number of appliances reported from the GB data is less than one-third of the number of total appliances measured. This would seem to indicate the vendors have more confidence in the 10-second data. When a vendor used both the 10-second and GB data, the GB accuracy was better in two out of six cases.

The vendors chose which appliances they reported data for, the frequency of the reported data (daily and/or monthly was recommended), and whether to participate in the post-appliance second stage and preliminary result third stage. Further, the vendors were not required to provide estimates for the entire timespan, but could submit only selected days. As would be expected with so many choices, the vendors each did different things making relative comparisons difficult. It might also be inferred that the vendors self-selected the appliances that they felt most confident with and therefore the results represent the best they can do.

The NILM study focused on the EPRI accuracy value (which is why our study also started with that as our primary metric), but also reported several other more typical statistical metrics. Overall there were a total of four (out of 25 total from 10-second and GB datasets) Stage 1 results that had a good EPRI accuracy value: one in EV, two in pool pumps, and one in refrigeration. The average of those four accuracy values was .755, with the cutoff being .7 to be considered good. Stage 2 saw three more good EPRI accuracy values, all in the same three appliance categories, one of which was by the same vendor for the same appliance but at the monthly reporting level. Stage 2 also had three cases of reduced accuracy. Stage 3 had six reported values; two showed a decrease in accuracy and three had a significant increase in accuracy.

In general, the NILM study found that for large or regularly running appliances (electric vehicle, pool pump, and refrigerator) the vendors performed well and thought the NILM algorithms were "promising." They also acknowledged that accuracy varies widely across homes and that the vendors did poorly on many of the appliances.

There were two big differences between the NILM study and our study. First, our study relied on 1-minute data from the PG&E smart meter data. This was accomplished thru a firmware update to existing smart meters and could be done without any customer involvement. PG&E determined that 1-minute was the shortest practical interval for their smart metering system. The NILM study relied on additional gateway hardware to achieve 10-second energy measurements. This approach would require the cooperation of each homeowner. Second, the NILM study focused on the estimate of use for selected appliances. Our study tested the ability of vendors to estimate standardized end uses, which in many cases combined use from a number of devices associated with the same end use. It is possible that the shorter interval measurements used in the NILM study would allow for improved results for standardized end uses, but this would have to be confirmed by additional testing. What both studies make clear is that the vendors still have a lot of room for improvement.

3.4.2. Pecan Street Study

In January 2015, Pecan Street published a paper on behalf of EEme titled *Setting the Benchmark for Non-Intrusive Load monitoring: A Comprehensive Assessment of AMI-based Load Disaggregation*. We will refer to this as the Pecan Street Study. The paper recommends some alternative metrics for evaluating smart meter disaggregation.

First, there is an acknowledgment that the field validation of smart meter disaggregators by various researchers places accuracy at around 55%. The point is made that small sample sizes and short time spans are factors that will limit evaluation studies. To this end, The Pecan Street Study applied EEme's dissagregation algorithm to one year of 15-minute data from 264 homes. Direct measurements were made on 12-24 circuits in each home with one-minute sampling intervals.

Error ratios were calculated in two forms:

$$Absolute\ Error = \frac{Inferred\ appliance\ use - Measured\ use}{Measured\ use}$$

$$Relative\ Error = \frac{Inferred\ appliance\ use - Measured\ use}{Total\ measured\ home\ use}$$

The argument is made that absolute error has a skewing effect on reported error. Larger errors for small total end uses may produce a larger impact than a small error on a large end use. Relative error on the other hand, will normalize the contribution of any particular appliance to the total energy use of the home. Figure 21 is a reproduction of a table from the Pecan Street Study that shows the median monthly errors over a full year are presented for four end uses.

	Absolute error	Relative error
HVAC	-0.31	-0.11
Refrigerator	-0.28	-0.02
Clothes Dryer	-0.45	-0.02
Dishwasher	0.33	0.003

Table 1

Figure 21: Pecan Street Study Results for Median Monthly Absolute and Relative Error

The absolute error is defined in the same manner that we defined accuracy for PG&E's test homes. However, Pecan Street reports median values. The absolute error achieved by EEme is similar to the accuracy achieved by some of PG&E's vendors for some end uses, however, their use of median values makes a comparison difficult. We designed our comparisons of measured to vendor estimated end use so all end uses appear of the same scale. The Pecan Street relative metric achieves a similar purpose, but as shown in their study it is still necessary to consider the absolute error for each end use.

