Premium Power Grades

Developing Site Selection Criteria for DPQ Phase II

Technical Report

Premium Power Grades

Developing Site Selection Criteria for DPQ Phase II

1005921

Final Report, February 2002

EPRI Project Manager
A. Sundaram

EPRI • 3412 Hillview Avenue, Palo Alto, California 94304 • PO Box 10412, Palo Alto, California 94303 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com

DISCLAIMER OF WARRANTIES AND LIMITATION OF LIABILITIES

THIS DOCUMENT WAS PREPARED BY THE ORGANIZATION(S) NAMED BELOW AS AN ACCOUNT OF WORK SPONSORED OR COSPONSORED BY THE ELECTRIC POWER RESEARCH INSTITUTE, INC. (EPRI). NEITHER EPRI, ANY MEMBER OF EPRI, ANY COSPONSOR, THE ORGANIZATION(S) BELOW, NOR ANY PERSON ACTING ON BEHALF OF ANY OF THEM:

(A) MAKES ANY WARRANTY OR REPRESENTATION WHATSOEVER, EXPRESS OR IMPLIED, (I) WITH RESPECT TO THE USE OF ANY INFORMATION, APPARATUS, METHOD, PROCESS, OR SIMILAR ITEM DISCLOSED IN THIS DOCUMENT, INCLUDING MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, OR (II) THAT SUCH USE DOES NOT INFRINGE ON OR INTERFERE WITH PRIVATELY OWNED RIGHTS, INCLUDING ANY PARTY'S INTELLECTUAL PROPERTY, OR (III) THAT THIS DOCUMENT IS SUITABLE TO ANY PARTICULAR USER'S CIRCUMSTANCE; OR

(B) ASSUMES RESPONSIBILITY FOR ANY DAMAGES OR OTHER LIABILITY WHATSOEVER (INCLUDING ANY CONSEQUENTIAL DAMAGES, EVEN IF EPRI OR ANY EPRI REPRESENTATIVE HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES) RESULTING FROM YOUR SELECTION OR USE OF THIS DOCUMENT OR ANY INFORMATION, APPARATUS, METHOD, PROCESS, OR SIMILAR ITEM DISCLOSED IN THIS DOCUMENT.

ORGANIZATION(S) THAT PREPARED THIS DOCUMENT

EPRI PEAC Corporation

ORDERING INFORMATION

Requests for copies of this report should be directed to EPRI Orders and Conferences, 1355 Willow Way, Suite 278, Concord, CA 94520, (800) 313-3774, press 2 or internally x5379, (925) 609-9169, (925) 609-1310 (fax).

CITATIONS

This report was prepared by

EPRI PEAC Corporation
942 Corridor Park Blvd.
Knoxville, Tennessee 37932

Principal Investigator(s)
C. Melhorn
A. Maitra
A. Mansoor

This report describes research sponsored by EPRI.

The report is a corporate document that should be cited in the literature in the following manner:

Premium Power Grades: Developing Site Selection Criteria for DPQ Phase II, EPRI, Palo Alto, CA: 2002. 1005921.

REPORT SUMMARY

With the proliferation of embedded processors and other sensitive digital loads, the need for increased understanding of power quality (PQ) event characteristics in electrical power systems has never been greater. Since the completion of the EPRI Distribution Power Quality (DPQ) project in 1995, several utilities have implemented system-wide PQ monitoring programs both in distribution and transmission systems. The wealth of data collected since the DPQ project provides a unique opportunity to synthesize meaningful information regarding variability of grid power—specifically on voltage sag rates—based on system characteristics. It also presents a unique opportunity to compare DPQ project results with data from these monitoring programs. This report develops the framework for conducting a second-round DPQ project, using existing data collected by utilities with permanent power quality monitoring systems.

Background

In the early 1990s, EPRI initiated a project called the Distribution Power Quality (DPQ) project that resulted in power quality monitoring at 277 distribution sites statistically chosen throughout the United States. The project's aim was to gain a better understanding of the frequency and severity of power quality events. Data were collected for a period of more than two years and analyzed to understand the frequency and severity of different types of power quality events.

Objectives

To develop the framework for conducting a second round of DPQ investigations using existing data collected by utilities with permanent power quality monitoring systems.

Approach

To reduce the amount of data required from each utility, the project team first collected a limited amount of data from each utility's monitoring site. These data characterized the type of PQ monitor, the amount of data, and the characteristics of the monitoring sites. This information was used to develop a site selection process. The intent of the site selection process was to establish a statistically valid methodology for sampling a portion of available data that would characterize the population of monitoring sites at different utilities. The project team used the stratified random sampling technique to develop the sampling procedure. Cell boundaries were then identified for each stratification variable using the cumulative root frequency method. Finally, variations in optimal sample sizes were evaluated as a function of desired precision levels.

Results

Out of a total of 1,861 PQ monitors (target population) deployed by the participating utilities, a majority are on the distribution level. Approximately 278 locations were either at transmission

busses or at customer busses fed directly from the transmission system. Data from these monitors will provide an indication of voltage sag rates originating from the transmission system since the exposure for these locations is primarily the transmission system. These data will provide unique insight into voltage sag rates for customers who are fed directly from the transmission system. The amount of data collected by a monitor over a given period typically is represented in monitor-years worth of data. For example, data from 5 monitors for a period of 2 years signifies 10 monitor-years of data. As a reference point, Phase I of the DPQ project collected approximately 900 monitor-years worth of data. The amount of data that represents the population of the DPQ Phase II project is approximately 4,000 monitor-years, which is almost 5 times greater than data collected during the DPQ Phase I effort.

Based on the data of the 12 participating utilities, the total number of PQ monitors (target population) deployed in each system was 1,861. To reduce the amount of data that is initially required from each utility, the overall site selection process used several stages of sampling with the details of sampling unit information increasing at each successive level. 95% confidence and 5% precision levels were used in determining sample sizes (based on stratified random sampling techniques). Voltage class and lightning flash density were used as the primary power quality descriptor during the initial stages of sampling the monitoring sites.

Once the initial sampling stages in the overall site selection process have been completed, or in other words, smaller “statistically valid” monitoring locations across the United States have been accurately identified, more detailed data from the selected sites will have to be requested. The information collected during this stage will be used in the final analysis for establishing baselines of power quality at a national level and for sampling at the next stage.

EPRI Perspective

Understanding the frequency and characteristics of power quality events, which are part of the normal electrical environment, is critical for designing end-use equipment immunity. This information also is important for defining baseline levels of power quality that can be expected by a customer and sets the stage for providing premium power services. The DPQ Phase II effort will result in a knowledgebase that will help realize the goal of enhancing compatibility of sensitive equipment with the electrical environment.

Keywords

Distribution power quality
Power quality monitoring
Statistical sampling
Voltage sag
Electrical environment

ACKNOWLEDGMENTS

The authors would like to thank Marek Waclawiak of United Illuminating Company, Shelton, Connecticut, for his invaluable technical assistance throughout the project.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

BACKGROUND

In the early 1990s, EPRI initiated a project called Distribution Power Quality (DPQ) that resulted in power quality monitoring at 277 distribution sites statistically chosen throughout the United States to gain valuable knowledge regarding the frequency and severity of power quality events. The data collected for a period of more than two years was analyzed to understand the frequency and severity of different types of power quality events. This DPQ Phase I data is the only comprehensive database that is currently available in the world that characterizes the power quality level in distribution systems.

As part of the Premium Power Grades project in 2000, an assessment of the performance range of grid power was characterized using the DPQ Phase I monitoring data to define base power quality levels in distribution systems. Since the completion of the DPQ Phase I in 1995, several utilities have implemented system-wide PQ monitoring programs both in distribution and transmission systems. The wealth of data collected since the DPQ project provides a unique opportunity to synthesize meaningful information regarding variability of grid power, specifically on voltage sag rates, based on system characteristics. It also presents a unique opportunity to compare the results of the DPQ project with the data from these monitoring programs.

The goals of this project were to:

1. Determine which utilities have implemented system-wide PQ monitoring programs.
2. Determine what types of monitoring systems have been implemented and how much data was available.
3. Determine what types of systems (e.g., distribution, transmission, etc.) are being monitored.
4. Determine a statistical approach for selecting monitoring locations from existing sites on the member utility systems.
5. Determine what methods of data collection and aggregation are feasible for collecting monitoring data from large, independent monitoring systems.

To accomplish these goals, a survey (found in Appendix B) was developed and distributed to the funding member utilities. A total of twelve utilities responded to the survey. That survey information is included in this report. A statistical approach for selecting monitoring sites was developed and is presented in this report as well.

Future Work

Future work on this project includes gathering monitoring and system data for sites selected during the site selection process. Analysis of the power quality monitoring data will be performed for voltage variations. This information will be compared to the results of the DPQ project and presented in a technical report.

Report Organization

The objective of this report is to provide an overview of the techniques used to develop a site selection process for the DPQ II project. The various statistical techniques, survey, and ultimately the site selection process are presented in this report.

Chapter 2 of this report presents the results of the survey that was sent to all participating utilities. This survey was used to determine the types of monitoring systems employed by the participating utilities. Information from the survey was used to develop a statistical site selection process.

Chapter 3 of this report addresses the site selection process. More than one method of selection was explored, and each of these methods is presented in this section. This chapter also presents the actual sites selected for each participating utility.

Chapter 4 of this report describes the techniques and methodologies used to analyze voltage variations. This chapter also addresses the need for a consistent methodology used for voltage sag analysis.

2

SURVEY OF UTILITIES

Knowing the number of utilities that have installed large-scale power quality monitoring systems and knowing the number of monitors placed in each system, the project team decided that the amount of information collected from each utility needed to be minimized. In an effort to reduce the amount of data required from each utility, the first step of the site selection process was to collect a limited amount of data from each utility pertaining to their monitoring sites. This information was used during the site selection process to develop the stratification variables. Once the site selection process was complete, more detailed data from the reduced number of selected sites would be requested from the respective utilities.

This report section summarizes the survey and the characteristics of the sites where utilities have installed permanent power quality monitors since the completion of the DPQ project in the early 1990s. The actual survey sent out to the participating utilities may be found in Appendix B.

Causes of Voltage Sags

It was decided early in the project that the main power quality concern for most utilities and their customers was voltage sags. Therefore, the survey was tailored to obtain information from the utilities with regard to voltage sags.

Voltage sags are generally caused by faults. Faults typically occur within the customer facility or on the utility transmission and distribution (T&D) systems. Starting large loads can also cause voltage sags, often lasting for several seconds. However, fault-induced sags are much more severe than sags due to the starting of large loads. To gauge the impact of weather and other external factors on voltage sags, the participating utilities were asked to rank different cause codes according to their relative impact on voltage sags. These cause codes were ranked on a scale of 1 (least severe) to 10 (most severe). This was a subjective assessment since even within a same utility area, causes for voltage sags may differ from one area to another.

Figure 2-1 shows the cause codes that were identified by the surveying utilities as the primary reason for fault-induced voltage sags. Equipment failure, lightning, animal contact, and tree limbs were the most common causes of faults identified by the surveying utilities. For some utilities, individual cause codes such as wind were rated highly, but when averaged with all other utility responses, their relative weight was not as high as the other cause codes.

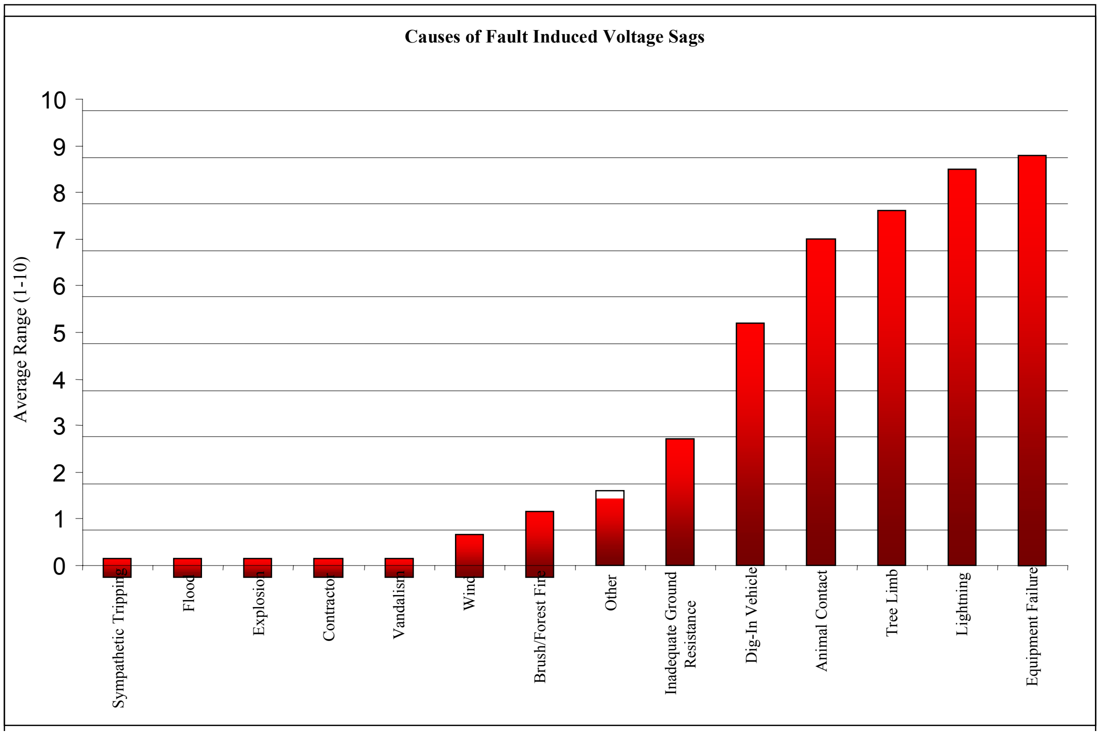


Figure 2-1
Causes of Fault-Induced Voltage Sags

Monitor Locations

Voltage sag rates will be different for transmission systems compared to distribution systems. The voltage sag rates will even be different between distribution systems depending on whether the system is a radial, loop, or network type system. To identify the number of monitors in each location, utilities were asked to provide the number of monitors in each location based on a generic monitor location diagram as shown in Figure 2-2.

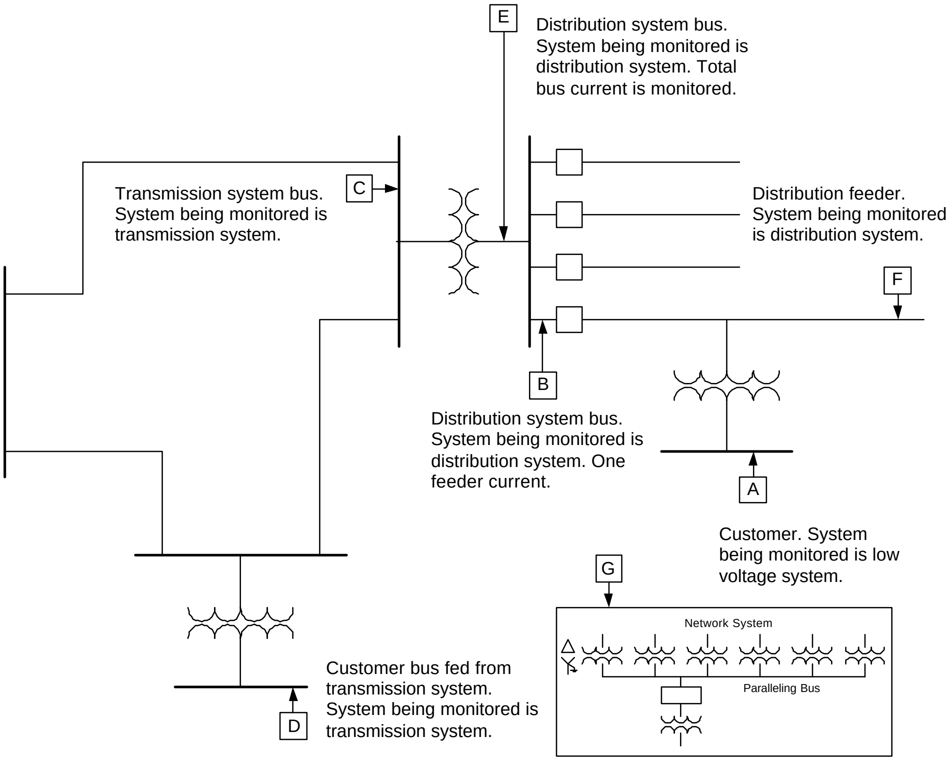


Figure 2-2
Different Generic Categories of Monitor Locations

The different location categories that are identified in Figure 2-2 are:

- **A** – Customer Bus Distribution System
- **B** – Feeder Monitor Distribution System
- **C** – Transmission Bus Transmission System
- **D** – Customer Bus Transmission System
- **E** – Distribution Bus Distribution System
- **F** – Distribution Feeder Distribution System
- **G** – Underground Network
- **H** – Other (explain)

Figure 2-3 shows that, out of the total 1,861 PQ monitors that have been deployed by the participating utilities, a majority of them are on the distribution system. Approximately 278 locations were either at transmission busses or at customer busses fed directly from the transmission system. Data from these monitors will provide an indication of voltage sag rates originating from the transmission system since the exposure for these locations is primarily the transmission system. This will provide some unique insight into voltage sag rates for customers who are directly fed from the transmission system.

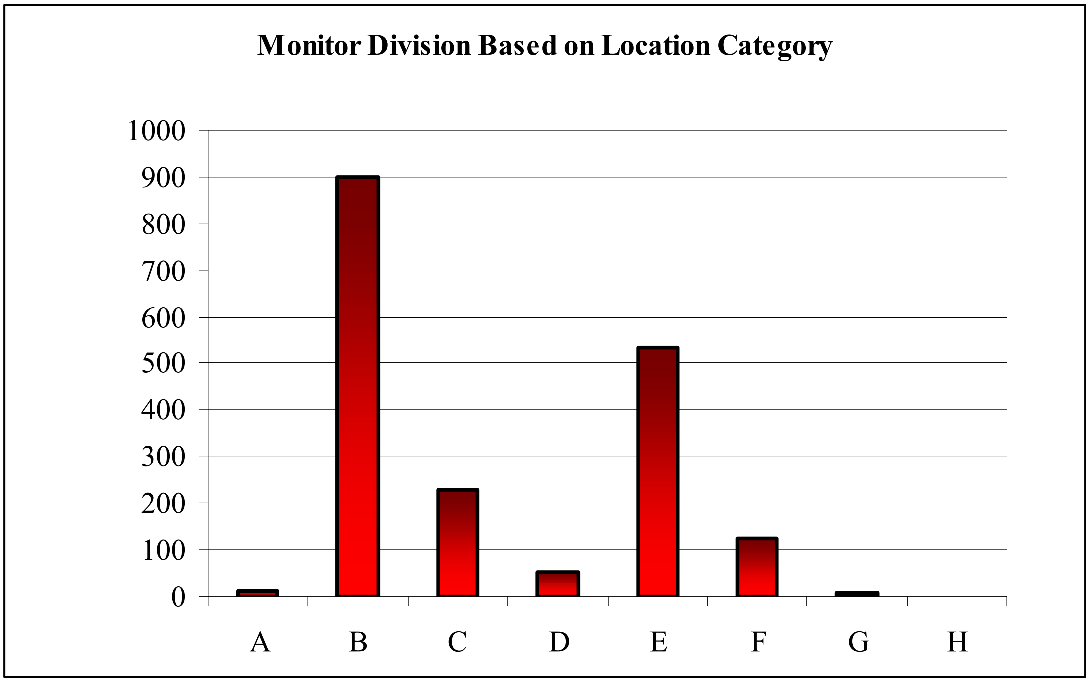


Figure 2-3
Number of PQ Monitors in Different Locations

Figure 2-4 illustrates the breakdown of monitor locations as a percentage of the total number of monitors for transmission, distribution, and secondary network type systems that are common in downtown areas.

