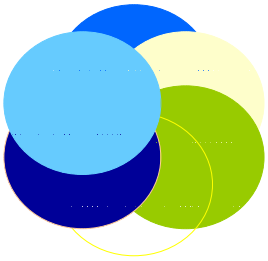
# Multi-Dimensional Sample Design

# Applied to the Small General Service Class of Customers



**Northeast Utilities Service Company**
**Load Research Department**

**Deborah J. Olisky**
**Senior Load Research Analyst**

**Presented to AEIC Load Research Workshop**
**Orlando, Florida**
**April 12, 2005**

Northeast Utilities Service Company (NU) is in the processing of implementing a new load research sample for the small general service class of customers in two jurisdictions of the company, Connecticut Light and Power Company (CL&P) and Western Massachusetts Electric Company (WMECO). CL&P is the largest electric utility in Connecticut serving nearly 1.2 million residential and commercial customers in 149 cities and towns. Western Massachusetts Electric Company serves 200,000 customers in 59 communities in western Massachusetts.

Of the two million customers that NU serves, approximately 193,000 are commercial and industrial (C&I) customers. The largest customers whose demand is over 1,000 KW have interval load data recording meters. There is a current initiative in progress to install recording meters on all C&I customers with a demand greater than 350 KW. The multi-dimensional sample design described throughout this paper is targeted at the population of the small general service, commercial and industrial customers, whose demand is less than 350 KW.

The advantage of a multi-dimensional sample design is that a population can be stratified by two or more variables. When variables, such as energy and demand are stratified properly, this technique categorizes the load shapes tendencies within stratum. With a single stratification variable, the degree of classification may not be possible. The disadvantage is that it can add more sample points than the traditional, simple random sample design. The data collection, implementation and maintenance can be more difficult when obtaining the within-stratum variance and precision levels. Therefore the costs may increase depending on the criteria and objectives selected.

At NU, the Load Research Department understands the importance of establishing clear objectives before designing the sample. Much effort is made to communicate with other groups and individuals to clearly define and identify their needs and uses of the load research data. From experience, this time spent

planning can save the expense of a second sample in the future. Discussions were held with the Rates Department and the Retail Access Planning and Support Department. The Rates Department uses load data for the cost of service studies and rate design, and the retail planners use the load profiles in their daily settlement process with the ISO.

The Load Research Department established objectives to meet the requirements of the individual groups. The sample design incorporates acceptable levels of accuracy and precision by rate class for regulatory and cost of service purposes and by class of customers for retail access purposes.

The cost of service studies uses a methodology based on 12 coincident peaks, system peak, class peak and the twelve monthly peaks. These parameters must be in the sample design. For retail market settlement, it was important for the sample to emphasize the load shape characteristics of the population not just examine the peak days and the criteria for rate making.

Like most commissions, our state regulators still use the PURPA standards established in 1978, even though the strict federal standards ended in 1992. The criteria is different for each jurisdiction. CL&P has an overall accuracy of +/- 10% at a 90% confidence level at the system peak and WMECO has a higher accuracy of +/- 5% at a 95% confidence level.

| | Jurisdiction | |
|---|---|---|
| Class | CL&P | WMECO |
| Small (0-100 KW) | Rate 30 | Rate G0 |
| Intermediate (>100-350KW) | Rate 35 | Rate G2 |
| Churches Schools | Rate 40 | - |