It is certainly desirable to perform an accuracy test for a large group of homes and using a complete year of measurements. However, PG&E's objective was to determine whether vendor accuracy was sufficient to report a comprehensive end use breakdown as part the monthly electric bill sent to individual homes. The average accuracy of many homes does not directly address PG&E's objective.

3.4.3. Energy Information Administration Study

This study was conducted for the U.S. Department of Energy – Energy Information Administration (EIA) by Leidos. It reports on interviews with utilities, regulators and sub-metering equipment vendors. It reports on the prevalence of smart meters and sub-metering across the country and draws conclusions about likely trends. One sub-metering company was interviewed that uses data from a whole-house meter to estimate end uses. The other sub-metering companies rely on direct measurements of circuits or plug loads.

This report does not explore the accuracy of whole-house disaggregation methods. It describes data collected by Pecan Street and the wiki they have setup to share that data with researchers, and says that the data can be used in "AMI meter disaggregation algorithm verification."

Unfortunately, this report does not provide any information on the accuracy of disaggregation products.

4. Conclusions

In this chapter, we summarize our conclusions from this test. These address the accuracy of vendor end use estimates, findings from the review of similar studies, the feasibility of disaggregated monthly electric bills, and lessons learned about end use measurement.

4.1. Accuracy of Vendor Estimates

We reached the following conclusion regarding accuracy.

Overall Accuracy

We combined all end uses across the six homes and five months of the test and found:

- Overall, the green vendor comes closest to estimating the measured end uses. Green nearly matched (111%) Spa_Pool Pump end use, which can be large when present in a home. However, they misestimated by more than a factor of two other end uses that account for significant shares of total use, such as Other, Other Pump and Space Heater.
- The Blue and Red vendors misestimated by more than a factor of two for a larger number of significant end uses. For Blue, those end uses are Lighting, Other, Refrigerator_Freezer, Other Pump, Space Heater and Cooker. For Red, those end uses are Lighting, Other, Spa_Pool Pump, Electric Vehicle, Other Pump and Cooker.

Accuracy by Month

For each end use we combined all homes, but examined the accuracy by month. As a general pattern, all vendors were less accurate in estimating end use for individual months, even when we averaged their accuracy over all test homes. This was true even for the outdoor temperature sensitive loads, Space Heater and Air Conditioner. Measured lighting use varied substantially over the test period, even though we would not expect it to be a function of outdoor temperature. For the two vendors who reported Lighting estimates, their accuracy was not related to the size of the measured use.

Accuracy by Site

For each end use we combined months, but examined the accuracy by site. When we looked at accuracy by site, we could also assess how well vendors did in detecting end uses, as some end uses, such as Spa_Pool Pumps and Electric Vehicle, are only present in certain homes. Spa or pool pumps were present in two homes.

- Blue detected this end use in all homes. Red detected this end use only in the two homes where it was present. Green detected this end use in those two homes, but also estimated substantial use in a third home that did not have this end use. Four of the six estimates of this end use, in homes where it was present, were within 40% of the measured use.
- Two of the homes had chargers for electric vehicles at the beginning of the test and one installed a charger part way through the test. Only Blue detected all three chargers. Red and

Green only detected this end use at one home. Red was within 20% of the measured use for one home and Blue was within 15% of the measured use for another home

Impact of Sharing Data on the Homes

After the vendors provided their initial estimates for the entire study period, they were given some of the measured data. The additional data was not universally helpful for the vendors. However, it did lead to substantial improvements for most of the large end uses. Vendor estimates improved for the Refrigerator_Freezer end use but worsened significantly for Space Heater. Among the substantial end uses, accuracy was most improved for Spa_Pool Pump and Electric Vehicle.

4.2. Findings from Similar Studies

We examined three recent studies that addressed end load disaggregation. Only two of those studies estimated accuracy for tests based on comparison of vendor estimates to end use measurements.

