ROCKY MOUNTAIN POWER
A DIVISION OF PACIFICORP

Sampling Plans, Procedures and Selections For the Profile Metering Sample of

Utah Residential Distributed Generation 2014

Prepared and Published by the
Load Research Group
May 2014

May 9, 2014

Executive Summary

This sample design was prepared in support of the Load Research Company Cost-of-Service commitment, with the intent of installing a load study on the Company's Utah Residential Distributed Generation Class. All sample designs were prepared in accordance with PURPA standards and, as such, are expected to provide estimates of system peak demand that achieve, at a minimum, ±10% precision at the 90% confidence level.

The recommended sample design for this study incorporates four strata and calls for the installation of **62** load recorders. Based on the level of recorders installed, the sample design estimates an achieved precision level of ±10% at the 95% confidence level. The strata boundaries for this sample are based upon the "cumulative square root of *f*" rule as defined in studies by Dalenius/Hodges. Appendix 1 contains a listing of both primary and alternate sample sites selected for this study. These sites have been cross referenced against current installations, and any duplicates have been noted.

Scott D. Thornton
Manager, Load Research

Sample Parameters & Design

Utah Residential DG (2014)
Load Recorder Study
Sampling Procedures

This paper describes the procedures used to develop the 2014 Utah Residential Distributed Generation Load Study. This study will provide load data for use in support of cost studies and price filings before the Utah Public Service Commission, and for use in studies of customer demand characteristics. The goal of this sample design is to provide relative precision of ± 10% at the 95% confidence level for an estimate of demand at the time of the monthly system peak hours.

Recorders will be placed in service effective no later than August 1st, 2014, and will be monitored on a continuous basis to insure no significant deviation from billing records.

Sampling Plan for Utah

This sampling plan includes several steps:

1. Formalization of the sample parameters;
2. Specification of the target variable;
3. Choice of the stratification variable;
4. Choice of method for estimating kW;
5. Choice of the number of strata;
6. Construction of the strata boundaries
7. Allocation of sample points to each stratum;
8. Selection of primary sample sites;
9. Selection of alternate sample sites.

Formalization of the sample parameters

This is a new load study, designed to provide estimates of load characteristics for the residential distributed generation population in Utah. Input data to be utilized in this design includes billing data for the period June 2013 through March 2014. The design will be based on a stratified random, single-dimensional sampling schema.

In this approach, customers with similar characteristics are grouped together into non-overlapping, homogeneous groups called "strata," and individual samples are selected from each stratum.

The strata are defined according to a user-specified demographic or usage variable called the "design variable." For continuous variables such as usage, the Dalenius-Hodges rule is used to define the strata boundaries. The Neyman allocation procedure is used to determine the optimum sample size for each stratum. (In Neyman allocation, the sample size for each stratum is determined according to its population proportion and the standard deviation. Data from prior load research studies, if available, may also be used to determine the mean and the standard deviation.) A simple random sample is then selected from each stratum.

Because customer-to-customer variation is the basic determinant of sample size (the more the variation, the larger the sample), fewer sampling units need to be selected from a population that has been stratified into homogeneous groups than if the units were merely selected from the entire population at random. In other words, because the variation within a stratum is less than for the entire population, fewer sample points are required to obtain the same accuracy level.

Stratification is a good choice when you need to economize with a smaller sample size, yet maintain a specified level of accuracy. It is also useful when you need data for specific demographic sets within the population (types of business, location, etc.). However, stratification has some aspects which may make it inappropriate for certain situations, i.e., since not all customers have the same chance of being selected, the sample may not be as flexible. Therefore, if you wish to use the sample to perform analyses and answer questions not anticipated in the original design, you may have to employ Domains Analysis to ensure that original sample weights are taken into consideration.

Also, over time, some customers will change their characteristics and will migrate out of their strata. However, the strata assignments must remain fixed throughout the analysis period. For that reason, samples must be replaced periodically to keep them up to date.

Specification of the target variable

Load studies in the state of Utah are used primarily to support cost allocation studies. For this current study, a sample design was prepared based on average delivered customer energy (billed kWh) over a designated 10 month period.

Bill frequency counts, by usage level, are summarized into standardized usage blocks to identify the ideal monthly breakpoints for the design. Utilizing the process defined by Dalenius-Hodges, these breakpoints are then averaged into strata to facilitate further analysis (see Table A).