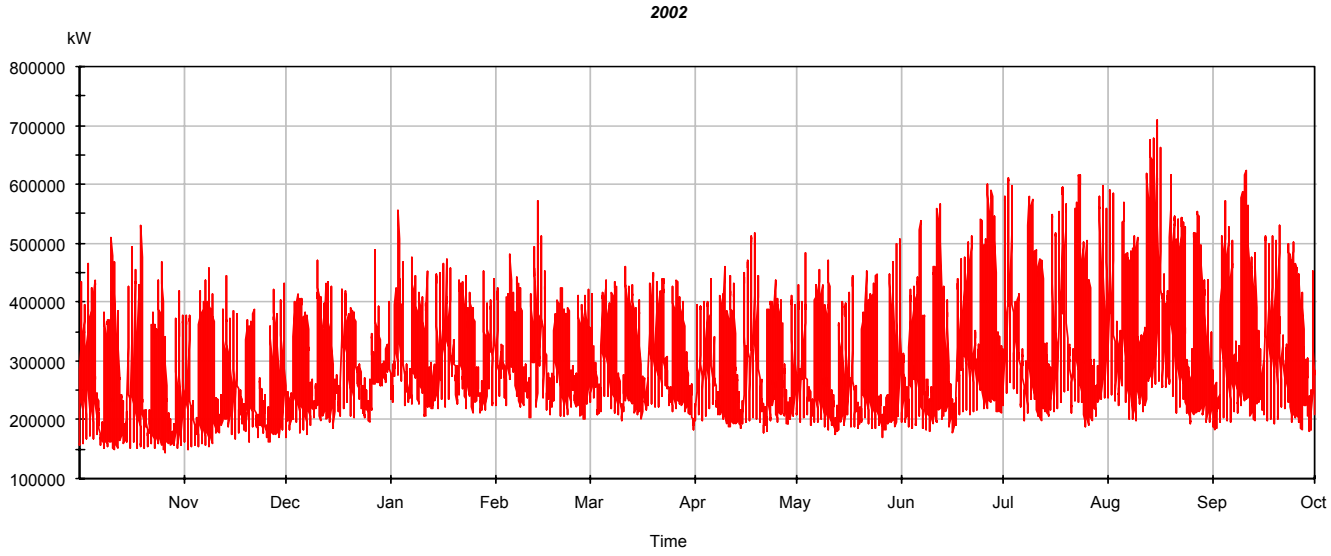
With the objectives established, historical interval load data and current billing data for the population is used to design the new sample. The process uses

Model Based Sample Design (MBSS®) to obtain the most efficient, optimal sample size and to determine the number of stratum and the stratum boundaries to achieve the desired level of accuracy within stratum and on an overall basis.

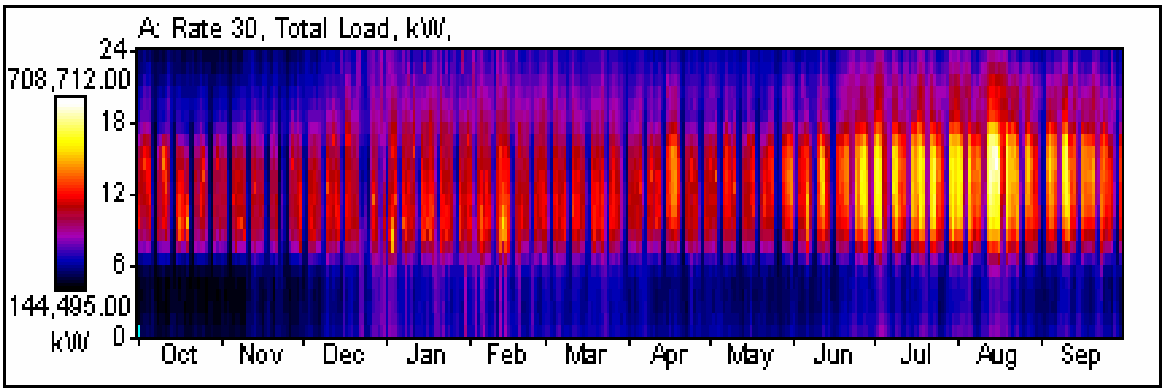
A sample for each rate class is developed and run through the MBSS® model using the ratio estimation techniques for sample design and analysis. Using the Visualize_IT® software developed by RLW Analytics, each customer load profile is plotted on a graph and a decision is made whether or not to include them in the sample population. This sample data along with the current billing data is then used in MBSS® to give a load shape fore each rate class. The parameters within the rate class are analyzed for precision levels, gammas, error ratios, and error bounds. This analysis is repeated using different grouping of the rate classes by jurisdiction and by demand and usage levels.