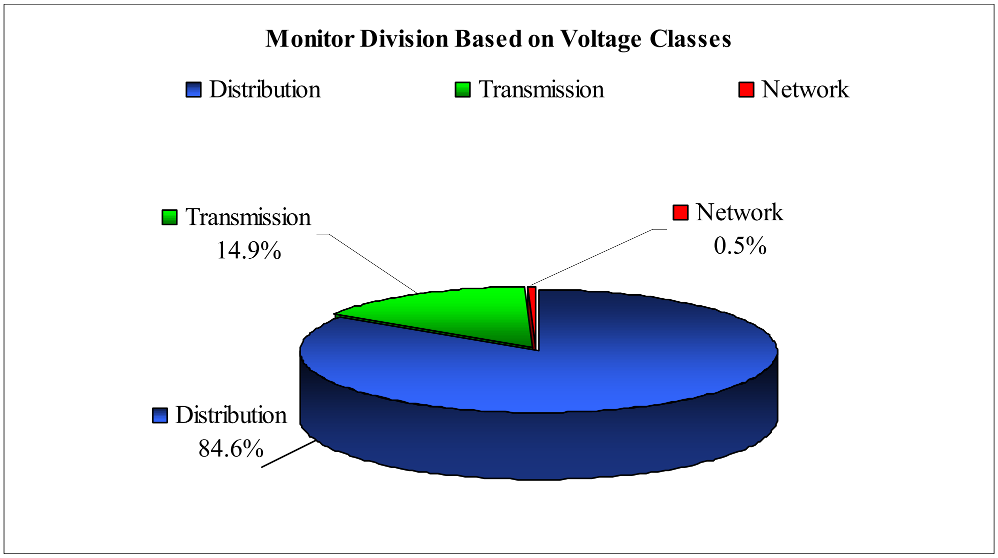


Figure 2-4
Percentages of Monitors in Transmission, Distribution, and Secondary Network Locations

Characteristics of a Distribution System

There are a number of characteristics of a distribution system that may impact the voltage sag rate at a particular site. To evaluate the diversity of distribution system characteristics for the monitor locations, utilities were asked to identify some preliminary characteristics of the distribution system based on the following criteria:

1. Distribution feeder characteristics
   - Overhead/underground/mixed
   - Radial/primary selective/secondary selective/spot
2. Distribution feeder load density
   - Urban
   - Rural
   - Mixed
3. Customer characteristics
   - Mostly residential
   - Mostly commercial
   - Mostly industrial
   - Mixed

Figure 2-5 through Figure 2-8 show the breakdown of the monitor locations based on these characteristics of a distribution system where the monitors are located.

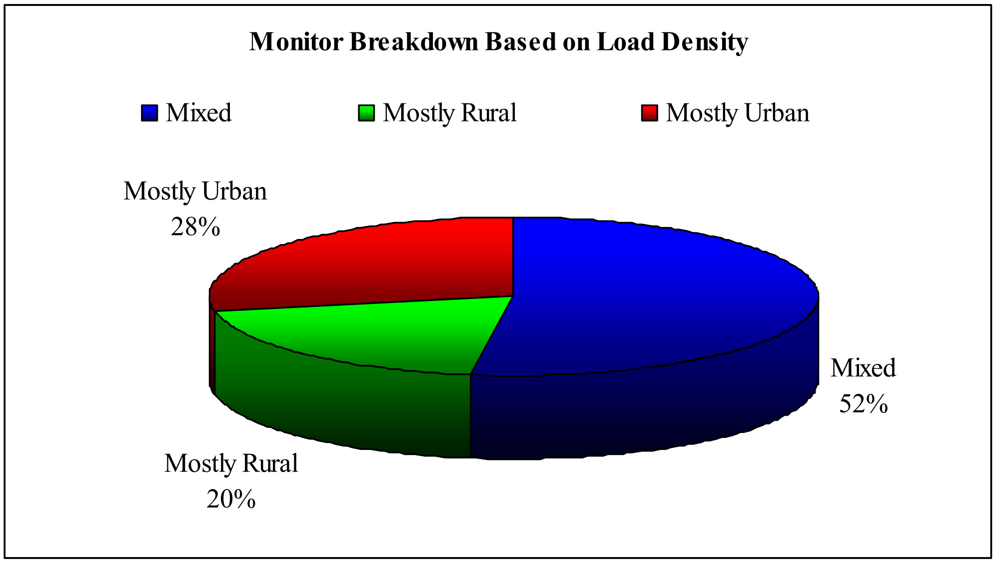


Figure 2-5
Breakdown of Monitor Locations in Distribution Systems Based on Load Density

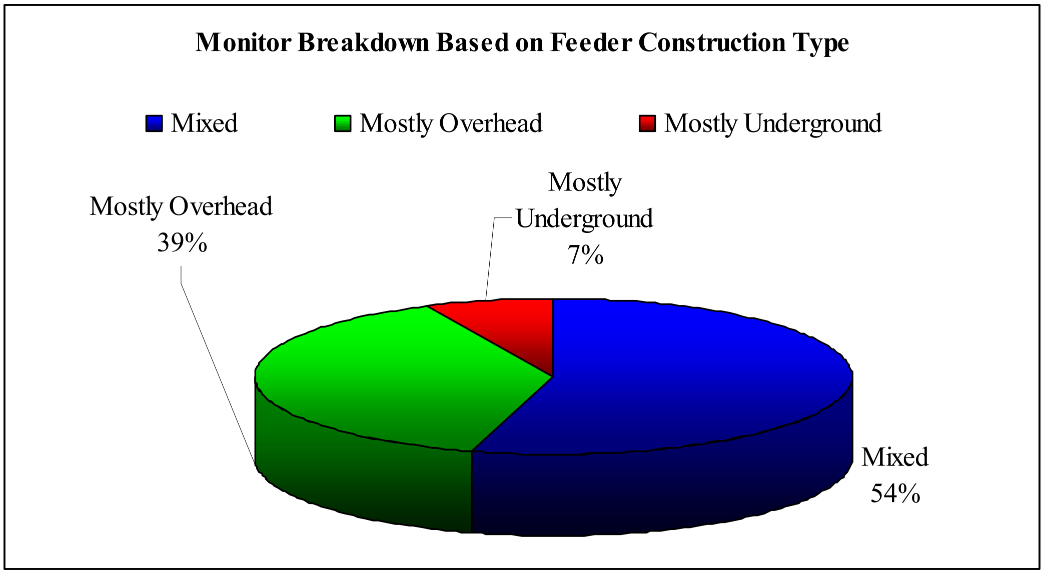


Figure 2-6
Breakdown of Monitor Locations in Distribution Systems Based on Feeder Construction Type

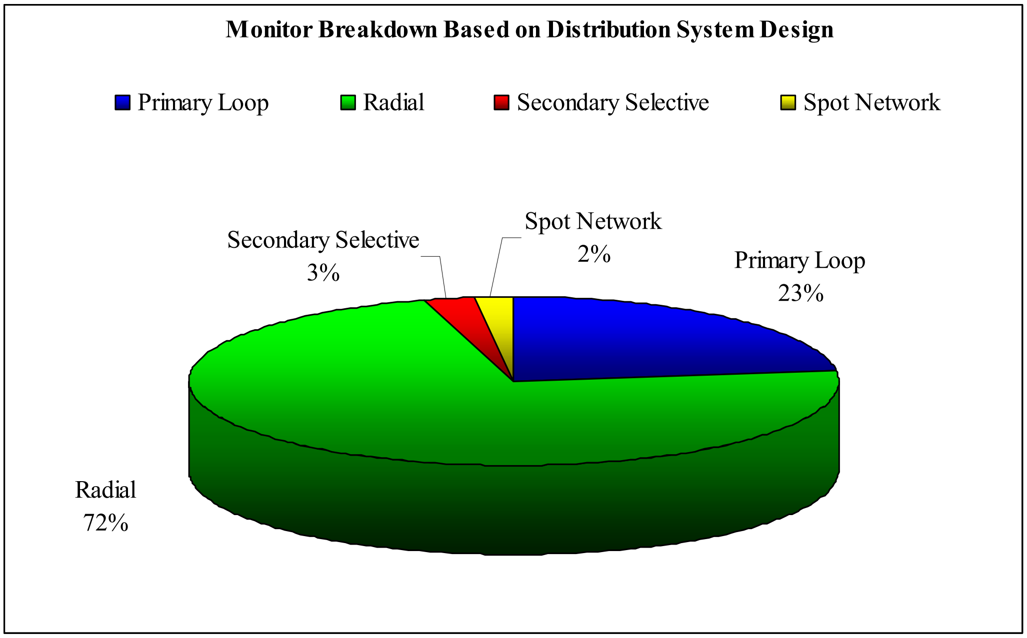


Figure 2-7
Breakdown of Monitor Locations in Distribution Systems Based on Distribution System Design

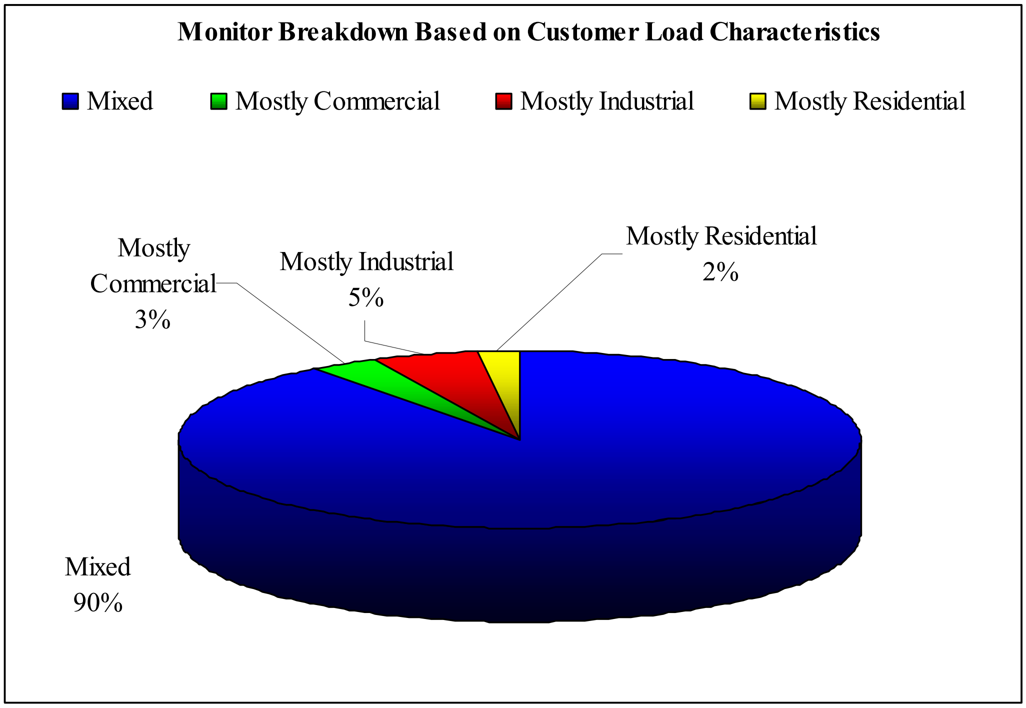


Figure 2-8
Breakdown of Monitor Locations in Distribution Systems Based on Customer Load Characteristics

Lightning Density

As shown in Figure 2-1, lightning represents one of the primary causes of fault-induced voltage sags on the utility T&D system. Transmission and distribution exposure in high lightning density areas will generally result in a greater number of voltage sags compared to a location with low lightning exposure. To identify the lightning exposure at the monitor locations, utilities were asked to provide a subjective breakdown of the number of monitors that represents a location with the following lightning density characteristics:

- High Lightning Density (>8 flashes/sq km/year)
- Medium Lightning Density (>3 and <8 flashes/sq km/year)
- Low Lightning Density (<3 flashes/sq km/year)

Figure 2-9 shows the breakdown of the number of monitors that fall within each one of the lightning categories identified above.

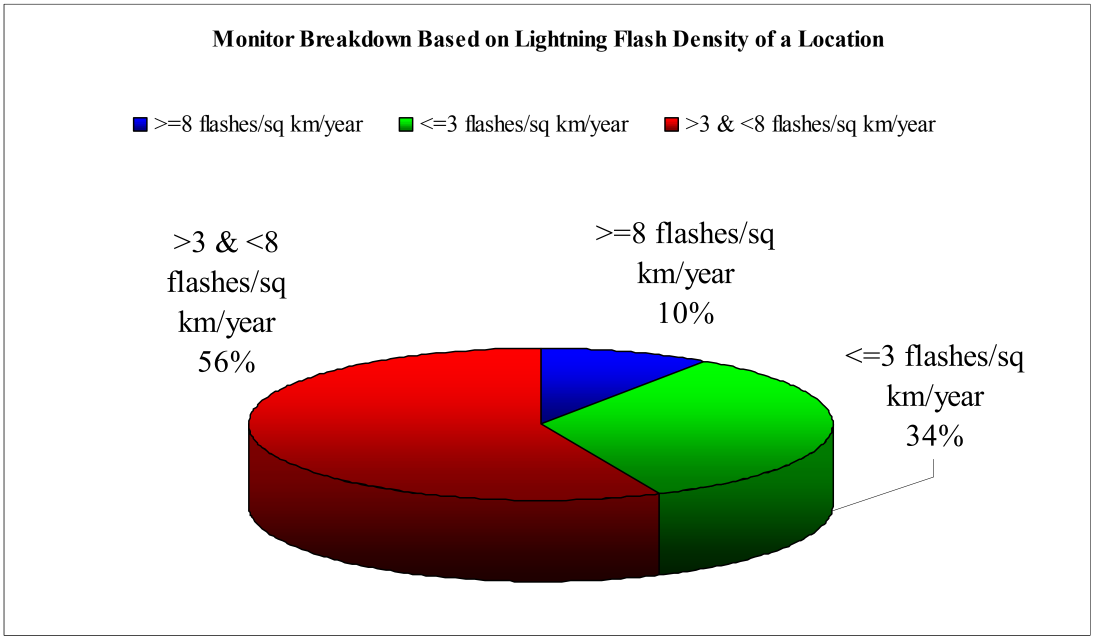


Figure 2-9
Breakdown of Monitor Locations Based on Lightning Density of the Location

The lightning density at the location of the monitor is a good indication of the lighting activity in the general area of exposure that will result in voltage sag at the monitoring location. However, in some cases, the T&D exposure at locations remote to the monitor location may be quite different than the lightning density at the monitor location. This will result in an inaccurate indication of the impact of lightning on the voltage sag rate at a given monitoring site. A more accurate way to gauge the lightning density for each monitoring location would be to conduct a detailed area of vulnerability calculation for each monitoring location and then evaluate the lightning exposure of the area of vulnerability for each location. Understandably, conducting such an analysis for all 1,600 or more possible locations would not be feasible. Therefore, classifying monitor locations based on the lightning density at the site where the monitor is connected was chosen as an alternative.

Total Number of Monitor-Years of Sag Data

The amount of data collected by a monitor over a given period is typically represented in monitor-years worth of data. For example, data from 5 monitors for a period of 2 years signifies 10 monitor-years worth of data. As a reference point, Phase I of the DPQ project, conducted during the 1994-95 timeframe, collected approximately 900 monitor-years worth of data. To identify the amount of data that has been collected by utilities since the completion of DPQ Phase I, the utilities were asked to provide an indication of the number of monitors with less than two years worth of data, between two and four years worth of data, and more than four years worth of data. This represents approximately 4,000 monitor-years worth of data, which is almost 5 times in magnitude more than was collected as part of DPQ Phase I.

Conclusion

The next step of this initiative will be to collect PQ data from a select sample of monitoring sites that adequately represents the range of characteristics seen on this site survey. Data will be collected from the participating utilities from this select number of sites and analyzed to quantify different statistical parameters of voltage sags and evaluate the impact of site characteristics on voltage sag rate. This DPQ Phase II effort, to characterize voltage sags, represents approximately more than 4,000 monitor-years of data covering different network topologies (radial, loop, network, etc.) and voltage classes (distribution, transmission, etc.). Analysis of this data set will provide opportunities to:

- Validate results from the DPQ Phase I that was performed in the early 1990s
- Provide design guidelines for manufacturers to design equipment with proper immunity for electrical environments
- Quantify the impact of important site characteristics on the voltage sag rate
- Evaluate the technical considerations for using power quality in a performance-based rate structure
- Compare the power quality trend from the early 1990s to recent years

3

A SITE SELECTION SCHEME – STATISTICAL APPROACH

Defining Premium Power

The term “premium power” implies an enhanced level of service required by customers that are particularly sensitive to power disturbances. The disturbances that affect these customers do not affect the majority of customers served by the utility. For a majority of utility customers, including residential, commercial, and some industrial customers, quality of power as measured by traditional reliability indices such as SAIFI and SAIDI are adequate to define the quality of power. For an ever-increasing group of customers, most notably those with a high population of variable-speed drives and/or computer controls, the requirement of power quality is defined in terms of the number of momentary interruptions and voltage sags that customers receive over a given time period. For these customers, quality of power is as important as reliability of power, and providing this required quality serves as the basis of a premium power offering.

Defining grades of power in terms of power quality indices was proposed in earlier EPRI technical reports [1,2]. Defining different grades using these power quality indices ($SARFI_{70}$, $SARFI_{ITIC}$, $SARFI_{SEMI}$, etc.) would allow customers to choose what grades of power they would require and the cost implications for obtaining that grade versus the normal or basic grade. However, because of the variation of grid power from site to site, defining a power quality level that will fit all possible combinations of geographical location and electrical system characteristics is difficult, if not impossible.