Billing data for the 10 months ending March 2014 were used to determine appropriate stratification.

Choice of the Stratification Variable

A potential stratifying variable, according to Cochran, should meet four criteria[1]:

1. *The population is composed of institutions varying widely in size.*
2. *The principle variables to be measured are closely related to the sizes of the institutions.*
3. *A good measure of size is available for setting up the strata.*

Average monthly billing kWh (KWH_MNTH), which is the average monthly energy registered over a given consecutive month period, was selected as the best available variable for this purpose. As reporting of monthly customer usage for this group presents the netted amount (delivery to the customer – delivery from the customer), it does presents issues not normally dealt with. Customer usage may be reflected as a negative value for instance. Or the much more likely scenario in which the usage delivered is understated because of power delivered back to the Company. This will make validation of sample results difficult. Nonetheless, the variable is readily available for all customers in this class, with a range from -1,058 to 16,008 kWh for any given customer in this group.

Choice of Method for Estimating kW

To estimate a peak demand for a population using MPU, the mean peak demand value from the sample is multiplied by the number of elements in the entire population. Use of the MPU method provides an unbiased estimate.

For ratio estimation, the ratio of the target variable over the auxiliary variable is calculated for the sample. This ratio is then multiplied by the total annual billed kWh for the population to get the estimated total group peak demand. Because energy usage and peak demand are correlated, a ratio estimate will have a smaller variance than a MPU estimate. However, a ratio estimate may be slightly biased.

With stratified sample designs, ratio estimators can be computed in two ways: separately for each stratum, or a combined ratio can be computed over all strata. Separate ratio estimation tends to result in smaller variance. However, the combined ratio method is more appropriate when stratum sample sizes are small, because the risk of bias is reduced.

Table B details the sample size required for the Utah Residential Distributed Generation Load Study using a mean-per-unit method, assuming a four strata design, with modified allocation utilizing the Tschprow/Neyman method.

---

[1] William G. Cochran, "Sampling Techniques", Third Edition, Wiley, pg.101

Choice of the Number of Strata

As the number of strata increases, precision of the estimate of the total contribution to demand (kW) at system peak also increases. However, the increase in precision per additional stratum diminishes after a relatively small number of strata[2]. Desire for simplicity and a reasonable number of sites in each stratum lead to a preference for a small number of strata. If a minimum number of sites policy is followed (eg.10 sites minimum per stratum), then the addition of strata can actually lead to more, rather than fewer, total sites. If such a policy is not followed, the result can be strata with so few recorders that confidence in sample estimates is at risk from unexpected data problems, variance estimates may not be sufficiently precise for future sample design purposes, and the sample may not be robust enough to be useful when analysis needs change.

A final decision on the number of strata requires actual cost comparison of potential stratification schemes to evaluate effectiveness versus cost. For this study, a four strata scheme was employed. The method described below was used to compare stratification approaches.

Construction of Strata Boundaries

Various methods might be used for definition of strata boundaries. Cochran found the "cumulative square root of f"[3] rule, as defined by Dalenius and Hodges (1959), to be superior in a comparative study of such methods applied to actual distributions exhibiting a range of skewness.

With the Dalenius-Hodges procedure, the program divides the population in the Frequency Distribution File into short intervals. Each interval has frequency f and interval length u. The quantity $\sqrt{uf}$ is summed over all the intervals, and this cumulative $\sqrt{uf}$ is divided by a user-defined number of strata to give the optimum length of each stratum.

Steps in calculating strata boundaries under the "cumulative $\sqrt{uf}$" rule are as follows. First, tabulate frequencies of the stratifying variable. For these studies, average monthly energy (KWH_MNTH) from customer billing records for the ten months ending March, 2014 were used. All Utah Residential DG customers, whose month end status was active, were included in this procedure, and in population figures for the sample design. Second, multiply the number of customers in each interval by the interval factor. Third, take the square root of these frequencies. Fourth, cumulatively sum the square roots. The resulting distribution of adjusted cumulative square roots of frequency is then partitioned into equal intervals by dividing by the number of strata. The final stratification scheme of four strata is presented in Exhibit 1, and shows the optimal boundaries resulting from the above procedure, after adjustments made to accommodate prior cost analysis requirements (if any).