Preliminary analysis of CL&P Rate 30 are in the following tables:

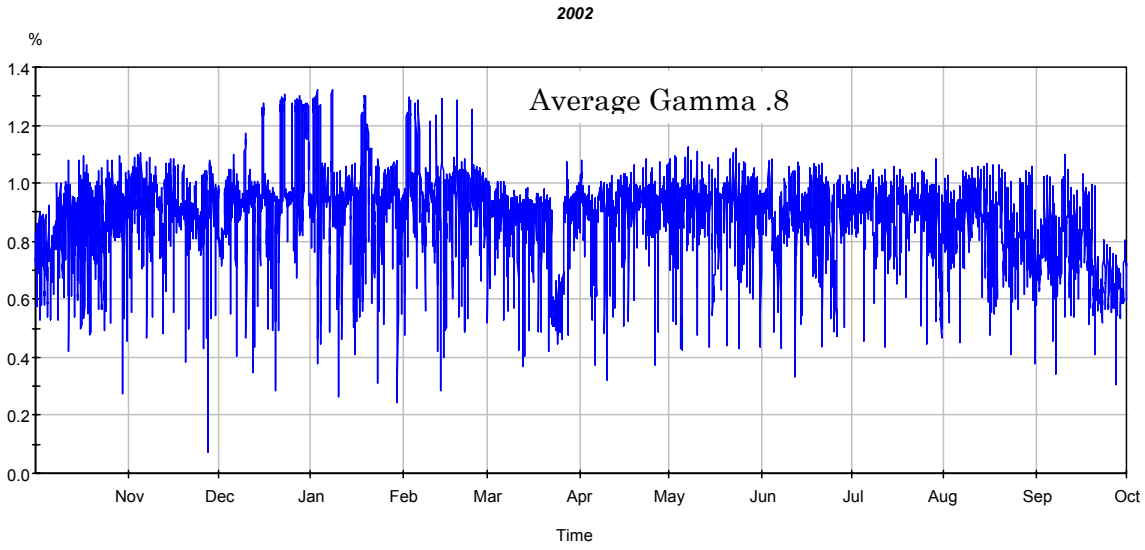
Total Load of Rate 30 Class – Expanding sample data to the current