For instance, it would not be useful to develop one set of power quality levels that includes transmission and distribution systems (with different voltage ratings) with different feeder characteristics (overhead with either radial feeders, or underground with radial feeders, etc.) all together. This would make it difficult to evaluate concerns for performance at individual utility locations or for overall systems. To effectively address the power quality levels, performance should be evaluated for portions of the overall system with similar design and power quality characteristics. Moreover, weather-related events are a major cause of voltage sags, and the number of voltage sags to some extent depends on the exposure of the distribution and transmission lines to lightning events. Expecting the same sag rate at two different locations with a large difference in flash density rate is not feasible. Therefore, lightning flash density also becomes a variable that affects voltage sag rate.

Application of Statistical Techniques for Site Selection

Over the past decade, electric utilities and related research organizations throughout the world have been engaging in power quality benchmarking projects to determine baselines of service. There are several ways to do this. One is to use software and system models to simulate the power quality levels on the system. A second method is to monitor at every substation and at various points on each feeder. In a world of unrestricted resources, a utility interested in measuring PQ at its substations would simply install a PQ monitor at each of its substations. However, the cost of monitoring all substations is prohibitive, thus, a utility must limit monitoring to a portion of its substations.

Technically, the most definitive approach to define the expected quality of power (as defined by not only the voltage sag indices, but also indices for other PQ phenomenon such as harmonics, voltage unbalance, or transients) involves long-term power quality monitoring by *statistically* choosing sites that represent a system under consideration [1-3]. However, it is impractical to wait several years before providing customers information regarding the quality level of the power that they are receiving. In such cases, if utilities (from different geographic locations) have power quality monitoring data from *statistically* valid system-wide benchmarking projects, a site selection procedure would then have to be formalized to identify a population of monitoring locations whose data would provide an *unbiased* representation of the types of transmission and distribution feeders present across the United States. Data from these selected monitoring locations can then be used to determine baseline power quality at the national level. However, it is important to realize that any national or system-wide benchmarking data will not necessarily reflect the condition at a particular site, but only provide a *typical* range that may be expected for sites with similar design and power quality characteristics.

DPQ Phase II – A Next Level to Evaluate Electrical Environment

Since the completion of the Distribution Power Quality (DPQ) project in 1995, several utilities have implemented system-wide PQ monitoring programs both in distribution and transmission systems. The wealth of data collected since the DPQ project provides a unique opportunity to synthesize meaningful information regarding variability of grid power, specifically on voltage sag rates, based on system characteristics. It also presents a unique opportunity to compare the results of the DPQ project with the data from these monitoring programs.

One of the major efforts of the Premium Power project for 2001 was to determine the types and sizes of monitoring systems employed by different utilities since the completion of the DPQ project in 1995. Based on this information, a site selection process to identify a smaller subset of monitoring locations from an existing population of monitors (deployed by participating utilities since the completion of the first phase of the DPQ project in 1995) will be developed in this report. At present, 11 utilities in the United States (see Figure 3-1) and 1 utility outside the United States (China Light & Power) agreed to provide the data (China Light & Power is not shown on the map).

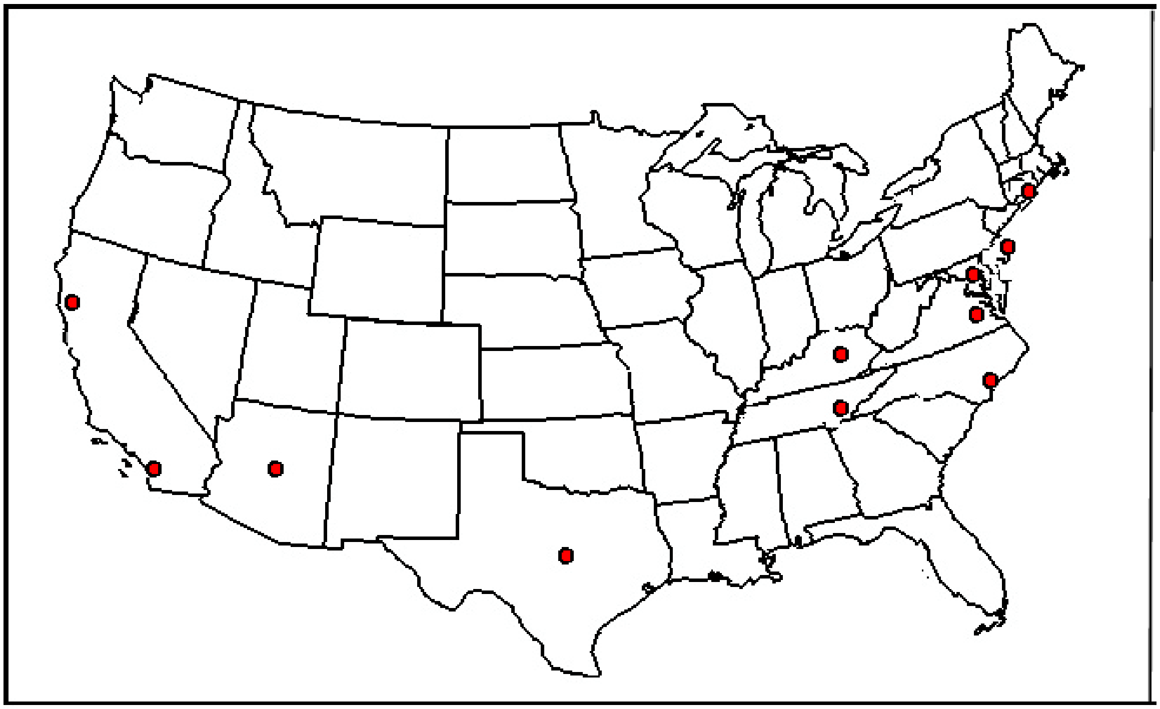


Figure 3-1
Locations of the 11 Participating Utilities That Volunteered

To determine whether the site selection procedure is accurate, two objectives will have to be met. First, to accomplish the objective of modeling power quality as a function of influencing factors, the monitoring locations need to provide coverage for the range of power quality conditions in the United States. Second, the data from these selected monitor locations must provide statistically valid estimates of power quality for transmission and distribution systems nationally. Consistent with the first objective, as stated above, a smaller subset of monitoring locations that would result as an outcome of a site selection process should adequately represent the range of characteristics seen on transmission and distribution feeders in the United States. This required that some form of a controlled probability allocation scheme (while distributing statistical valid samples among participating utilities) be adapted for the site selection process to ensure that both smaller and larger utilities, and common and uncommon feeder characteristics (at a national level) were well represented in the "statistically valid" monitors (an outcome of the final site selection process).

One challenge, however, was to limit the burden on utilities regarding the amount of data initially required. For large utilities with hundreds of pre-installed monitors, it would be unnecessarily burdensome to initially ask for a great deal of data for each monitor location. A tradeoff had to be made between having information for every possible candidate monitor location and enough information from individual utilities regarding all of their monitoring locations to maintain a certain level of statistical validity.

In an effort to reduce the amount of data initially required from each utility, the overall site selection process would therefore have to be performed using several stages of sampling with the details of sampling unit information increasing at each successive level. A possible approach to prioritize individual candidate stages is to evaluate the importance of power quality site descriptors and how they affect the voltage sag rates at a particular site. These relationships can be demonstrated by using different statistical significance tests (such as leverage plots) that are used to determine the effect of a variable on the regression model of voltage-sag rate [1]. Leverage plots are used to determine how much of an effect a given independent variable has on the dependent variable. The dependent variable is the SARFI_{70} value associated with a specific site, and the independent variable is a specific descriptor value that may help to explain the SARFI_{70} value. Examples of specific power quality descriptors include 1) voltage base rating, 2) feeder length, 3) feeder types (overhead or underground), 4) lightning flash density, 5) transformer MVA size, 6) load density, and 7) feeder configurations. Detailed descriptions and comparison results to demonstrate the effects of individual power quality descriptors on voltage sag rates (SARFI_{70}) are provided in the preceding sections.

As an example, a possible way to classify monitor locations (for individual utilities) during the first sampling stage could be in terms of “base voltage ratings.” Base voltage rating is the voltage rating of the feeder to which the monitor was connected at the substation. The voltage rating of a site is considered to be an important variable. Distinct design philosophies are sometimes used at different voltage levels, which affects the overall sag performance of the site. In addition, voltage levels also dictate the total circuit length and exposure level in some cases, which can affect the overall voltage-sag rate. The clustering for this stage, therefore, can be done on the basis of voltage classes (distribution levels – 0-5kV, 5-15kV, 15-34.5kV; transmission levels – 34.5-69kV, 69-230kV, >230kV). Some form of probability sampling would then have to be performed to identify a smaller subset of monitoring locations (corresponding to different voltage classes) from an existing population of monitors deployed by participating utilities since the completion of the first phase of the DPQ project in 1995.

By keeping track of the successive probabilities of a monitor’s chances of being selected at individual sampling stages, sampling weight factors can also be developed. These factors, at successive sampling stages, provide a means to extrapolate the data from a few “statistically valid” monitors to estimate baselines of power quality at a national level. Sampling weighting factors are calculated as reciprocal of the selected probabilities. Weighting factors associated with individual monitor locations are indicative of the fraction of the population the location represents. Details of the theoretical frameworks required for implementing some form of controlled probability sampling techniques are provided in the preceding sections.

Once the initial sampling stages in the overall site selection process have been completed, and smaller sets of “statistically valid” monitoring locations across the United States have been accurately identified, more detailed data from these selected sites will then have to be requested. The information collected during this stage will then be used in the final analysis for establishing baselines of power quality at a national level and for sampling at the next stage.

Proposed Probability Sampling Technique – A Brief Overview

The core value of statistical methodology assists one in making inferences about a large group (a population) based on observations of a smaller subset of that group (a sample). Using a combination of powerful statistical tools known as "inferential statistics" and "unbiased sampling techniques," one can collect data that actually represent the views of the entire population from which the sample was taken. However, it should not be forgotten that the real purpose is to benchmark power quality for the whole population.

To gain a high confidence level for ensuring that the sample represents the whole population, the researcher has two primary objectives.

- It is desirable to produce estimates that are unbiased
- It is desirable to produce estimates that are precise (for a certain confidence level)

Bias as a statistical term means error. This means that the expected value of mean PQ estimates derived from monitoring equals the true mean in the population. To avoid biases, researchers might employ random sampling procedures. The use of random sampling ensures that the resulting sample is representative of the monitoring location's population as a whole, and does not lead to biased estimates of population PQ measures.

Statistical precision on the other hand, is a function of two factors--the variability of the underlying population data and the size of the sample. In statistics, precision is defined as the percentage deviation of the true population mean (unknown to the researcher) from the sample mean (determined from the monitoring results) at some specified level of confidence. The smaller the percentage deviation, the more precise the sample mean is as an estimator for the population mean. In designing a sample, the researcher has no control over the variability of the underlying data and must accept it as a given. However, the researcher can increase the sample size to improve precision. In addition, the researcher can employ stratified sampling designs (over simple random sampling) that improve precision without requiring increases in the overall sample size.

Sampling Methods (Random Versus Non-Random)

As previously noted, equally important to the size of the sample is the determination of the type of sampling to be done. Non-random sampling techniques such as systematic sampling, cluster sampling, incidental sampling, or purposive sampling will always produce larger sampling errors than true random techniques (for the same sample size). This is because non-random techniques generate the expected random sampling error on each selection in addition to errors related to the non-random nature of the selection process. Moreover, these sampling techniques usually do not produce samples that are representative of the general population from which they are drawn. The greatest error occurs when the researcher attempts to generalize the results of the survey obtained from the sample to the entire population. Such an error is insidious because it is not obvious from merely looking at the data, or even from looking at the sample.

The easiest way to recognize whether a sample is representative of the entire population is to determine if the sample was selected randomly. Random sampling *always* produces the smallest possible sampling error. In a very real sense, the size of the sampling error in a random sample is affected only by random chance. The following criteria must be met to be considered a random sample.

- Every member (monitor location) in the total population has an equal chance of being selected for the sample
- Selection of one monitor location should in no way influence the selection of another. It will have to be used with a homogenous population. (That is, one composed of members who will possess the same attribute in which the researcher is interested.)

One of the most useful statistical-based random-sampling techniques is the stratified random-sampling method. This report outlines the use of this technique for formalizing an overall site selection process and how it may be used independently to determine the optimal PQ monitors required for establishing baselines of PQ at a national level. Mathematical procedures required for implementing these techniques are also addressed.

Design Procedures and Site Characteristics (Descriptions and Data) Used for Overall Method Development – Monitor Location Selection

Sampling Frame

The first step in developing a site selection design process is to define the sampling frame, or the population from which the survey sample will be chosen. The sampling frame usually corresponds fairly closely to types and sizes of monitoring systems employed by participating utilities since the completion of the DPQ project in 1995. The precise definition of the sampling frame will be guided by the ultimate uses of the power quality estimates.

The participating utilities were asked to provide the number of monitors in each location based on a generic monitor location diagram shown in Figure 3-2. It was assumed that the voltages greater than 34.5kV were considered to fall under the transmission voltage level class.

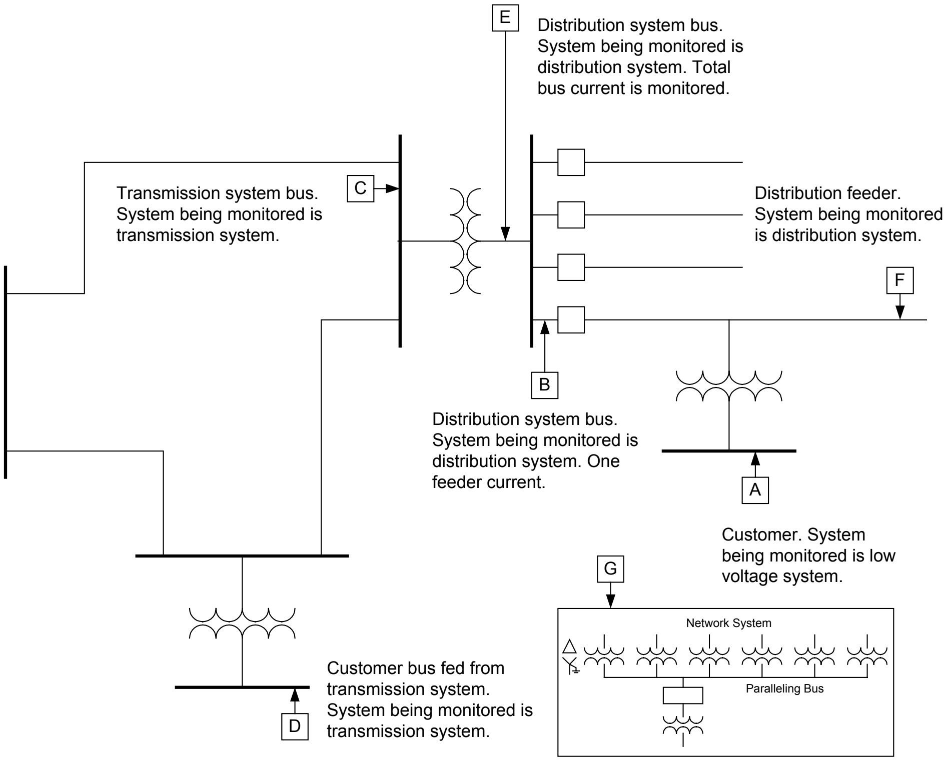


Figure 3-2
Different Generic Categories of Monitor Locations

The different location categories identified in Figure 3-2 are provided below.

A – Customer Bus Distribution System

B – Feeder Monitor Distribution System

C – Transmission Bus Transmission System

D – Customer Bus Transmission System

E – Distribution Bus Distribution System

F – Distribution Feeder Distribution System

G – Underground Network

H – Other (explain)

Based on the data provided by the different utilities, the breakdowns of monitors in the above categories are provided in Figure 3-3.

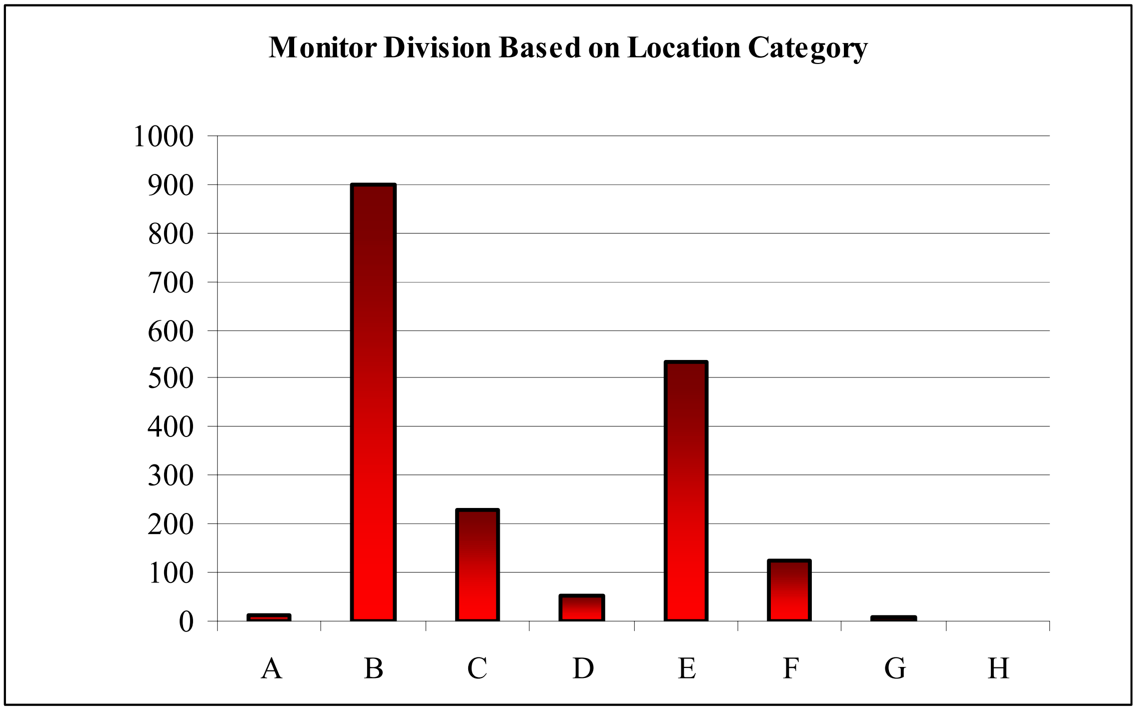


Figure 3-3
Breakdown of Monitors Based on Location Category

Out of a total of 1,861 PQ monitors (target population) that have been deployed by the participating utilities, a majority of them are on the distribution level. Approximately 278 locations were either at transmission busses or at customer busses fed directly from the transmission system. Data from these monitors will then provide an indication of voltage sag rates originating from the transmission system since the exposure for these locations is primarily the transmission system. This will provide some unique insight into voltage sag rates for customers who are fed directly from the transmission system.