---

[2] William G. Cochran, "Sampling Techniques", Third Edition, Wiley, Pg. 132
[3] William G. Cochran, "Sampling Techniques", Third Edition, Wiley, Pgs. 129-130

Allocation of Sample Points to Each Stratum

Once the stratum boundaries have been determined, sample points (i.e., load recorders) must be assigned to the strata. The Tschprow-Neyman allocation procedure[4] allocates an optimal sampling rate to each stratum. Optimal allocation techniques minimize the variance of the population estimates by increasing the sample proportion in the strata having larger variances. This produces a sampling rate for each stratum which is proportional to the standard deviation within the stratum. The analogous procedure for a ratio sampling plan is allocation in proportion to the square root of the residual variance.

Average billing energy was selected as both the target and stratification variables. These data were used to provide estimates for the new Utah Residential DG sample design. For the mean-per-unit method, the variance within each stratum was the ordinary variance of the mean.

Minimum recorder allocations and data loss adjustments are required for each stratum to maintain adequate data in case of recorder failure and to provide data for analysis of load characteristics other than the primary target variable, should such analysis be necessary. Minimums ranging from 5 to 15 sites per stratum have been used in past studies. In the present studies, a minimum of 10 sites was used. A minimum on the high side was selected, despite improvements in data quality due to solid state recording equipment, because changing requirements for load research and other areas using this data may require unanticipated applications, and because overall sample efficiencies are bringing these studies in well below the budgeted number of sites, even with the 10 site minimum. The final allocation of recorders reflected an additional ten percent data loss adjustment per stratum over the optimal or minimum allocation.

Budget approval was received which allowed us to install 62 network meters for this study. An analysis of customers selected to participate in this load study indicates that 0 sites currently have load profile metering installed. The four strata design selected calls for the installation of 45 recorders to meet design standards. We supplemented this amount to reflect total installations of 62 meters. This design selected should achieve ± 10% Relative Accuracy at the 95% Confidence Level on estimates of the target variable.

Sample Selection

Systematic sample selections were used for each stratum to ensure a representative distribution. For practical reasons, inactive customers and customers with no kWh meter installed (usually certain types of lighting customers with very predictable demand and consumption, indicated by absence of a kWh meter number) were eliminated from the sampling frame. Eligible customers were then sorted by stratum and by average monthly billed energy (KWH_MNTH) within stratum. The number of customers available in the sampling frame for each stratum was then divided by the number of recorders allocated to that stratum (N_h/n_h), yielding the sampling interval size. A five digit random number

[4] William G. Cochran, "Sampling Techniques", Third Edition, Wiley, pgs. 96-99

between 0 and 1 was chosen for each stratum, and multiplied by the stratum interval size to obtain the starting selection point for each stratum (Table C). Beginning with this site, additional sites were selected at the given sampling intervals to obtain the desired number of sample sites. This procedure was repeated four times to provide a list of alternate selection sites.

The list of primary and alternate selection sites for this sample are contained in Appendix 1. This list was compared against current Utah profile metering installations to check for duplicates. Duplicates between the design and production systems were noted and updated in the Appendix.

Tables

Utah Residential DG DH Worksheet
Four Strata

Range			Customer Count f	Interval Factor μ	μf	$\sqrt{\mu f}$	cum $\sqrt{\mu f}$	
0	to	50	107	1	107	10.3	10.3	
50	to	100	61	1	61	7.8	18.2	
100	to	150	78	1	78	8.8	27.0	
150	to	200	107	1	107	10.3	37.3	
200	to	250	95	1	95	9.7	47.1	
250	to	300	133	1	133	11.5	58.6	
300	to	350	93	1	93	9.6	68.3	
350	to	400	87	1	87	9.3	77.6	761
400	to	450	96	1	96	9.8	87.4	
450	to	500	80	1	80	8.9	96.3	
500	to	600	129	2	258	16.1	112.4	
600	to	700	88	2	176	13.3	125.7	
700	to	800	71	2	142	11.9	137.6	
800	to	900	63	2	126	11.2	148.8	527
900	to	1000	55	2	110	10.5	159.3	
1000	to	1100	44	2	88	9.4	168.7	
1100	to	1200	32	2	64	8.0	176.7	
1200	to	1300	21	2	42	6.5	183.1	
1300	to	1400	24	2	48	6.9	190.1	
1400	to	1500	17	2	34	5.8	195.9	
1500	to	1750	32	5	160	12.6	208.6	
1750	to	2000	11	5	55	7.4	216.0	236
2000	to	2250	12	5	60	7.7	223.7	
2250	to	2500	13	5	65	8.1	231.8	
2500	to	2750	5	5	25	5.0	236.8	
2750	to	3000	7	5	35	5.9	242.7	
3000	to	3250	2	5	10	3.2	245.9	
3250	to	3500	3	5	15	3.9	249.7	
3500	to	3750	2	5	10	3.2	252.9	
3750	to	4250	1	10	10	3.2	256.1	
4250	to	4500	1	5	5	2.2	258.3	
4500	to	5000	1	10	10	3.2	261.4	
5000	to	5500	2	10	20	4.5	265.9	
5500	to	6000	2	10	20	4.5	270.4	
6000	to	7500	1	30	30	5.5	275.9	
7500	to	9000	1	30	30	5.5	281.3	
9000	to	15000	1	120	120	11.0	292.3	54
Total N			1,578					1,578