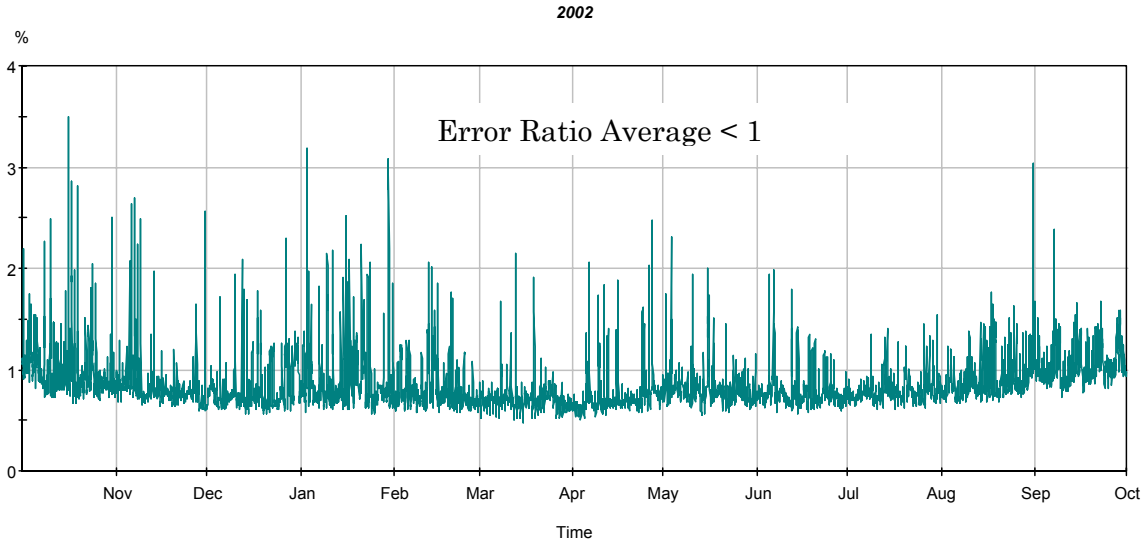


Total Load of Rate 30 Class – Energy Print

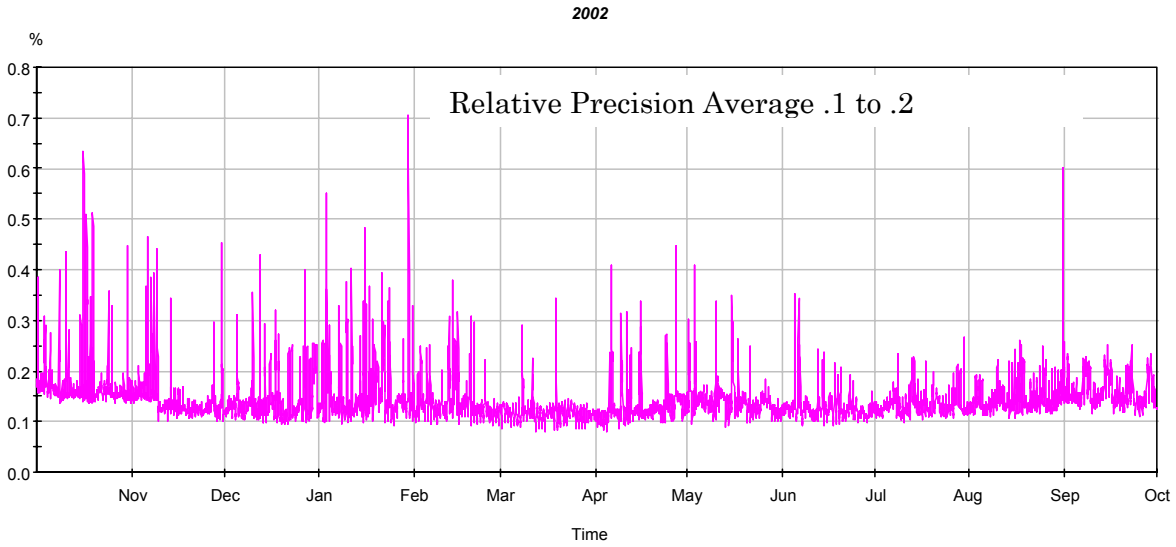



The parameters of the Rate 30 are the following:



The parameters of the Rate 30 are the following:




After the statistics were known for each class, this information was used to begin the new sample designs using the MBSS® model. For each rate class the optimal number of stratum and the stratum boundaries needed to be found for the dependant variables of demand and usage. To begin, the traditional five strata were chosen to get the breakpoints. By examining the summary and descriptive statistical results from the model, a determination was made on the stratum boundaries and the ranges within each boundary. This became an iterative process until satisfied with the statistical results and the number of strata.

Then the model was used to find the most appropriate number of sample points for the desired precision levels for the overall sample both at the time of the peak and 12 coincident peaks. The dependant variables are demand and energy and the gamma of .8 was used based on the historical data. For a simple random sample, the model calculates the number of sample points based on the equation:

$$n = \left[ \frac{Z * CV}{RP} \right]^2$$

Applying this to a simple random sample with the Rate 30 Class:
n = number of sample points

Z statistic depending on the precision level selected = 1.645

CV = error ratio = .6598 for Rate 30

RP= Relative Precision is 10%

$$n = \left[ \frac{1.645 * .6598}{.10} \right]^2$$

n = 118

Using a more sophisticated design having five strata based on annual kilowatt hours, the model calculates 170 sample points to obtain the 90%, +/-10% precision level with 34 sample points per strata.