The breakdown in terms of total number of monitor locations present at the transmission, distribution, and secondary network type systems (downtown areas) is shown in Figure 3-4. This represents approximately 4,000 monitor-years of data, which is almost 5 times more in magnitude than was collected as a part of the DPQ Phase I effort. The breakdown of monitors based on monitoring period is shown in Figure 3-5.

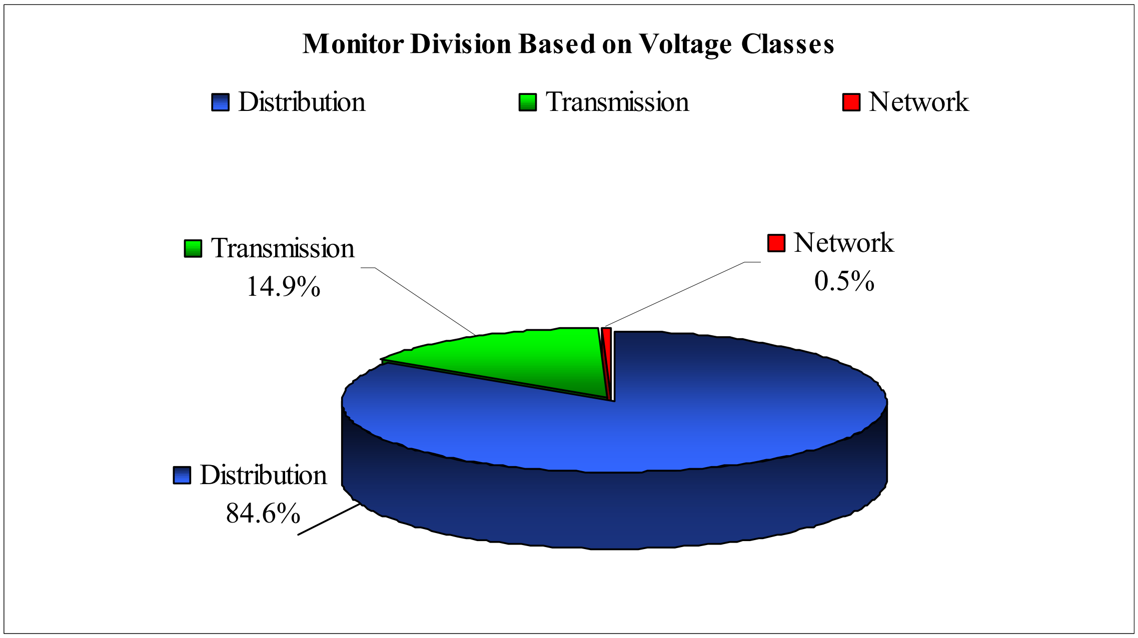


Figure 3-4
Breakdown of Monitors Based on Voltage Classes

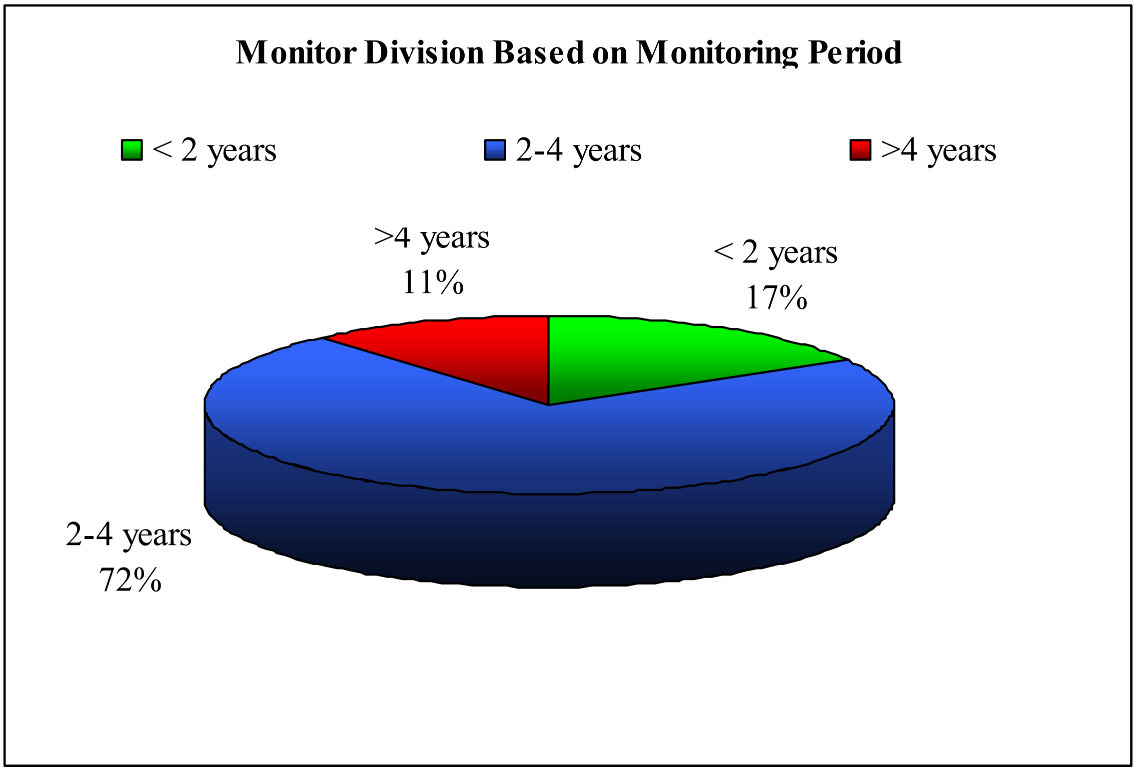


Figure 3-5
Breakdown of Monitors Based on Monitoring Period

Develop Sample Design - Power Quality Site Descriptors

Once the sampling frame has been defined, the next step is to design the sample. The design process in turn has three component tasks:

1. An ability to identify one or more power quality variables that have high correlations with the number of voltage sags. Voltage sag rates at the transmission levels will be different from the voltage sag rates at the distribution systems. These rates will even be different depending on whether the feeders are either overhead/underground, or radial/loop, or urban/rural/suburban. For instance, rural sites have more voltage sags and momentary interruptions as compared to urban or suburban. A regression analysis using leverage plots was first proposed in reference [1] to grade the potential usefulness, and to classify potential candidate power quality site descriptors from those specified in Table 3-1. Moreover, a decision regarding which site descriptor to use in the consecutive stages of sampling would then be made from this importance list. Some of the site descriptors are correlated with other stratification variables (provided in Table 3-1), and are therefore somewhat substitutable.

2. An ability to choose a sampling method. Depending on the availability of other information (available data and precision), the researcher may wish to perform either a simple random sampling or a stratified random sampling. If a stratified sampling method were to be employed, the researcher would also need to identify which sample allocation techniques to use. Care should be taken in the decision to choose the most suitable allocation technique. Otherwise, inaccurate representations would be made, particularly while choosing utilities that have fewer monitors installed in their system. For example, even though Utility 1 has 20 monitors in their distribution network with uncommon feeder characteristics compared to Utility 2 with 900 monitors with common feeder characteristics, use of the proportional allocation scheme would tend to weigh more for the utility with the greater number of monitors. As a result, out of the total statistically valid monitors that are chosen as a final outcome from the site selection process, the maximum number of monitors would correspond to Utility 2. To avoid this type of misrepresentation, some form of controlled probability allocation scheme will have to be adapted for the site selection process that ensures that both smaller and larger utilities, and common and uncommon feeder characteristics are well represented.

3. An ability to define the boundaries for each of the selected power quality descriptors. For example, for stratified random sampling, if base voltage ratings are chosen as the descriptor in the first stage of sample design, the researcher needs to determine:
    - How many strata (within each descriptor variable) will be adequate
    - What the subjective break-points will be between them
    - The mean and standard deviations of these individual ranges if stratified random sampling is used

Power Quality Descriptors – Identification and Classification

There are a number of power quality variables that may impact the voltage sag rate at a particular site. A possible list of important stratification variables that could have been used is shown in Table 3-1.

Table 3-1
List of Candidate Variable Categories to Consider

Power Quality Descriptor	**Range and Units**
Base Voltage Rating (Substation Primary and Secondary)	In kV
Total Three Phase Feeder Length	In kilometers
Approximate Load Density	Rural, Urban, Mixed
Size of Substation	In MVA
Substation Grounding Feeder Configuration	Delta, wye, wye grounded, etc.
Lightning Flash Density	In strokes/sq km
Type of Service Systems	Radial, Loop, Mixed
Number and Size of Substation Transformers	
Feeder Characteristics	Mostly Overhead, Underground, Mixed
Customer Load Characteristics	Industrial, Residential, Commercial
Monitor Duration Period	<2 years, 2-4 years, >4 years

Note that the primary interest in stratification is to ensure adequate representation of variables relevant to power quality modeling. Moreover, it is a means to improve a survey precision and at the same time control sample size for important subpopulations. However, to develop a realistic sample design, the number of site classifications would have to be 1) realistic, 2) manageable, and 3) where participating utilities could readily obtain the data. This simply suggests that only a limited number of power quality descriptors of Table 3-1 could be chosen for overall site classification and sample design. With all this in mind, the researcher needs to determine—with the aid of engineers, system operators, end users, and statistical tests for significance such as multi-variable regression (such as leverage plots)—the "best" descriptor(s) or stratification variables for the first stage of the site selection process.

Figure 3-6 shows the comparison of probability >F for each of the descriptors, indicating which descriptors have significance and which are not significant in explaining the dependent variable. Probability >F is a statistical measure used in leverage plots to determine how much of an effect a given independent variable has on the dependent variable. The dependent variable is the $SARFI_{70}$ value associated with a specific site, and the independent variable is a specific descriptor value that may help to explain the $SARFI_{70}$ value.

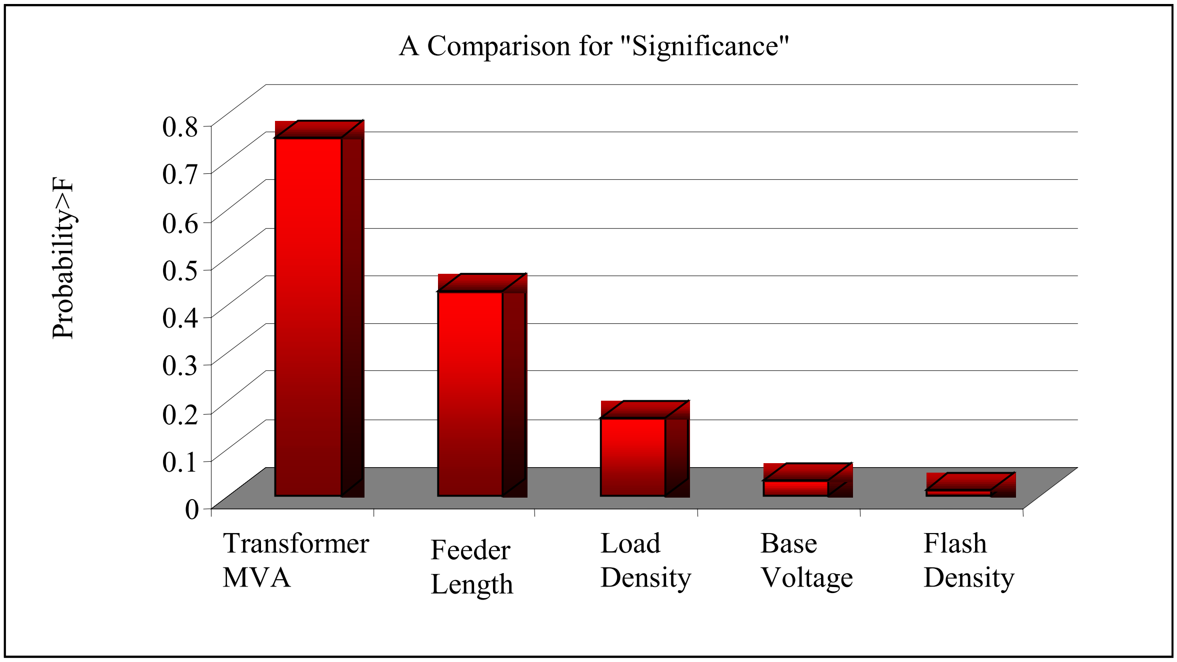


Figure 3-6
Importance of the Five Power Quality Site Descriptors and How They Affect $SARFI_{70}$

A probability list for the five descriptors in the order of significance is shown in Table 3-2. Lower values of probability >F (as indicated in Figure 3-6) or higher percentage value in probability (as indicated in Table 3-2) for lightning flash density and nominal base voltage classes indicate that they are likely to have a greater effect on voltage sag rate as compared to the other descriptors shown in Figure 3-6.

Table 3-2
Probability List for the Five Power Quality Site Descriptors and How They Affect $SARFI_{70}$

Site Descriptors	Probability of “Significance”
Transformer MVA	25%
Feeder Length	57%
Load Density	83%
Nominal Base Voltage	97%
Flash Density	99%

Moreover, the dependent relationships of the variables will also have to be evaluated, and correlation analysis among different site descriptors will have to be performed. In an ideal sampling, there would be no correlation between independent variables. However, it is possible that there might be some correlation between individual descriptors. The correlation coefficient, referred to as the Pearson product-moment correlation coefficient, measures the strength of the linear relationship between two variables. If there is an exact linear relationship between two variables, then the correlation is 1 or –1, depending on whether the variables are positively or negatively related. If there is no linear relationship, the correlation will be zero.

For the five descriptors shown in Figure 3-6, there is only one strong correlation that emerges from computing the correlation coefficients. That correlation between transformer size and voltage class is 0.6198. It is expected, from a physical standpoint, that the larger the transformer, the higher the voltage class. The correlation matrix for the five variables, shown in Figure 3-6, is illustrated in Table 3-3.

Table 3-3
Correlation Matrix for Five Variables

Descriptor	Flash Density	Xfmr MVA	Feeder Length	Nominal Base Voltage	Load Density
Flash Density	1	-0.0495	0.1592	0.0639	0.0123
Xfmr MVA	-0.0495	1	-0.1723	0.6198	0.0384
Feeder Length	0.1592	-0.1723	1	-0.1876	0.0477
Nominal Base Voltage	0.0639	0.6198	-0.1876	1	-0.192
Load Density	0.0123	0.0384	0.0477	-0.192	1

Based on the results obtained in Figure 3-6, Table 3-2, and Table 3-3, the obvious choice for the power quality descriptor to be used in the first and second stage of the site selection process would be nominal base voltage and lightning flash density. Voltage rating of a site (transmission and distribution) is considered to be an important variable because distinct design philosophies are sometimes used at different voltage levels, which affects the overall sag performance of the site. In addition, voltage levels also dictate the total circuit length, exposure level, and load density in some cases, which can affect the overall voltage sag rate. Data concerning the number of monitors located at different voltage levels can be readily obtained. This, therefore, was considered to be the site descriptor during the first stage of sampling.

Lightning represents one of the primary causes of fault-induced voltage sags on the transmission and distribution system of most utilities. Transmission and distribution exposures in high lightning density areas will result in a greater number of voltage sags as compared to a location with low lightning exposures. However, in some cases, the T&D exposure at locations remote to the monitor location may be quite different than the lightning densities at the monitor location. This will result in an inaccurate indication of the impact of lightning on the voltage sag rate at a given monitoring site. A more accurate way to gauge the lightning density for each monitoring location would be to conduct a detailed area of vulnerability (AOV) calculation for each monitoring location and then evaluate the lightning exposure of the AOV for each location. Understandably, conducting such an analysis for all 1,861possible locations would not be feasible, and classifying monitor locations based on the lightning density at the site where the monitor is connected was chosen as an alternative. To identify the lightning exposure at the monitor locations, utilities were asked to provide a subjective breakdown of the number of monitors that represents a location with the following lightning density characteristics:

- High Lightning Density (>8 flashes/sq km/year)
- Medium Lightning Density (>3 and <8 flashes/sq km/year)
- Low Lightning Density (<3 flashes/sq km/year)

For the majority of the participating utilities who failed to provide information that identified their monitors based on AOV estimation to evaluate lightning flash density, it is assumed that all of their monitors have the general lightning density based on the geographical area of their service territory [4] (as shown in Figure 3-7). Due to a greater probability for data uncertainty, this was used as the power quality descriptor in the second stage of the site selection process. The data are corrected for estimated detection efficiency of 70% by multiplying all values by 1.4.

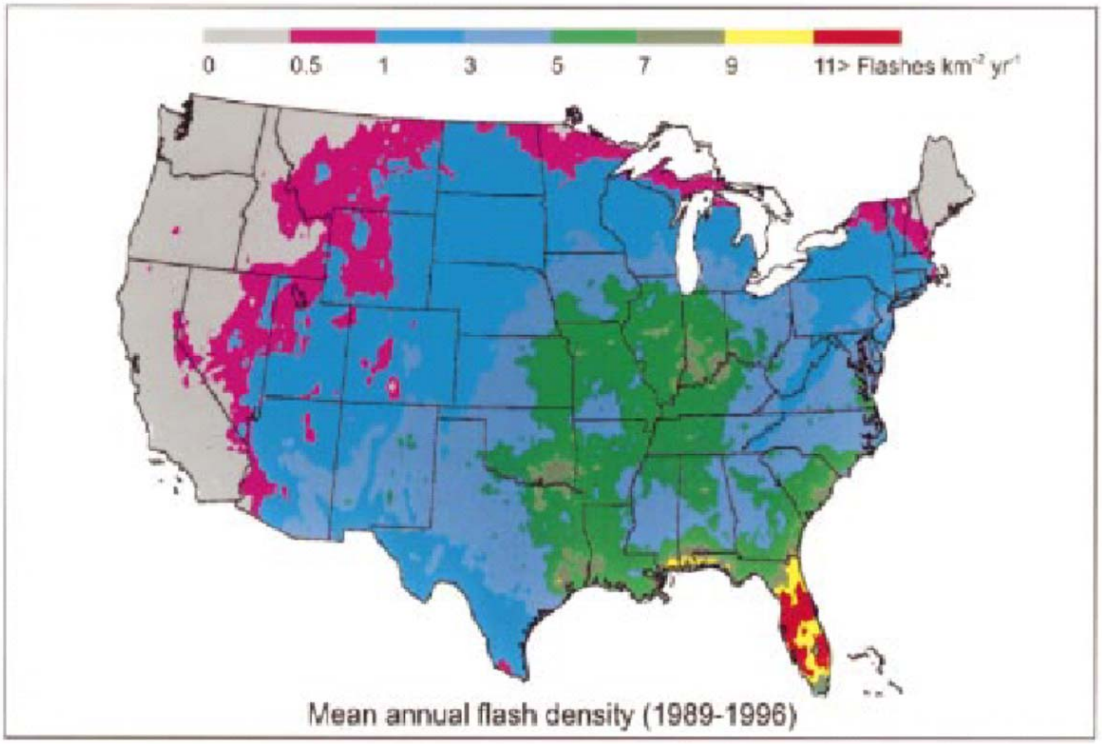


Figure 3-7
Mean Annual Flash Density Map for the United States

Even though descriptors like feeder characteristics (mostly overhead, mostly underground, mixed), types of service systems (radial, loop), and customer load characteristics (industrial, residential, commercial) might also have significant impact on voltage sag rates, these were not used during the evaluation of "significance" among power quality site descriptors due to the lack of sufficient data. Once the initial two sampling stages in the overall site selection process have been completed, and smaller, statistically valid monitoring locations across the United States have been accurately identified, more detailed data from these selected sites regarding the above descriptors will then have to be requested. The information collected during this stage will then be used in the final analysis for establishing baselines of power quality at a national level and for sampling in the next stage.