- In August of 2014 a study of disaggregation vendor performance sponsored by the San Diego Gas and Electric Company (SDG&E) was released. Vendors were given access to 10-second data and the study focused on estimating use for specific appliances and not for standardized end uses, as was done in the PG&E test. In general, the study found that for large or regularly running appliances (electric vehicle, pool pump, and refrigerator) the vendors performed well. The study also acknowledged that accuracy varied widely across homes and that the vendors did poorly on many of the appliances.
- In January 2015, Pecan Street published a paper on behalf of EEme titled *Setting the Benchmark for Non-Intrusive Load monitoring: A Comprehensive Assessment of AMI-based Load Disaggregation*. The absolute error achieved by EEme is similar to the accuracy achieved by some of PG&E's vendors for some end uses. However, even though the Pecan Street defines absolute error in the same way we did for the PG&E test, the chose to report the median accuracy across the site tested, which makes it hard to compare to our findings.

4.3. Feasibility of Disaggregated Monthly Bills

Based on the data obtained and analyzed under this effort, our results suggest that the tested products available at the onset of this study would not be able to consistently and accurately disaggregate residential home energy usage given the same constraints as the test environment established in this project. While our findings indicate that vendor estimates were close to actual measured values in certain limited instances, these instances represented end uses that constituted a small fraction of overall home energy use. Focusing on only the largest energy end uses, our results indicate that PG&E would have to accept a greater than ±30% error tolerance in disaggregation estimates. The findings from this test indicate that product advancements are needed to improve accuracy.

4.4. Lessons Learned about End Use Measurement

Following are lessons learned in implementing the end use measurement systems for the test homes:

- **End use contamination**. Some measured devices, such as bathroom ventilation units, which contained both a fan and light fixture, were of mixed end use. We had to assign these devices to an end use based on an estimate of which end use accounted for the largest share of the device's use over the test period. Other examples were pumps that served both space and water heating or fans that served both space cooling and space heating.
- **Very small or temporary loads.** We did not attempt to measure the energy use of power strips. In addition, and probably more important, we only measured devices that were always plugged into a wall outlet or had a hard-wired connection to a circuit. Devices like vacuum cleaners stored in a closet were not directly measured. However, it is possible their energy use was captured at the circuit level and included in one of our measured end uses.
- **Measurement resolution.** The resolution of our electrical measurements was .001 kWh. Some devices used less than this amount in a minute. For such devices, the use would be reported after sufficient time has passed to accumulate at least .001 kWh, making the intervening one-minute intervals appear to have no energy use.
- **Check-sum Measurements.** Check-sum measurements, provided by the PG&E smart meter and obtained by our team in measuring total use for electrical panels, are critical to resolving installation errors. However, it is not always possible to measure the total use of certain breaker panels.
- **Measurement Planning.** It takes substantial time to determine what circuit powers each device in a home. Further time is required to characterize each the load for each of these devices, which is needed to develop effective quality test for the end use measurements.
- **Managing Changes**. Regular contact with homeowners participating in the test was necessary in order to identify changes to devices in the homes (moves, additions and removals) that would require updates to the measurement system.
- **Wireless Home Networks**. It is challenging and time-consuming to install and debug Z-wave and Zigbee networks in homes and there are few options for controllers that will connect to both types of wireless networks and that support conventional circuit-level power monitoring equipment.

5. Recommendations

In this chapter, we offer recommendations that may help guide future research on end use disaggregation and other applications of data collected from the test homes.

Reserve test data for future vendor comparison

The data collected for this test could be used to test other vendor products. Vendors would be provided the PG&E smart meter data for some or all of the 166 homes for which 1-minute readings were obtained. For a blind test of accuracy, PG&E should not disclose the identity of the homes with end use measurement systems. The identity of the homes was disclosed to the three vendors that participated in this test. PG&E should request that those vendors destroy all of the information they received during this test so that it does not contaminate future blind tests of accuracy. All of the routines (created in R or Excel) for comparing vendor estimates and measure end uses are included in the data and documentation that accompany this report and can be re-used in future tests.