NU's sample design has two way stratification, three strata for demand and five strata for usage. The model allocates 200 sample points and assigns them with 75 meters for the low demand strata, 50 meters for the medium demand strata, 75 meters for the highest demand strata. The loss factor of 15% is then added in, increasing the total to 230 meters (30 additional meters added in for loss). This can be seen in the following table:

| Two Dimensional Strata Design | **Rate 30 (Tests B1, B2, & B3)** | | | | |
|---|---|---|---|---|---|
| 3 Demand Strata 5 Usage Strata | Usage Strata | Cut-Points Maximum Annual Use (kwh) | Population Count (N) | Required Sample Size (n) | Case Weights (N/n) |
| 1- Demand < 25 KW (Test B1) | 1 | 11,966 | 38,981 | 17 | 2293.00 |
| | 2 | 21,530 | 14,337 | 17 | 843.35 |
| | 3 | 34,299 | 9,424 | 17 | 554.35 |
| | 4 | 52,893 | 6,654 | 17 | 391.41 |
| | 5 | 300,000 | 4,377 | 17 | 257.47 |
| | Totals | | 73,773 | 85 | |
| | Usage Strata | Cut-Points Maximum Annual Use (kwh) | Population Count (N) | Required Sample Size (n) | Case Weights (N/n) |
| 2- Demand ≤ 25 KW < 50 KW (Test B2) | 1 | 61,656 | 3,577 | 12 | 298.08 |
| | 2 | 84,036 | 2,231 | 12 | 185.92 |
| | 3 | 107,371 | 1,798 | 12 | 149.83 |
| | 4 | 139,674 | 1,478 | 12 | 123.17 |
| | 5 | 500,000 | 1,142 | 12 | 95.17 |
| | Totals | | 10,226 | 60 | |
| | Usage Strata | Cut-Points Maximum Annual Use (kwh) | Population Count (N) | Required Sample Size (n) | Case Weights (N/n) |
| 3- Demand ≥ 50 KW (Test B3) | 1 | 112,237 | 1,177 | 17 | 69.24 |
| | 2 | 147,750 | 716 | 17 | 42.12 |
| | 3 | 182,548 | 594 | 17 | 34.94 |
| | 4 | 235,461 | 497 | 17 | 29.24 |
| | 5 | 2,000,000 | 329 | 17 | 19.35 |
| | Totals | | 3,313 | 85 | |
| Total Population | | | 87,312 | | |
| Total Sample Size Required | | | 230 | | |

| **Relative Precision** | | | |
|---|---|---|---|
| ( Based on 200 sample points) | 12 Cp | System Pk | Demand Strata |
| Overall Relative Precision | 4.8 | 8.6 | Strata 1 precision |
| Relative Precision -Test B1 | 8.2 | 13.7 | Strata 2 precision |
| Relative Precision -Test B2 | 7.8 | 15.1 | Strata 3 precision |
| Relative Precision -Test B3 | 5.8 | 12.0 | |

Rate 35 has a design with two demand strata and five usage strata. The design gives the desired overall precision level and within-strata precision level.