BOUNDARIES INDICATED FOR STRATA:

	3	4	5	6
1	97.4	73.1	58.5	48.7
2	194.9	146.2	116.9	97.4
3		219.2	175.4	146.2
4			233.8	194.9
5				243.6

SAMPLING ST/	Avg. kWh[1]	Mean kW[2]	St. Dev [1]
1	204.1		115.8
2	594.3		141.8
3	1,229.5		266.5
4	3,317.1		2,078.2
5			
6			

1 Billing records for April 2013 through March 2014

Table A

UTAH RESIDENTIAL DG LOAD STUDY DESIGN OPTION (2014)
FOUR STRATA, MEAN-PER-UNIT DESIGN

		a	b	c	d	e	f	g	h	i	j
		Sample Mean kW	Sample Mean kWh	Pop N	Variance of Mean	Standard Deviation	Wtd. Devtns. c*e	Proprtn. row f/ sum f	Optimal Allocation g*h total	Optimal with Attrition	Final with Attrition
STRATUM 1	0 -400 kWh		204.1	761	13409.6400	115.800	88124	0.2607	12	12	15
STRATUM 2	401 - 900 kWh		594.3	527	20107.2400	141.800	74729	0.2211	10	10	14
STRATUM 3	901 - 2,000 kWh		1,229.5	236	71022.2500	266.500	62894	0.1861	8	10	12
STRATUM 4	GT 2,000 kWh		3,317.1	54	4318915.2400	2078.200	112223	0.3321	15	15	21
EST POP MEAN (wtd by N)		0.000	594.298	1,578			337969	1.0000	45	47	62
										Sample Estimate	**Adj Sample Estimate**
										45	62

RELATIVE PRECISION OF SAMPLE KW ESTIMATE

		TOTAL KW Optimal n (col. h)	TOTAL KW Adjusted n (co	TOTAL KW Final (col. J	MEAN KW Adj. n
Variance	1	694,849,754	694,849,754	543,766,649	279.046712
contributed	2	608,710,945	608,710,945	418,154,803	244.453980
by strata:	3	545,937,890	420,893,636	341,320,020	169.027887
	4	649,688,250	649,688,250	384,815,348	260.910174
Total Variance		2,499,186,838	###########	##########	953.438753
Standard Error		49991.86772	48725.17403	41085.9687	30.87780357
Desired Conf. Level		95%	95%	95%	95%
(z two tailed)		1.96	1.96	1.96	1.96
Conf. Interval		97984.06073	95501.34111	80528.4986	60.520495
MPU Est of kW		0	0	0	0.0000
Relative Conf. Int.		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!

Table B

Utah Residential DG Sample Selection Parameters

Active Customers with kWh Meters
For the 12 Months Ending March 2014

Stratum	1	2	3	4	5	6
Sampling Frame	761	527	236	54		
Sample	15	14	12	21		
Interval	50.73	37.64	19.67	2.57		
Random Starts						
Primary						
Random No.(1)	0.28885	0.60446	0.93179	0.74182		
Start	15	23	18	2		
Alternate 1						
Random No.(1)	0.00035	0.14860	0.01623	0.47069		
Start	1	6	1	1		
Alternate 2						
Random No.(1)	0.62603	0.02792	0.35359	0.75281		
Start	32	1	7	2		
Alternate 3						
Random No.(1)	0.21875	0.83566	0.79521	0.20098		
Start	11	31	16	1		
Alternate 4						
Random No.(1)	0.89793	0.64829	0.61813	0.14148		
Start	46	24	12	1		

(1) Random numbers from Excel's random function.

Table C