Constructing Stratum Boundaries

The next step in the design process for stratified sampling techniques is to experiment with cell boundary definitions and number of strata to determine the level of precision that would result from each. One general strategy in choosing stratum boundaries and the number of strata is to choose these boundaries in such a way that the resulting variance, is minimized [5-8]. This method, commonly called cumulative root frequency ($cum\sqrt{f(y)}$, where $f(y)$ is the frequency distribution of the power variable used for stratification) is based on:

- Grouping the stratification variables into a number of smaller classes. As the stratification variable chosen in the first stage of sampling is the voltage levels, one possible way would be to divide them into seven categories as illustrated in the first two columns of Table 3-4.
- Determining the frequency distribution for each of these categories (second column of Table 3-4) for the power quality descriptor chosen in the first stage of sampling. This is demonstrated in the third column of Table 3-4.
- Determining the cumulative frequency distribution of the square root of these individual categories. This is demonstrated in columns four and five of Table 3-4.

Table 3-4
Calculating the Stratum Boundaries Using the Cumrootf Approach

Original Categories	**Break Ups**	**Monitors**	**CumFreq.**	**Cumrootf**	**Assume Strata=2** Optimal alloc. of stratas	**Assume Strata=3** Optimal alloc. of stratas	**Assume Strata=4** Optimal alloc. of stratas	**Assume Strata=5** Optimal alloc. of stratas	**Assume Strata=6** Optimal alloc. of stratas
1	5	8	8	2.8284	1	1	1	1	1
2	12.5	400	408	23.0274	1	1	1	1	1
3	15	169	577	47.0483	1	1	1	2	2
4	23	819	1396	84.4113	1	2	2	3	3
5	26.4	137	1533	123.5649	2	2	3	3	4
6	34.5	50	1583	163.3518	2	3	4	4	5
7	500	278	1861	206.4911	2	3	4	5	6

The next step in the selection of the boundary requires an assumption be made regarding the number of strata to be used. Assuming that the number of strata to be used is five, the next step is to determine whether this represents the best characteristics to construct the boundary. Since the total of the Cumrootf is 206.4911, an equal division in interval based on the $cum\sqrt{f(y)}$ scale would result in points divided into: 41.292, 82.596, 123.895, 165.193, and 206.4911. These represent the upper boundaries for the five strata and the boundary intervals will now have to be derived from these values. The result is shown in the ninth column of Table 3-4 where boundary ranges for the three strata will be between 0-12.5; 12.5-15; 15-26.4; 26.4-34.5; and >34.5. Note that a similar thought process can be adopted if the number of strata other than five is selected.

Evaluating the Optimal Number of Strata

The next step is to evaluate the optimal number of strata required for final sample design. This can be obtained by observing the reduction in variance affected by the addition of another stratum. That is, the variance from L strata is compared with the variance resulting from L-1 strata. However, at this pre-stratification stage, the optimal sample size has not yet been decided. Thus, without undergoing the final stage of sampling, L is selected in the algorithm based on the lowest decreasing variance rate. This is illustrated in Equation 3-1 and a comparison in variance ratios is provided in Table 3-5.

$$\text{Variance Ratio} = \frac{VR_{L\text{-}1}}{VR_L} = \frac{\sum_{h=1}^{L}\left(\frac{N_h}{N}\right) * \sigma_h^2}{\sum_{h=1}^{L-1}\left(\frac{N_h}{N}\right) * \sigma_h^2}$$ **Equation 3-1**

where:

N is the total population;

Z is the confidence level;

N_h is the total number of points (population) in the h^{th} strata;

σ_h is the standard deviation of the variable of interest in the h^{th} strata;

L is the total number of strata;

Table 3-5
Comparison in Variance Ratios Between L-1 and L Strata

$\frac{VR_{L=2}}{VR_{L=3}}$	$\frac{VR_{L=3}}{VR_{L=4}}$	$\frac{VR_{L=4}}{VR_{L=5}}$	$\frac{VR_{L=5}}{VR_{L=6}}$
1.12	1.15	**0.98**	1.17

It is clear from the values of Table 3-5 that a stratum size equal to five results in the lowest variance ratio level. This would suggest that, to obtain the best characteristics for stratum construction, the number of strata one would need would have to be five. This was then used to derive an optimal number of samples using the controlled allocation scheme in the first stage of sampling.

Choosing a Random Sampling Technique

The researcher next needs to select which random sampling technique to adopt during the successive stages of sample design. Detailed mathematical procedures of the stratified random sampling using proportional allocation and stratified random sampling using Neyman allocation are provided in Appendix A. Results obtained during the successive stages for the overall site selection process were provided previously. A comparison between the two sampling methods (Proportional and Neyman allocation) is only provided for the first sample stage.

Overall Sampling Plan Execution – A Three-Stage Site Selection Scheme

As a part of DPQ Phase II, a site selection approach to assess power quality levels at a national level using the data provided by participating utilities was developed in this report. Moreover, studying the power quality issues separately for one participating utility is not that different from studying issues nationally. Most objectives of the site selection process required for an individual power quality study could fit into the sample design process proposed in this report. The major precept is that random sampling techniques are used to the extent possible because they are statistically valid, widely accepted, and always would provide an *unbiased* representation of the monitors located at transmission and distribution levels. The selected monitors can then be used to establish baselines of power quality at a national level. The preceding sections provide a step-by-step approach to formalize the overall site selection process.

First Stage in Sample Design – Voltage Levels are the Only Site Descriptor Used

Table 3-6 illustrates the allocation of monitors among different voltage clusters or bins that were used as inputs based on the data provided by the participating utilities. Other input variables, namely mean and standard deviation (based on SARFI_{70}), are also evaluated using proxy data from the DPQ study for the distribution system, and proxy data from the SEMI F47 study [9] for the transmission system. While not as extensive as the DPQ study, data from the SEMI F47 study is representative of 16 sites with 30 monitors and serves as a good resource on the quality of transmission level service. As mentioned in the previous section, a stratum size of five was considered adequate.

Table 3-6
Sampling Distribution of Monitors for Population 1,861 (Example Interpretation: When the Voltage Level is Between 0 and 12.5kV, the Average Number of Sags Below 70% is 8.73 per Year)

Stratum	Voltage Level Categories	Number of Monitors (Population N=1,861)	μ	σ
1	0-12.5	408	8.73	15.87
2	12.5-15	169	22.49	21.21
3	15-26.4	956	27.81	17.87
4	26.4-34.5	50	6.69	5.35
5	>34.5	278	3.8	4.36

Comparison Between Proportional and Neyman/Optimal Stratified Sampling

Determining Adequate Sample Size – Proportional Sample Design

Figure 3-8 illustrates the graphical variations in the optimal sample size and the precision levels for three different confidence levels using Proportional stratified random sampling techniques. These values were derived using the generic form shown in Equation (A-2) and Equation (A-5). The total population in this case was 1,861. Under this design, the 1,861 monitors are distributed among the 5 sampling cells in proportion to the number of monitors in the population residing in each cell. By doing so, it ensures the greatest precision possible from any given sample size. Table 3-7 provides the amounts by which the samples were proportioned for the 95% confidence level.

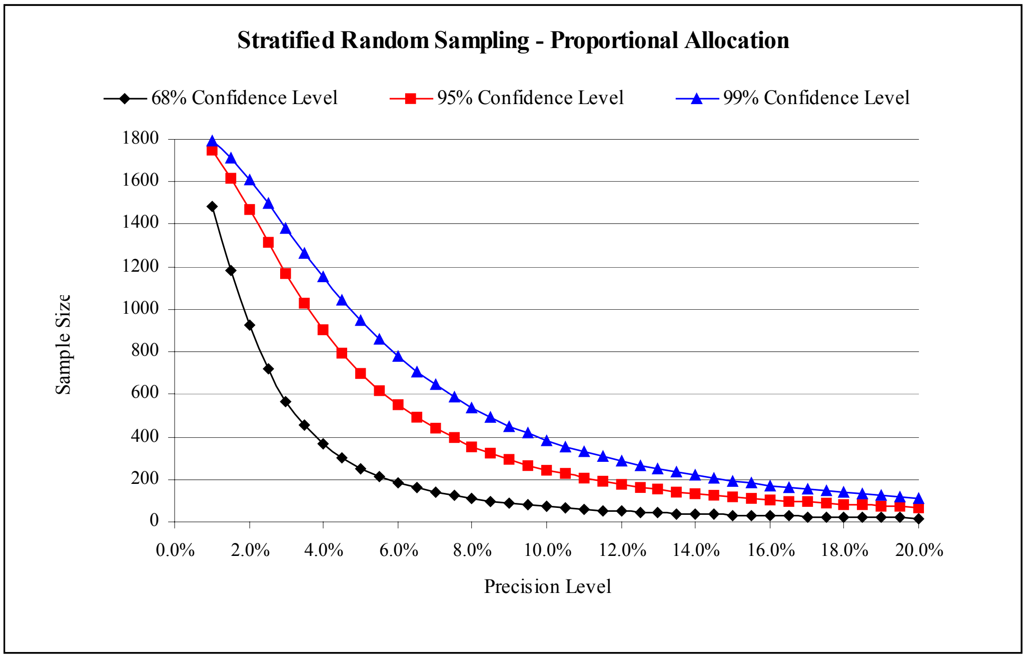


Figure 3-8
Variations in Optimal Sample Sizes Versus Precision Levels (Total Population N = 1,861–Proportional Allocation Based Stratified Random Sampling

Table 3-7
Distribution of Samples Inside the Individual Categories (Confidence Level = 95%)

		Category – Voltage Classes				
		0-12.5 kV	12.5-15 kV	15-26.4 kV	26.4-34.5kV	>34.5kV
Precision (%)	**Total**	n_{h1}	n_{h2}	n_{h3}	n_{h4}	n_{h5}
2.00%	1471	323	134	756	40	220
2.50%	1316	289	120	676	35	197
3.00%	1166	256	106	599	31	174
3.50%	1027	225	93	528	28	153
4.00%	904	198	82	464	24	135
4.50%	795	174	72	408	21	119
5.00%	701	154	64	360	19	105
5.50%	620	136	56	318	17	93
6.00%	550	121	50	282	15	82

Determining Adequate Sample Size – Neyman/Optimal Allocation

Figure 3-9 illustrates the graphical variations in the optimal sample size and the precision levels for three different confidence levels using Neyman stratified random sampling techniques. These values were derived using the generic form shown in Equation (A-2) and Equation (A-6). The total population in this case was also 1,861.

As Equation (A-6) suggests, the Neyman design allocates the total number of sample points based on each sampling cell's population share, N_h, as well as the variation in the variable of interest (standard deviation). By doing so, the Neyman design ensures the greatest precision possible from any given sample size. Table 3-7 provides the amounts by which the samples were proportioned for the 95% confidence level. Since the voltage class category "15-26.4" is the most populous as well as having high variability, this cell receives the largest percentage of the monitors. In contrast, voltage category ">34.5" gets a lower number of samples as compared to that obtained during proportional allocation.

In general, transmission sites have a much better power quality as compared to distribution sites. Therefore, the $SARFI_{70}$ number at the transmission level will be lower. As a result, having a fewer number of monitors at the transmission sites--as obtained during the Neyman allocation--is logical. In general, the greater the variation in the standard deviation, the greater the improvement a Neyman design will yield over a proportional design, as well as simple random sampling techniques. Based on geographical reasons, as greater variability in $SARFI_{70}$ numbers at a national level for either transmission or distribution levels can be expected, Neyman's stratification scheme was preferred over proportional design in the successive stages of site selection process. Neyman allocation also requires fewer samples to meet a given level of precision and accuracy. This would indicate that a lower number of monitors would be adequate to estimate baselines in power quality at a national level. However, the sample size may need to be sufficiently large to measure not only the extent of power quality, but also changes over time in power quality.

As mentioned earlier, one of the most important considerations in any sample design is that of determining how large a sample is needed to be reliable as well as to meet the objectives of any survey. Risk, as it relates to sample size determination, is specified by two interrelated factors:

- Confidence level - to minimize risk, one needs a high confidence level
- Precision/reliability range – to minimize risk, one needs a low precision level
- Available budget

A compromise between budget and the other two levels will therefore need to be made during the initial stages of any site selection process. An easier solution to this problem could be made by being conservative in terms of precision and confidence levels required for estimating the most optimal sample size. For example, 95% confidence and 5% precision levels are most frequently specified in the determination of sample sizes for sample surveys.

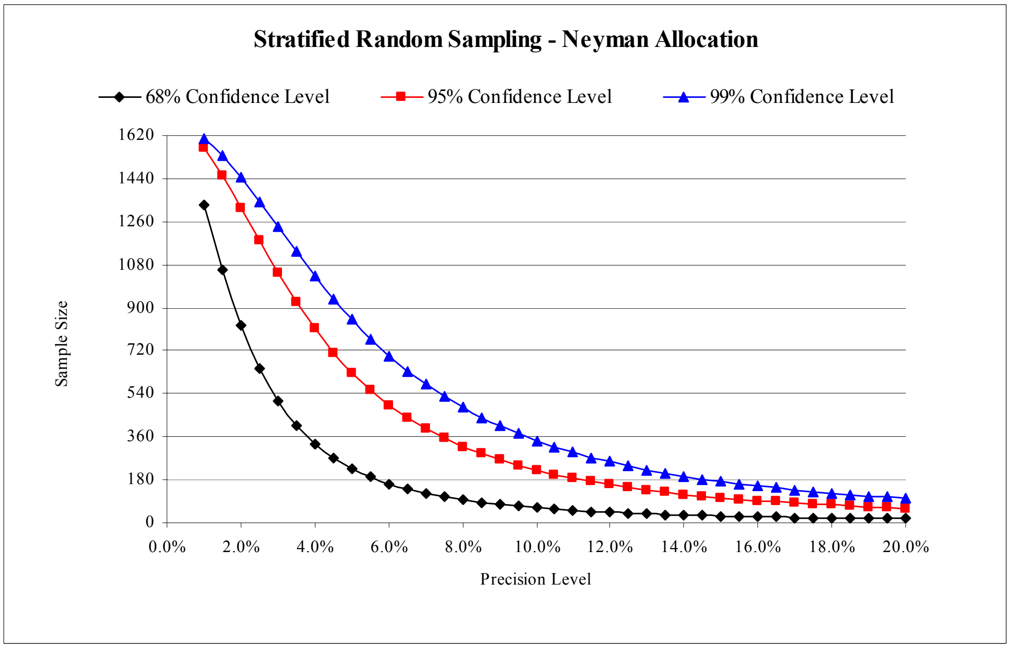


Figure 3-9
Variations in Optimal Sample Sizes Versus Precision Levels (Total Population N = 1,861) – Neyman Allocation Based Stratified Random Sampling

Table 3-8
Distribution of Samples Inside the Individual Categories (Confidence Level = 95%)

		Category – Voltage Classes				
		0-12.5 kV	12.5-15 kV	15-26.4 kV	26.4-34.5kV	>34.5kV
Precision (%)	Total	n_{h1}	n_{h2}	n_{h3}	n_{h4}	n_{h5}
2.00%	1320	299	165	788	12	56
2.50%	1181	267	148	705	11	50
3.00%	1046	237	131	624	10	44
3.50%	922	209	115	550	9	39
4.00%	811	183	101	484	8	34
4.50%	713	161	89	425	7	30
5.00%	629	142	79	375	6	27
5.50%	556	126	70	332	5	24
6.00%	493	112	62	294	5	21

Distribution of Individual Utility Monitors Using Controlled Allocation Scheme

Once the optimal number of sample sizes for the individual clusters or bins (distribution and transmission voltage levels) have been accurately identified, as shown in Table 3-8, the next step in the sample plan involves distributing them within individual participating utilities. In lieu of using a proportional allocation scheme, as described in Equation (A-5), a controlled allocation scheme to distribute monitors within participating utilities during the first stage of the site selection process is proposed and described by Equation (A-7).

Recall that the overall goal of the site selection process was to identify a smaller subset of "statistically valid" monitoring from which estimates to define baselines of power quality at a national level is sought. The data collected from these monitors will then serve as a source of EPRI's concept of combined electric system reliability and power quality indices at a national level. More importantly, the data collected from these monitors will also provide information to electric utilities that want to implement premium power services. As a result, using a controlled scheme during the initial stages of site selection, so that both smaller and larger utilities as well as common and uncommon feeder characteristics (at a national level) are well represented in the statistically valid monitors, is logical. The distribution of "statistically valid" monitors for the different utilities based on different voltage clusters is provided in Table 3-9 through Table 3-13.