Exploit device-level data for load and potential studies

We obtained measurements of energy used by many specific devices, such a pumps, fans, heaters, televisions, and game consoles. Although, these are not statistically significant samples of such devices, the energy use patterns may be useful in understanding the impact of such devices in load forecasts or the assessment of energy efficiency potential.

Apply full test data to calibrate and test building efficiency models

Although the test of vendor accuracy was limited to five months, end use measurements continued for the six test homes for a full year. In addition, gas end uses were measured in all homes. These use measurements, in combination with detailed data on devices measured, could be useful in improving building efficiency modeling tools. The data could be used for model calibration and for testing the accuracy of these models. These models, such as EnergyPro (used by Energy Upgrade California), also estimate various end uses. The data could be used to determine the accuracy of this type of model.

Leverage test data to develop improved disaggregation methods

The data we collected could be used to develop publicly available algorithms for disaggregation. If public domain algorithms were developed, it might not be necessary to conduct blind tests. Instead, multiple researchers, unimpeded by vested interest in the disaggregation products, could test the accuracy of the algorithms. PG&E should be cautious in committing resources to the development of new algorithms. There would be great benefits from algorithms that are proven accurate, but this and other recent tests demonstrate that it is a difficult problem to solve.

A. Appendices

A.1. Public Comments and Responses

PG&E posted a draft version of this report for public comment on January 6, 2017, with a two-week window for receiving feedback. As of the January 20 deadline, two comments from one party had been received. These comments are shown verbatim in Table 7, along with our responses.

Table 7: Public Comments and Responses

	Commentor	Comment	Response
1	Steve Schmidt (HEA)	**Why are prior public comments not included?** *Appendix A.1, page 52* This appendix is titled "Public comments and responses", but it is empty. I personally submitted 15 comments to the draft released in August 2016 (via this same EnergyDataWeb service; see them here: http://www.energydataweb.com/cpuc/comment.aspx?did=1599) and I see no responses online or in this document. Why would you exclude any public comments from this Appendix?	As noted in the errata on page ii, "*To fully address comments received after stakeholder review of the draft report in August 2016, SBW Consulting, Inc. reanalyzed all data and prepared a revised report with additional details. This current report reflects SBW's complete reanalysis of the data and results.*" As previously communicated, the original report was replaced by this revised report, and we only accepted comments on the revised report.
2	Steve Schmidt (HEA)	**Lighting end use of 27.4% is wrong** *Figure 7, page 29* This chart indicates the "Light" end use accounts for over a quarter of electric use across all six homes that were monitored. Perhaps this is a typo? If not: this figure is substantially higher than recent California estimates which show a much lower (and dropping) figure for lighting. This unreasonably high figure casts doubt on the other figures in the report. Since the majority of the lighting measurements (114 of 180, per Table 3 on page 16) were done using "Subtraction", it would be more accurate to label this category as "Lighting/Other", since there was no way to know whether the end use was lighting or some hidden plug load. This comment applies to Figure 10, and also	Figure 7 reflects the actual conditions measured during this metering effort. Review of the data for lighting showed typical use patterns, in addition to periods of higher lighting use at night at specific sites. The data did not reveal any systemic issues with the presented lighting results. Deviations from average or typical use profiles are commonplace throughout the PG&E territory, as detailed in internal PG&E data gathered outside of this project. As explicitly stated in Section 3.3 (Limitations of this Test) on page 43 of the draft final report, *"As discussed in Section 2.1.1, a diverse collection of six homes was selected for this test. However, an intentionally selected (non-random) sample of six homes is not statistically representative of the single-family home population served by PG&E. Thus, the*

Commentor	Comment	Response
	to the comment "For the two vendors who reported Lighting estimates, their accuracy was not related to the size of the measured use." If the measured figure is inaccurate (much too high, since it can include hidden plug loads), perhaps the vendor's estimates were much more accurate (two were significantly lower). In this case it seems the AMI analysis may provide better accuracy than the "subtraction" method used, and this possibility should be noted.	*accuracy findings cannot be taken as typical of what would be achieved if all such customers were tested. However, the objective of the EPIC project was to determine whether it was possible to accurately estimate end uses for individual homes and not on average across many homes. The six test sites provide a strong indication of whether or not this objective can be achieved.*”