| Two Dimensional Strata Design | | Rate 35 (Tests C1 & C2) | | | |
|---|---|---|---|---|---|
| 2 Demand Strata 5 Usage Strata | Usage Strat | Cut-Points Maximum Annual Use (kwh) | Population Count (N) | Required Sample Size (n) | Case Weights (N/n) |
| 1- Demand less than 108 KW (Test C1) | 1 | 204,035 | 992 | 17 | 58.35 |
| | 2 | 246,559 | 678 | 17 | 39.88 |
| | 3 | 294,404 | 591 | 17 | 34.76 |
| | 4 | 362,399 | 509 | 17 | 29.94 |
| | 5 | 1,000,000 | 402 | 17 | 23.65 |
| | Totals | | 3,172 | 85 | |
| | Usage Strat | Cut-Points Maximum Annual Use (kwh) | Population Count (N) | Required Sample Size (n) | Case Weights (N/n) |
| 2- Demand ≥ 108 KW (Test C2) | 1 | 448,565 | 1,091 | 17 | 64.18 |
| | 2 | 612,734 | 770 | 17 | 45.29 |
| | 3 | 836,920 | 602 | 17 | 35.41 |
| | 4 | 1,100,731 | 477 | 17 | 28.06 |
| | 5 | 40,000,000 | 348 | 17 | 20.47 |
| | Totals | | 3,288 | 85 | |
| Total Population | | | 6,460 | | |
| Total Sample Size Required | | | 170 | | |
| **Relative Precision** | | | | | |
| ( Based on 150 sample points) | 12 Cp | System Pk | | | |
| Overall Relative Precision | 3.4 | 7.6 | | | |
| Relative Precision -Test C1 | 5.5 | 13.6 | | | |
| Relative Precision -Test C2 | 4.1 | 9.0 | | | |

The churches and schools rate used a simple, one dimensional design with five strata based on kilowatt hour usage. The two dimensional design was not a good choice because of the demands and usage for the churches was much different than the schools. The precision levels did not increase using a two way stratification design because the characteristics were so unique to each group.

**Sample Design Document**

| One Dimensional Strata Design | Rate 40 (Test D) Church and School Rate | | | | |
|---|---|---|---|---|---|
| 5 Usage Strata | Usage Strata | Cut-Points Maximum Annual Use (kwh) | Population Count (N) | Required Sample Size (n) | Case Weights (N/n) |
| | 1 | 45,220 | 1,556 | 20 | 77.80 |
| | 2 | 112,847 | 466 | 20 | 23.30 |
| | 3 | 212,399 | 244 | 20 | 12.20 |
| | 4 | 324,907 | 162 | 20 | 8.10 |
| | 5 | 2,000,000 | 104 | 20 | 5.20 |
| | Totals | | 2,532 | 100 | |
| Total Population | | | 2,532 | | |
| Total Sample Size Required | | | 100 | | |
| **Relative Precision** | | | | | |
| ( Based on 90 sample points) | 12 Cp | System Pk | | | |
| Overall Relative Precision (Test D) | 4.6 | 14.6 | | | |

Before implementing the roll out of the new meters in the jurisdictions, it is essential to educate and clearly communicate with the Meter and Services Department. As a cost saving measure for the company, a conscious decision was made to keep the existing telephone AMR recording meters in the field, and they were to become part of the sample.

There are some installation problems due to the nature and movement within this general service class. New businesses move in and out before an installation can be completed. The substitutes have to be selected by an analyst to insure the same demand and usage strata is selected for a customer that can not be installed. The technology of the new Nertec meters uses the customer's exiting telephone lines which has added some delay in installation. The technicians found many telephone lines where the line and meter were too far apart and costly to install. There was not much rejection of participation on the part of the customers.

The following table show the meters required for each jurisdiction and each rate class:

| Rate | Meters Required | Existing AMRs | New Installations |
|---|---|---|---|
| Rate 30 | 230 | 48 | 185 |
| Rate 35 | 170 | 134 | 68 |
| Rate 40 | 100 | 4 | 96 |
| CLP Subtotal | 500 | 186 | 349 |
| | | | |
| Rate G0 | 230 | 12 | 218 |
| Rate G2 | 170 | 31 | 139 |
| WMECO Subtotal | 400 | 43 | 357 |
| Total | 900 | 229 | 706 |

There are some existing AMR meters installed in Rate 30 and Rate 35 that cause an oversmpling of the design requirements.

In essence then, NU is relying on the mult-dimensional stratified design to give the statistical results expected. This design meets all the objectives of the Load Research Department and our clients who count on us for their data needs.