Table 3-9
Distribution of "Statistically Valid" Monitors Among Participating Utilities (Voltage Cluster is Between 0-12.5kV)

Utility ID	N_{h1}	Distribution Level	Controlled Allocation	Sampling Probabilities
Utility 1	132	12.5	35	0.27
Utility 2	1	0.38	1	1.00
Utility 2	4	11	4	1.00
Utility 4	190	12	49	0.26
Utility 6	19	12	12	0.63
Utility 7	5	12	5	1.00
Utility 8	6	4.16	6	1.00
Utility 9	3	12	3	1.00
Utility 9	32	12	15	0.47
Utility 10	15	12.47	11	0.73
Utility 11	1	4.16	1	1.00
Total	**408**		**142**	0.35

Table 3-10
Distribution of "Statistically Valid" Monitors Among Participating Utilities (Voltage Cluster is Between 12.5-15kV)

Utility ID	Nh1	Distribution Level	Controlled Allocation	Sampling Probabilities
Utility 3	26	13.8	17	0.65
Utility 5	7	13.2	7	1.00
Utility 8	91	13.2	33	0.36
Utility 11	45	13.8	22	0.49
Total	**169**		**79**	0.47

Table 3-11
Distribution of "Statistically Valid" Monitors Among Participating Utilities (Voltage Cluster is Between 15-26.4kV)

Utility ID	Nh1	Distribution Level	Controlled Allocation	Sampling Probabilities
Utility 4	810	23	265	0.33
Utility 5	2	23	2	1.00
Utility 6	3	25	3	1.00
Utility 7	7	21	7	1.00
Utility 8	24	26.4	24	1.00
Utility 12	110	26	74	0.67
Total	**956**		**375**	0.39

Table 3-12
Distribution of "Statistically Valid" Monitors Among Participating Utilities (Voltage Cluster is Between 26.4-34.5kV)

Utility ID	Nh1	Distribution Level	Controlled Allocation	Sampling Probabilities
Utility 5	31	34.5	3	0.10
Utility 11	19	34.5	3	0.16
Total	**50**		**6**	0.12

Table 3-13
Distribution of "Statistically Valid" Monitors Among Participating Utilities (Voltage Cluster is >34.5kV)

Utility ID	Nh1	Transmission Level	Controlled Allocation	Sampling Probabilities
Utility 1	13	69	2	0.15
Utility 1	30	345	2	0.07
Utility 1	137	138	8	0.06
Utility 2	3	400	1	0.33
Utility 2	6	123	1	0.17
Utility 3	3	115	1	0.33
Utility 7	1	115	1	1.00
Utility 8	3	230	1	0.33
Utility 8	4	69	1	0.25
Utility 8	5	138	1	0.20
Utility 10	3	115	1	0.33
Utility 10	20	69	2	0.10
Utility 12	14	69	2	0.14
Utility 12	36	500	3	0.08
Total	**278**		**27**	0.10

Sampling Weight Factors and Selection Probability Computation

Out of a total of 1,861 monitors, 629 monitors were selected during the first stage of the site selection process. To account for over-sampling during successive stages, the chances of selecting a monitor for individual utilities (based on the voltage level at which they are installed) can be determined using sampling weight factors. Sampling weighting factors are calculated as reciprocal of the selected probabilities. Equation 3-2 will be used to calculate the basic sampling weight factor at individual stages of site selection.

$$W_F(n) = \frac{1}{(p_1 p_2 \ldots p_n)}$$ **Equation 3-2**

where $p_1 \ldots p_n$ at the selection probabilities during the "n" stages of site selection

The probability (p) of selecting any monitor out of the fraction of the population (as demonstrated in the second column of Table 3-9 through Table 3-13), based on different voltage clusters during the first stage of sample design, is provided in the fifth column of Table 3-9 through Table 3-13. These factors, at successive sampling stages, provide a means to extrapolate the data from a few "statistically valid" monitors to estimate baselines of power quality at a national level. As an example, using controlled allocation, 35 out of 132 monitors placed at a distribution level of 12.5kV for Utility 1 would have to be selected. If 35 monitors out of 132 were selected for the next stage, one would set p_1 equal to (35/132), or 0.27. Therefore, there was a 27% chance that any monitor out of the 35 would have been picked out of the 132.

Second Stage in Sample Design – Voltage Levels and Lightning Flash Density are the Site Descriptors Used

Instead of asking participating utilities to accumulate detailed information on all 629 monitors selected during the first stage of sampling, which would be an unnecessary burden, the next stage of the site selection process was performed using lightning flash density and voltage levels as the stratification variables. To identify the lightning exposure at the monitor locations--along with providing information concerning the number of monitors located at different voltage levels--utilities were asked to provide a subjective breakdown of the number of monitors that represent a location with the following lightning density characteristics: 1) High Lightning Density (>8 flashes/sq km/year), 2) Medium Lightning Density (>3 and <8 flashes/sq km/year), and 3) Low Lightning Density (<3 flashes/sq km/year).

For the majority of the participating utilities who failed to provide information that identified their monitors based on AOV estimation to evaluate lightning flash density, it is then assumed that all of their monitors have the general lightning density based on the geographical area of their service territory (as shown in Figure 3-7).

Stratification techniques were again applied to the 15 combinations among these variables. Table 3-14 illustrates the allocation of monitors among 10 combinations that were used as inputs based on the data provided by the participating utilities. Five combinations had no monitors corresponding to that cell and were not used during stratification.

Mean and standard deviation were evaluated using proxy data from the DPQ study for the distribution system, and proxy data from the SEMI F47 study [9] for the transmission system. The data is shown in Table 3-15.

Table 3-14
Data Used for the Second Stage of Site Selection

Flash Density Voltage Level	F1 LOW	F2 MEDIUM	F3 HIGH
V1 (0-12.5Kv)	69	73	
V2 (12.5-15kV)		62	17
V3 (15-26.4kV)		301	74
V4 (26.4-34.5kV)		6	
V5 (>34.5kV)	7	12	8

Table 3-15
Sampling Distribution of Monitors for Population 629*

Combinations	Strata	Number of Monitors (Population N=629)	μ	σ
1	V1&F1	69	9.05	13.62
2	V1&F2	73	15.47	16.70
3	V2&F2	62	22.35	19.36
4	V2&F3	17	24.31	21.96
5	V3&F2	301	25.01	17.69
6	V3&F3	74	26.97	20.29
7	V4&F2	6	14.45	11.43
8	V5&F1	7	6.59	7.86
9	V5&F2	12	13.00	10.94
10	V5&F3	8	14.97	13.54

*Example Interpretation: When the Voltage Level is Between 0 and 12.5kV and Flash Density is Low, the Average Number of Sags Below 70% is 9.05 Per Year

Determining Adequate Sample Size – Neyman/Optimal Allocation

Figure 3-10 illustrates the graphical variations in the optimal sample size and the precision levels for three different confidence levels using Neyman stratified random sampling techniques. These values were derived using the generic form shown in Equation (A-2) and Equation (A-6). The total population in this case was 629.

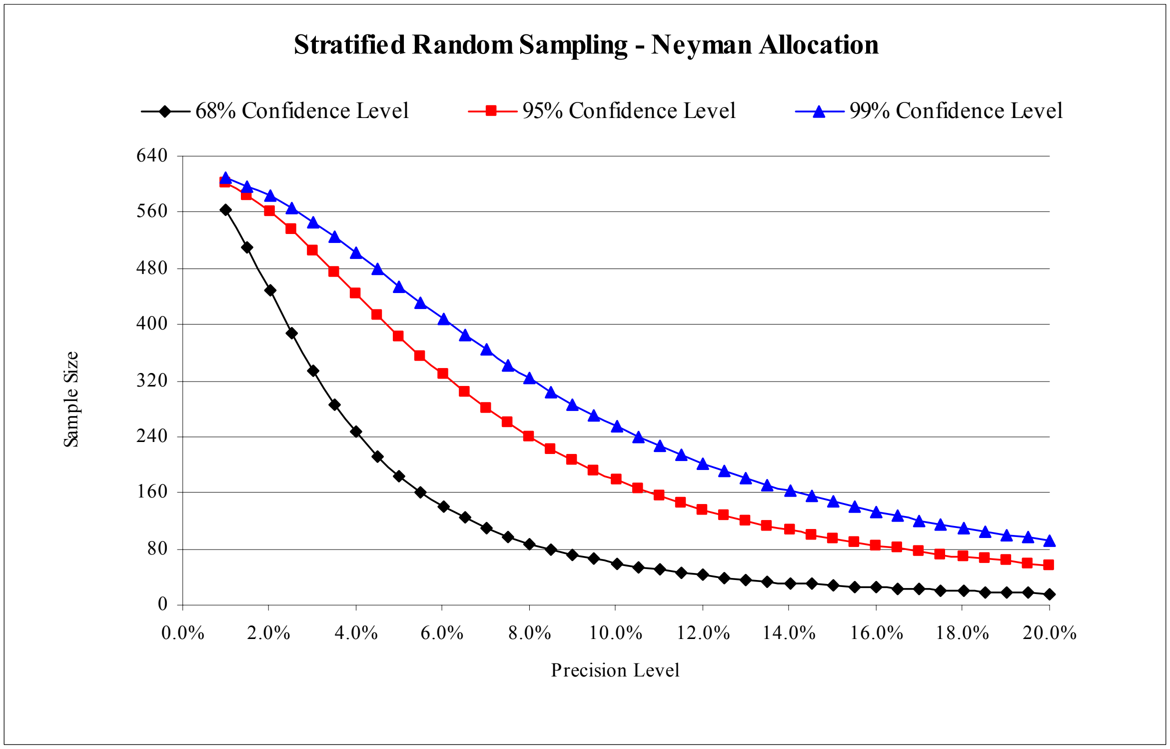


Figure 3-10
Variations in Optimal Sample Sizes Versus Precision Levels (Total Population N = 629) – Neyman Allocation Based Stratified Random Sampling

Table 3-16
Distribution of Samples Inside the Individual Categories (Confidence Level = 95%)

Combinations		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Precision (%)	**Total**	n_{h1}	n_{h2}	n_{h3}	n_{h4}	n_{h5}	n_{h6}	n_{h7}	n_{h8}	n_{h9}	n_{h10}
2.00%	562	48	63	62	19	274	77	4	3	7	6
2.50%	535	46	60	59	18	261	74	3	3	6	5
3.00%	506	44	56	56	17	246	70	3	3	6	5
3.50%	475	41	53	52	16	231	65	3	2	6	5
4.00%	443	38	49	49	15	216	61	3	2	5	4
4.50%	412	35	46	45	14	201	57	3	2	5	4
5.00%	384	33	43	42	13	187	53	2	2	5	4
5.50%	354	31	40	39	12	173	49	2	2	4	4
6.00%	328	28	37	36	11	160	45	2	2	4	3

Distribution of Individual Utility Monitors Using Proportional Allocation Scheme

Once the optimal number of sample sizes for the individual clusters or combinations has been accurately identified, as shown in Table 3-16, the next step in the sample plan involves distributing them within individual participating utilities. Even though a controlled allocation scheme could have been used, the proportional allocation scheme will be used to distribute monitors within participating utilities during this stage as well as for the later stages. The distribution of “statistically valid” monitors for the different utilities based on different voltage and lightning flash density clusters and their selection probabilities are provided in Table 3-17 throughTable 3-26.

Table 3-17
Distribution of “Statistically Valid” Monitors Among Participating Utilities*

Utility ID	N_{h1}	Distribution Level	Proportional Allocation	Sampling Probabilities
Utility 1	35	12.5	17	0.49
Utility 2	1	0.38	1	1.00
Utility 2	4	11	2	0.50
Utility 4	0	12	0	
Utility 6	0	12	0	
Utility 7	0	12	0	
Utility 8	0	4.16	0	
Utility 9	3	12	1	0.33
Utility 9	15	12	7	0.47
Utility 10	11	12.47	5	0.45
Utility 11	0	4.16	0	
Total	69		33	0.48

*Cluster is When Voltage Level is between 0-12.5kV and Flash Density is Low

Table 3-18
Distribution of "Statistically Valid" Monitors Among Participating Utilities*

Utility ID	N_{h1}	Distribution Level	Proportional Allocation	Sampling Probabilities
Utility 1	0	12.5	0	
Utility 2	0	0.38	0	
Utility 2	0	11	0	
Utility 4	49	12	28	0.57
Utility 6	12	12	7	0.58
Utility 7	5	12	3	0.60
Utility 8	6	4.16	4	0.67
Utility 9	0	12	0	
Utility 9	0	12	0	
Utility 10	0	12.47	0	
Utility 11	1	4.16	1	1.00
Total	**73**		**43**	0.59

*Cluster is When Voltage Level is between 0-12.5kV and Flash Density is Medium

Table 3-19
Distribution of "Statistically Valid" Monitors Among Participating Utilities*

Utility ID	Nh1	Distribution Level	Proportional Allocation	Sampling Probabilities
Utility 3	0	13.8	0	
Utility 5	7	13.2	5	0.71
Utility 8	33	13.2	22	0.67
Utility 11	22	13.8	15	0.68
Total	**62**		**42**	0.68

Cluster is When Voltage Level is between 12.5-15kV and Flash Density is Medium

Table 3-20
Distribution of "Statistically Valid" Monitors Among Participating Utilities*

Utility ID	Nh1	Distribution Level	Proportional Allocation	Sampling Probabilities
Utility 3	17	13.8	13	0.76
Utility 5	0	13.2	0	
Utility 8	0	13.2	0	
Utility 11	0	13.8	0	
Total	**17**		**13**	0.76

*Cluster is When Voltage Level is between 12.5-15kV and Flash Density is High

Table 3-21
Distribution of "Statistically Valid" Monitors Among Participating Utilities*

Utility ID	Nh1	Distribution Level	Controlled Allocation	Sampling Probabilities
Utility 4	265	23	165	0.62
Utility 5	2	23	1	0.50
Utility 6	3	25	2	0.67
Utility 7	7	21	4	0.57
Utility 8	24	26.4	15	0.63
Utility 12	0	26	0	
Total	**301**		**187**	0.62

*Cluster is When Voltage Level is between 15-26.4kV and Flash Density is Medium

Table 3-22
Distribution of "Statistically Valid" Monitors Among Participating Utilities*

Utility ID	Nh1	Distribution Level	Controlled Allocation	Sampling Probabilities
Utility 4	0	23	0	
Utility 5	0	23	0	
Utility 6	0	25	0	
Utility 7	0	21	0	
Utility 8	0	26.4	0	
Utility 12	74	26	53	0.72
Total	**74**		**53**	0.72

*Cluster is When Voltage Level is between 15-26.4kV and Flash Density is High

Table 3-23
Distribution of "Statistically Valid" Monitors Among Participating Utilities*

Utility ID	Nh1	Distribution Level	Controlled Allocation	Sampling Probabilities
Utility 5	3	34.5	1	0.33
Utility 11	3	34.5	1	0.33
Total	**6**		**2**	0.33

*Cluster is When Voltage Level is Between 26.4-34.5kV and Flash Density is Medium

Table 3-24
Distribution of "Statistically Valid" Monitors Among Participating Utilities*

Utility ID	Nh1	Transmission Level	Controlled Allocation	Sampling Probabilities
Utility 1	2	69	1	0.50
Utility 1	0	345	0	
Utility 1	0	138	0	
Utility 2	1	400	0	
Utility 2	1	123	0	
Utility 3	0	115	0	
Utility 7	0	115	0	
Utility 8	0	230	0	
Utility 8	0	69	0	
Utility 8	0	138	0	
Utility 10	1	115	0	
Utility 10	2	69	1	0.50
Utility 12	0	69	0	
Utility 12	0	500	0	
Total	**7**		**2**	0.29

*Cluster is When Voltage Level is >34.5kV and Flash Density is Low

Table 3-25
Distribution of "Statistically Valid" Monitors Among Participating Utilities*

Utility ID	Nh1	Transmission Level	Controlled Allocation	Sampling Probabilities
Utility 1	0	69	0	
Utility 1	0	345	0	
Utility 1	8	138	**5**	0.63
Utility 2	0	400	0	
Utility 2	0	123	0	
Utility 3	0	115	0	
Utility 7	1	115	0	
Utility 8	1	230	0	
Utility 8	1	69	0	
Utility 8	1	138	0	
Utility 10	0	115	0	
Utility 10	0	69	0	
Utility 12	0	69	0	
Utility 12	0	500	0	
Total	**12**		**5**	0.42

*Cluster is When Voltage Level is >34.5kV and Flash Density is Medium

Table 3-26
Distribution of "Statistically Valid" Monitors Among Participating Utilities*

Utility ID	Nh1	Transmission Level	Controlled Allocation	Sampling Probabilities
Utility 1	0	69	0	
Utility 1	2	345	1	0.50
Utility 1	0	138	0	
Utility 2	0	400	0	
Utility 2	0	123	0	
Utility 3	1	115	1	1.00
Utility 7	0	115	0	
Utility 8	0	230	0	
Utility 8	0	69	0	
Utility 8	0	138	0	
Utility 10	0	115	0	
Utility 10	0	69	0	
Utility 12	2	69	1	0.50
Utility 12	3	500	**1**	0.33
Total	**8**		**4**	0.50

*Cluster is When Voltage Level is >34.5kV and Flash Density is High

Of a total of 629 monitors, 384 were selected during the second stage of site selection process. The probability of selecting any monitor out of the fraction of the population (as demonstrated in the second column of Table 3-17 through Table 3-26) based on different voltage levels and lightning flash density clusters during the second stage of sample design are provided in the fifth column of Table 3-17 through 3-26. As an example, the selection probability to choose monitors placed at a distribution level of 12.5kV for Utility 1 having low lightning flash density would then be the product of the two stages (P_1P_2), and is equal to (35/132)*(17/35) or 12.87%. For this example, we could say that there is a 1 in 8 chance that we would have selected a monitor placed at a distribution level of 12.5kV for Utility 1 having a low lightning flash density.

Third Stage in Sample Design – Voltage Levels, Lightning Flash Density, Feeder Characteristics, and Type of Service Systems are the Site Descriptors Used

Once the initial two stages in the overall site selection process have been completed, and manageably smaller "statistically valid" monitoring locations across the United States have been accurately identified, more detailed data from these selected 384 sites will then have to be requested. This section assumes that if sufficient data were obtained and if two other site descriptors, namely feeder characteristics (mostly overhead, mostly underground, mixed) and types of service systems (radial, loop, mixed) were identified as the important power quality site descriptors, then the stratification techniques could again be applied. Data from these "statistically valid" monitors would then serve to establish baselines in power quality levels at a national level.

Table 3-27 illustrates the allocation of monitors among 23 combinations between the site descriptors during the final stages of site selection. Thirty-seven combinations or bins did not have any monitors corresponding to that cell and were not used during stratification. Notations used in Table 3-27 are as follow:

- V1LOW – voltage level between 0-12.5kV and low flash density
- V1MEDIUM – voltage level between 0-12.5kV and medium flash density
- V2MEDIUM – voltage level between 12.5-15kV and medium flash density
- V2HIGH – voltage level between 12.5-15kV and high flash density
- V3MEDIUM – voltage level between 15-26.4kV and medium flash density
- V3HIGH – voltage level between 15-26.4kV and high flash density
- V4MEDIUM – voltage level between 26.4-34.5kV and medium flash density
- V5LOW – voltage level >34.5kV and low flash density
- V5 MEDIUM – voltage level >34.5kV and medium flash density
- V5HIGH – voltage level >34.5kV and high flash density
- OVRD – feeder characteristics are mostly overhead and have feeder configurations that are radial
- OVLP – feeder characteristics are mostly overhead and have feeder configurations that are loops
- UGRD – feeder characteristics are mostly underground and have feeder configurations that are radial
- UGLP – feeder characteristics are mostly underground and have feeder configurations that are loops
- MXRD – feeder characteristics are mixed (partially overhead and partially underground) and have feeder configurations that are radial
- MXLP – feeder characteristics are mixed (partially overhead and partially underground) and have feeder configurations that are loops

Table 3-27
Monitor Placement

	OVRD	OVLP	UGRD	UGLP	MXRD	MXLP
VILOW	6		24	3		
VIMEDIUM	11		3			29
V2MEDIUM	7	28	2	5		
V2HIGH	9		2		2	
V3MEDIUM	68	15			104	
V3HIGH	50		3			
V4MEDIUM	1				1	
V5LOW	2					
V5MEDIUM	5					
V5HIGH	4					

Mean and standard deviation for these 23 combinations were evaluated using proxy data from the DPQ study for the distribution system, and proxy data from the SEMI F47 study [9] for the transmission system. The data is shown in Table 3-28.

Table 3-28
Sampling Distribution of Monitors for Population 629*

Combinations	Strata	Number of Monitors (Population N=384)	μ	σ
1	V1LOW_OVRD	6	13.53	15.71
2	V1LOW_UGRD	24	10.53	13.31
3	VILOW_UGLP	3	12.53	8.31
4	VIMEDIUM_OVRD	11	16.73	17.25
5	VIMEDIUM_UGRD	3	13.73	14.85
6	V1MEDIUM_MXLP	29	17.23	10.83
7	V2MEDIUM_OVRD	7	20.17	18.58
8	V2MEDIUM_OVLP	28	22.17	13.14
9	V2MEDIUM_UGRD	2	17.17	16.18
10	V2MEDIUM_UGLP	5	19.17	11.18
11	V2HIGH_OVRD	9	21.16	19.88
12	V2HIGH_UGRD	2	18.16	17.48
13	V2HIGH_MXRD	2	19.66	18.68
14	V3MEDIUM_OVRD	68	21.50	17.74
15	V3_MEDIUM_OVLP	15	23.50	12.30
16	V3MEDIUM_MXRD	104	20.00	16.54
17	V3HIGH_OVRD	50	22.49	19.05
18	V3HIGH_UGRD	3	19.49	16.65
19	V4MEDIUM_OVRD	1	16.22	14.62
20	V4MEDIUM_MXRD	1	14.72	13.42
21	V5LOW_OVRD	2	12.29	12.83
22	V5MEDIUM_OVRD	5	15.50	14.37
23	V5HIGH_OVRD	4	16.48	15.67

*Example Interpretation: When the Voltage Level is Between 0 and 12.5Kv, Flash Density is Low, Feeder Characteristics are Mostly Overhead, and Have Feeder Configurations that are Radial, the Average Number of Sags Below 70% is 13.53 Per Year

Determining Adequate Sample Size – Neyman/Optimal Allocation

Figure 3-11 illustrates the graphical variations in the optimal sample size and the precision levels for three different confidence levels using Neyman stratified random sampling techniques. These values were derived using the generic form shown in Equation (A-2) and Equation (A-6), and are provided in Table 3-29. The total population in this case was 384. However, the distribution of individual distribution monitors is not shown for the sake of brevity.

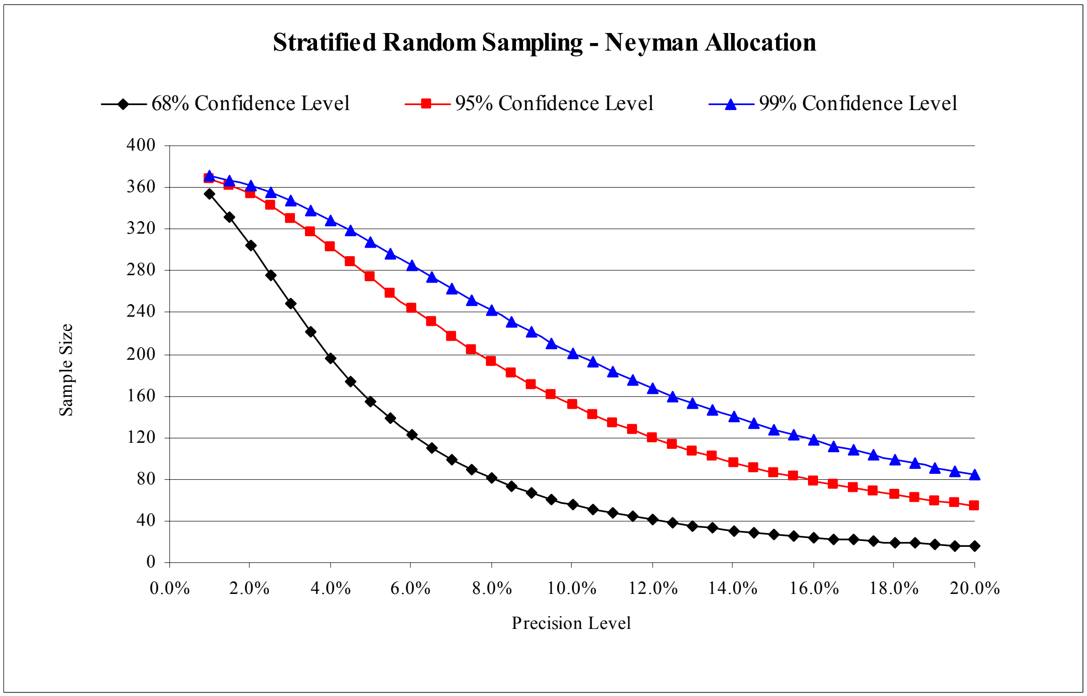


Figure 3-11
Variations in Optimal Sample Sizes Versus Precision Levels (Total Population N = 384) – Neyman Allocation Based Stratified Random Sampling

Table 3-29
Allocation of Samples in a Strata (Total Population N = 384) – Optimal Allocation Based Stratified Random Sampling

	Confidence Level 68%	**Confidence Level 95%**	**Confidence Level 99%**
Precision (%)	**Sample Size (n)**	**Sample Size (n)**	**Sample Size (n)**
2.00%	305	353	362
2.50%	276	342	355
3.00%	248	330	347
3.50%	221	317	338
4.00%	196	303	329
4.50%	174	288	319
5.00%	155	273	308
5.50%	138	259	297
6.00%	123	244	286

One of the initial study objectives was to develop a methodology for the site selection scheme so that limited sets of "statistically valid" monitors could be used to establish baselines of power quality at a national level. Statistical validity of extended results to the national level cannot be strictly defended because the study relied on only 12 participating utilities. However, statistically valid results were obtained for this group of participating utilities, and a comparison of this group with the nation showed that they covered a wide range of factors that affect power quality at a national level. The ranges were typical of the transmission and distribution systems present nationwide. As a result, the limited sets of monitors that are obtained as the final outcome of the overall site selection process will provide a good indication of power quality levels that can be expected over the rest of the nation.

Bibliography

[1] *Understanding Premium Power Grades: Final Report*, EPRI, Palo Alto, CA: 2000: EPRI TR-1000406.

[2] *An Assessment of Distribution System Power Quality*: *Volumes 1-3*, EPRI, Palo Alto, CA: TR-106294-V1, TR-106294-V2, TR106294-V3, 1996.

[3] *Review of CLP Power System Quality (PQ) of Voltage Dips*, EPRI, Palo Alto, CA: 2001: EPRI contract number-4500017036.

[4] Huffines, Gary R.; Richard E. Orville, "*Lightning Ground Flash Density and Thunderstorm Duration in the Continental United States: 1989-1996*," Journal of Applied Meteorology, Vol. 38, July 1999, pp. 1013-1019.

[5] Levy, S. Paul; Stanley Lemeshow, "*Sampling of Populations: Methods and Applications,"* 3rd edition, John Wiley & Sons Inc., New York, 1999.

[6] Huang S. R., "*Effectiveness of Optimum Stratified Sampling And Estimation in Monte Carlo Production Simulation,"* IEEE Transaction on Power Systems, Vol. 12, No. 2, May 1997, pp. 566-572.

[7] Cochran W. G., "*Sampling Techniques,*" 3rd edition, John Wiley & Sons, Inc., New York, 1977.

[8] Hess, I.; V.K. Sethi; and T.R. Balakrishnan, "*Stratification: A Practical Investigation*," Journal of the American Statistical Association, pp. 71-90.

[9] Short, Tom; Mansoor, Arshad; Sunderman; Wes and Sundaram, Ashok, "*Site Variation and Prediction of Power Quality*," unpublished paper, IEEE Transaction of Power Delivery, 2001.

4

VOLTAGE SAG DATA ANALYSIS AND PRESENTATION METHODOLOGY FOR DPQ PHASE II

A Need for Consistent Data Analysis and Presentation Technique

The primary power quality variables of interest for the DPQ Phase II project are voltage sags and momentary interruptions. These two power quality events are considered to be the most important power quality variation for end-use customers. Although these events are relatively infrequent, they are important because of their tremendous economic impacts to the end user. Voltage sag data can be analyzed and presented using several different methods as described in [1, 2, 3, 4]. For the purpose of the DPQ Phase II project, the project team chose three methods of analysis and presentation of voltage sag data. These three methods are:

1. The EPRI Reliability Benchmarking Indices - specifically the System Average Root Mean Square Frequency Variation (SARFI) index

2. The voltage sag magnitude and duration table developed by UNIPEDE[1] and later adopted in IEC 61000-2-8, Voltage Dips and Short Interruptions on Public Electric Power Supply Systems with Statistical Measurement Results

3. The voltage sag coordination chart recommended in IEEE 1346, Recommended Practice for Evaluating Electric Power System Compatibility with Electronic Process Equipment

In this chapter, the three methodologies that will be used for data analysis and presentation for the DPQ Phase II project are presented.

EPRI Reliability Benchmarking (RBM) Indices

The RBM technical report, Reliability Benchmarking Methodology (EPRI TR-107938) [1], defines indices for short-duration RMS voltage variations such as voltage sags and momentary interruptions. The RBM indices have provided a common basis and terminology for assessing and discussing service quality. These indices are currently used by several American utilities for reporting power quality performance both internally and externally.

---

[1] The International Union of Producers and Distributors of Electrical Energy, a major association representing European electric power companies.

The most basic RBM index for voltage sag performance is the System Average RMS (Variation) Frequency Index$_{\text{Voltage}}$ or SARFI$_X$. The SARFI$_X$ concept is the basis for most of the other RBM indices as well. SARFI$_X$ represents the average number of specified short-duration RMS variation measurement events that occurred over the monitoring period per customer served from the assessed system. For SARFI$_X$, the specified disturbances are those RMS variations with a voltage magnitude less than X for voltage drops or a magnitude greater than X for voltage increases. SARFI$_X$ is defined by Equation 4-1.

$$SARFI_x = \frac{\sum N_i}{N_T}$$

Equation 4-1

where;

$X \equiv$ RMS voltage threshold; any positive value is possible. However, some of the more common values include 140, 120, 110, 90, 80, 70, 50, and 10

$N_i \equiv$ Number of customers experiencing voltage deviations with magnitudes above X% for X>100 or below X% for X<100 due to measurement event i

$N_T \equiv$ Total number of customers served from the section of the system to be assessed

The original intent for the SARFI$_X$ calculation is similar to the calculation of the System Average Interruption Frequency Index (SAIFI) value that many utilities have calculated for years for reliability purposes. However, the SARFI index has evolved over the years from a system-wide index to a more site-specific index where the SARFI$_X$ equation reduces to a count of the number of sags that have a magnitude below the specified RMS voltage threshold (X).

The five RBM SARFI indices that will be evaluated as part of the DPQ Phase II project are, SARFI$_{70}$, SARFI$_{50}$, SARFI$_{10}$, SARFI$_{\text{ITIC}}$, and SARFI$_{\text{SEMI}}$. The rationale for choosing these five SARFI indices is as follows:

SARFI$_{70}$ – These are sags where the lowest retained voltage was less than 70% of the nominal voltage at the evaluated location. This sensitivity level corresponds to the level specified in the ITIC curve and in SEMI 2844 for events that last longer than 200 milliseconds. When characterizing system performance with a single index, SARFI$_{70}$ is the preferred index.

SARFI$_{50}$ – These are sags where the lowest retained voltage was less than 50% of the nominal voltage at the evaluated location. This level is important because the SEMI F47 standard "Specification for Semiconductor Process Equipment Voltage Sag Immunity" has specified that semiconductor manufacturing equipment should be able to ride through events with retained voltage down to 50% for up to 200 milliseconds (12 cycles on a 60 Hertz base). This is an important level for future characterizing of system performance.

SARFI$_{10}$ – This provides an estimate of momentary interruption performance. Voltage sags with a retained voltage less than 10% are defined as interruptions.

SARFI$_{ITIC}$ – This provides an estimate of the number of voltage sags outside the equipment voltage sag tolerance developed by the Information Technology Industry Council (ITIC) and is shown in Figure 4-1.

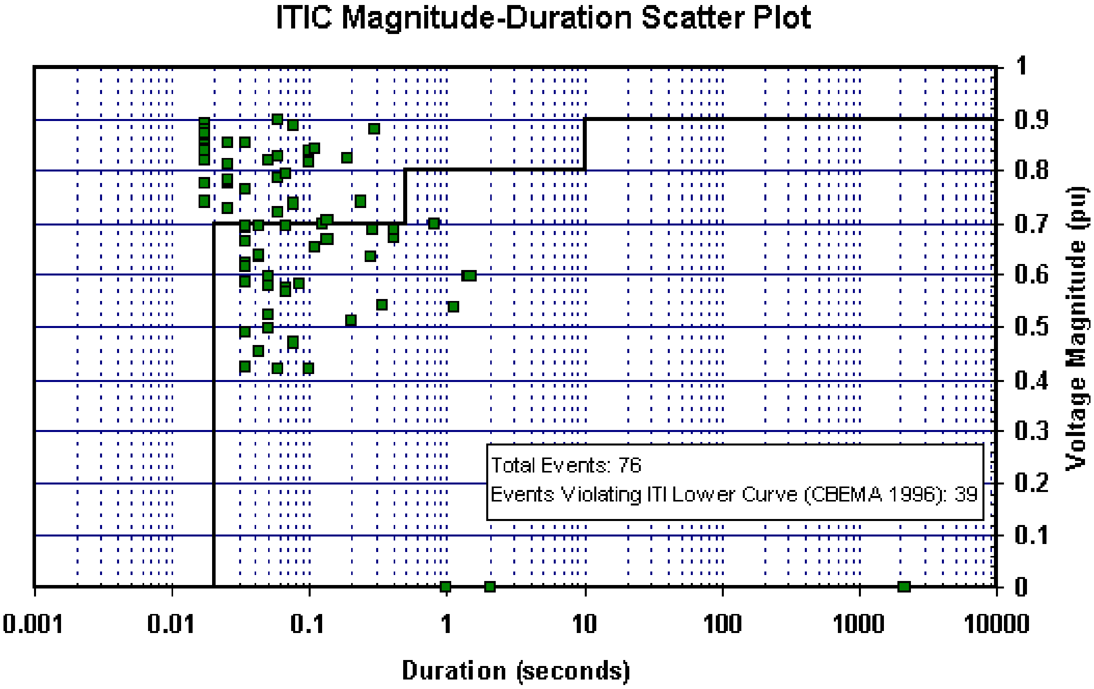


Figure 4-1
Example Comparison of Voltage Sag Data with ITIC Tolerance Curves

SARFI$_{SEMI}$ – This provides an estimate of the number of voltage sags outside the equipment voltage sag tolerance (See Figure 4-2) developed by the Semiconductor Equipment and Materials International (SEMI) organization as part of the SEMI F47 standard.

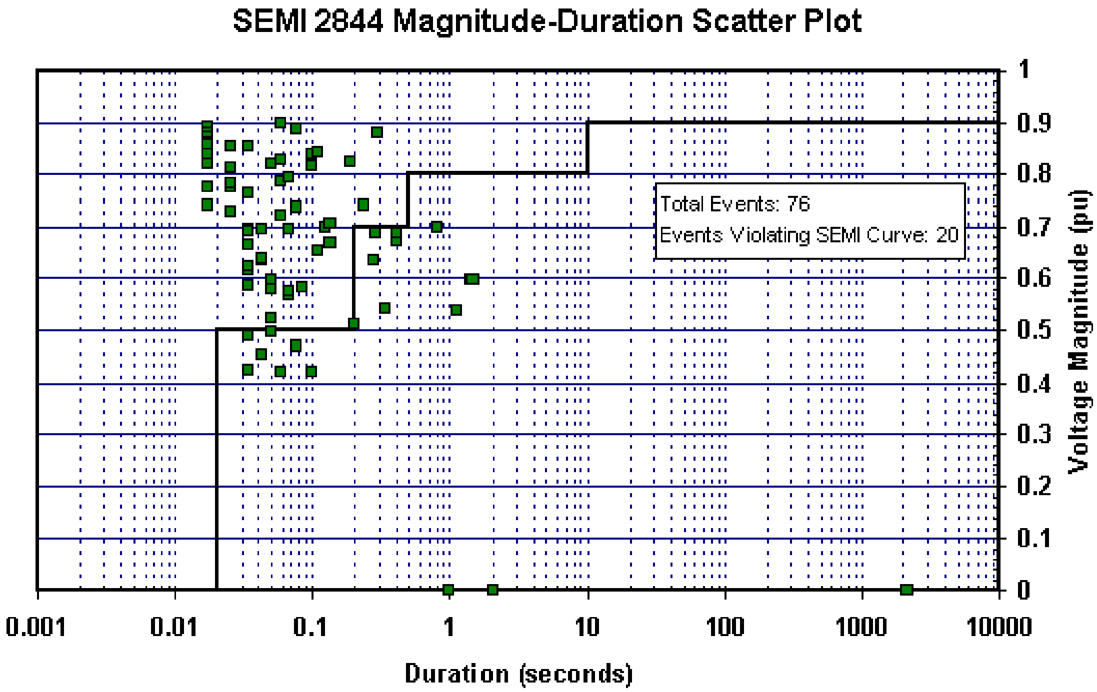


Figure 4-2
Example Comparison of Voltage Sag Data with SEMI F47 Tolerance Curves

IEC 61000-2-8 Voltage Sag Magnitude and Duration Table

A simplified way of quantifying the number of sags is through a table with magnitude and duration ranges often known as the sag density table. Table 4-1 shows the magnitude and duration bins of the sag density table that were adopted by IEC in the IEC 61000-2-8 document. This table is based on the UNIPEDE table with a major change in the last row that describes momentary interruptions. Originally, in the UNIPEDE table, voltage less than 1% was classified as momentary interruptions. However, realizing that motors and other loads most likely will support voltage higher than 1% even during a momentary interruption, this range was changed to 10% which agrees with the *IEEE Standard 1159-1995, Recommended Practice for Monitoring Electric Power Quality*.

Table 4-1
Example of Sag Density Table Based on IEC 61000-2-8 Requirement

	Duration (seconds)							
	0.01 < t ≤ 0.02	0.02 < t ≤ 0.1	0.1 < t ≤ 0.5	0.5 < t ≤ 1	1 < t ≤ 3	3 < t ≤ 20	20 < t ≤ 60	60 < t ≤ 180
90 > V ≥ 85	8	4	2	1	0.5	0.2	0.1	0
85 > V ≥ 70	6	4	3	2	1.5	2	1	0
70 > V ≥ 40	4	3	2.8	2	1	0.5	0	0
40 > V ≥ 10	3	2	1	0.8	0.6	0.4	0.2	0
10 > V ≥ 0	2	1	0.7	0.4	0	0	0	0

NOTE 1: Measurement results in the first column and first row are likely to be inflated by transients and load fluctuations, respectively.

NOTE 2: 0.01 and 0.02 s in the first two duration headings correspond to a half and one period of the 50 Hz voltage. For 60 Hz systems, corresponding values would be used.

Another variation in presenting voltage sag data in a magnitude duration table is to use cumulative numbers instead of numbers for each cell. Often, customers are not so much interested in the number of sags in a given magnitude and duration range, but rather the number of times voltage sag will be worse than a certain magnitude and duration. Each element *MD* of the cumulative sag table is defined in Equation 4-2.

$$F_{MD} = \sum_{m=0}^{M} \sum_{d=D}^{d_{\max}} f_{md}$$ **Equation 4-2**

where,

f_{md} = element *md* of the density table: the number of sags in the duration range *d* and the magnitude range *m*; and

F_{MD} = element *MD* of the cumulative table; the number of sags with duration longer than *D* and magnitude less than *M*.

The cumulative table obtained from the sag density table in Table 4-1 is shown in Table 4-2. For example, based on Table 4-2, a customer is likely to experience 25.4 sags with a magnitude below 70% of nominal and with a duration greater than 0.01 seconds.

Table 4-2
Example of Cumulative Sag Density Table Based on Table 4-1

	Duration (seconds)							
	0.01	0.02	0.1	0.5	1	3	20	60
90%	60.7	37.7	23.7	14.2	8	4.4	1.3	0
85%	44.9	29.9	19.9	12.4	7.2	4.1	1.2	0
70%	25.4	16.4	10.4	5.9	2.7	1.1	0.2	0
40%	12.1	7.1	4.1	2.4	1.2	0.6	0.2	0
10%	4.1	2.1	1.1	0.4	0	0	0	0

NOTE 1: Measurement results in the first column and first row are likely to be inflated by transients and load fluctuations, respectively.

NOTE 2: 0.01 and 0.02 s in the first two duration headings correspond to a half and one period of the 50 Hz voltage. For 60 Hz systems corresponding values would be used.

IEEE 1346 Voltage Sag Coordination Contour

The contour chart is recommended as a “voltage sag coordination chart” in IEEE 1346. In the voltage sag coordination chart, the contour chart defining the electrical environment voltage sag rate is superimposed with an equipment voltage sag tolerance curve to estimate the number of times the equipment will be impacted in the given electrical environment.

The contour chart starts with a voltage sag density table as shown in Table 4-1. From the voltage sag density table, a cumulative number for each bin signifying the number of voltage sags worse than the magnitude and duration of that bin has to be developed as shown in Table 4-2. The values shown in Table 4-2 can be seen as a two-dimensional function of the number of voltage sags versus magnitude and duration. A contour plot is a visual representation of this magnitude and duration plane. Figure 4-3 shows this plane derived from the data in Table 4-2.

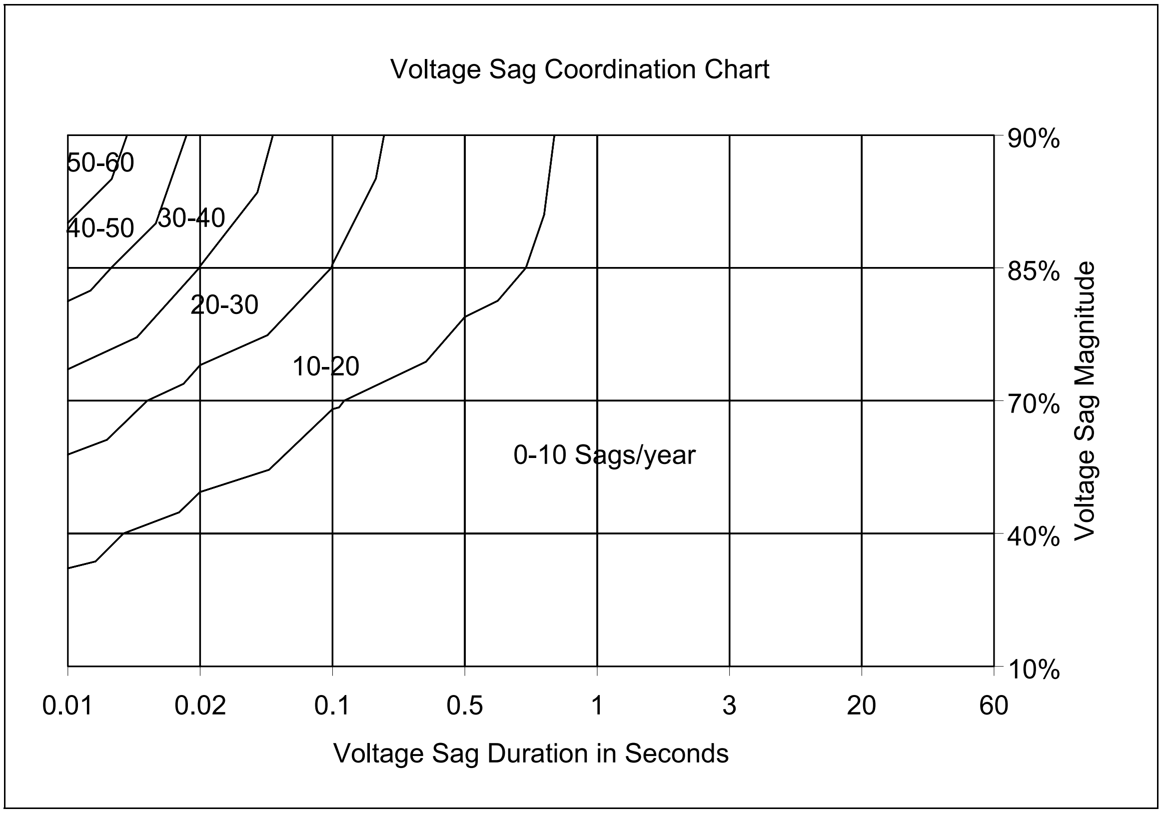


Figure 4-3
Voltage Sag Coordination Contour Plot

Figure 4-4 illustrates the superposition of the equipment sensitivity curve on the contour curve to identify the number of times equipment will be impacted in the electrical environment. For example, based on Figure 4-4, Equipment “A” will likely be affected 30 times a year, whereas Equipment “B” will be impacted 10 times a year because it is less sensitive than equipment “A.”

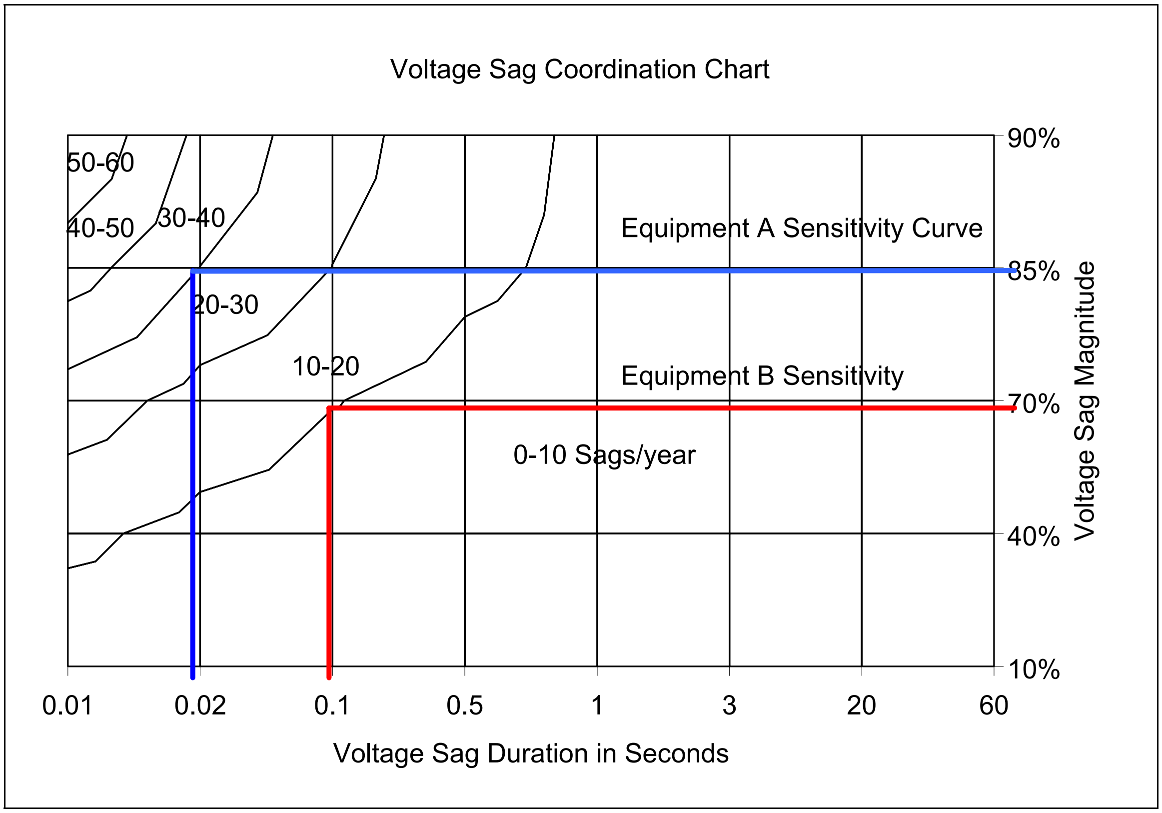


Figure 4-4
Equipment Sensitivity Curve Superimposed on Voltage Sag Contour Plot

Statistical Parameters for Multiple Site Sag Analysis

When voltage sag data from several sites is to be aggregated, as is the case for the DPQ Phase II project, suitable statistical parameters have to be calculated to aggregate individual site measurement data. For DPQ Phase II, the 95th percentile, 50th percentile, and the arithmetic average will be presented for the different RBM indices, voltage sag magnitude and duration tables, and contour plots. Figure 4-5 shows a sample calculation of the three parameters for 20 measurement sites for the SARFI_{70} index.

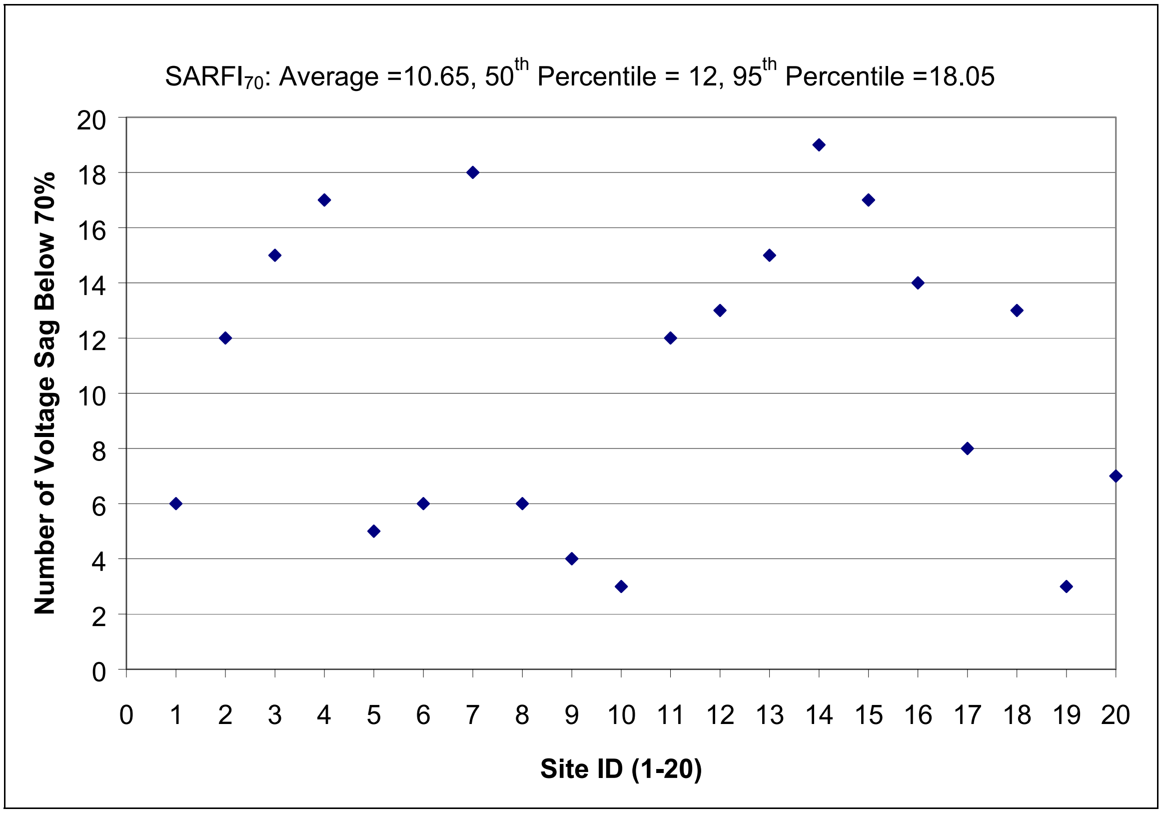


Figure 4-5
Statistical Parameters from Individual Site Data

References

[1] *Reliability Benchmarking Methodology*, EPRI, Palo Alto, CA: 1997. TR-107938

[2] *Reliability Benchmarking Application Guide for Utility/Customer PQ Indices*, EPRI, Palo Alto, CA: 1999. TR-100761

[3] IEC 61000-2-8, *Voltage Dips and Short Interruptions on Public Electric Power Supply Systems with Statistical Measurement Results*

[4] IEEE 1346, *Recommended Practice for Evaluating Electric Power System Compatibility with Electronic Process Equipment*

A

STRATIFIED RANDOM SAMPLING

Stratification refers to the classification of individual sampling units according to a set(s) of power quality site descriptors or stratification variables. In stratified random sampling, the population is heterogeneous rather than homogeneous. This form of sampling design provides a greater precision for a given sample size as compared to simple random sampling. Another reason for stratification is to control sample size for important subpopulations.

Stratified designs are obtained by segmenting the substation population into "L" mutually exclusive and exhaustive groups, commonly known as sampling cells or strata, and a simple random sample "n_h" elements is taken within each stratum "h" according to some stratifying variable or set of variables. This is illustrated in Equation (A-1), where the subscript "h" represents individual stratum.

$$Total\ population = N = N_{h1} + N_{h2} + N_{h3} + ... + N_{h_L}$$

$$Total\ samples = n = n_1 + n_2 + ... + n_h$$

Equation A-1

The strata are homogeneous categories of a heterogeneous population. If one group is proportionally larger than another, or has a greater variation in standard deviations, its sample size should also be proportionally larger. The optimal number of groups or strata to be considered can be determined by the characteristics of the population and was first presented by Dalenius-Hodges [5-8] who provided procedures to aid in developing the sampling cell boundaries. Gains in precision are achieved by choosing stratification schemes that divide the population into groups with similar power quality. Intuitively, the improved precision results from the fact that within-group variations in the variable of interest are smaller than the overall variation in the population. As a consequence, sample points can be allocated in a more optimal way among the sampling cells to increase precision without changing the overall sample size.

After dividing the population into groups, the researcher can then sample each homogenous group using any of the following techniques:

- Equal Allocation
- Proportional Allocation
- Neyman (Optimal) Allocation
- Controlled Allocation Using a Combination of Equal and Proportional Allocation

A generic formula to determine the optimal sample size required for a given set of precision and confidence levels under any types of stratified random sampling is shown in Equation (A-2). Symbols “ε” and “Z” represent the precision and confidence levels respectively.

$$n \geq \frac{\left(\frac{Z^2}{N^2}\right)\left(\sum_{h=1}^{L} \frac{N_h^2 \sigma_{hx}{}^2}{\pi_h \overline{\mathrm{X}}^2}\right)}{\varepsilon^2 + \left(\frac{Z^2}{N^2}\right)\left(\sum_{h=1}^{L} \frac{N_h \sigma_{hx}{}^2}{\overline{\mathrm{X}}^2}\right)}$$

where **Equation A-2**

$$\pi_h = \frac{n_h}{N}$$

where:

N is the total population;

ε is the precision level;

Z is the confidence level;

N_h is the total number of points (population) in the h^{th} strata;

n_h is the total number of points (sample) in the h^{th} strata;

σ_{hx} is the standard deviation of the variable of interest in the h^{th} strata;

L is the total number of strata;

$\bar{X}$ is the population mean.

Expressions for variance and standard errors are illustrated in Equation (A-3).

$$\mathrm{var}(\overline{\mathrm{X}}\ \mathrm{str}) = \left[\frac{1}{\mathrm{N}^2}\sum_{h=1}^{L}\left(N_h^2\right)\left(\frac{\sigma_{hx}^2}{n_h}\right)\left(\frac{N_h - n_h}{N_h}\right)\right]$$

$$S\mathrm{E}(\bar{\mathrm{x}}\ \mathrm{str}) = \sqrt{\mathrm{var}(\bar{\mathrm{x}}\,str)}$$

Equation A-3

The next step is to obtain the optimal value of “n_h” among the “L” cells from the different allocation schemes mentioned earlier. These are provided in the next section. However, the equal allocation technique is not discussed.

Stratified Equal Designs

In equal allocation, the same number of elements is sampled from each stratum. For each stratum h, the sample sizes are given as shown in Equation (A-4).

$$n_h = \frac{n}{L} \qquad \text{for h=1,...,L} \qquad \textbf{Equation A-4}$$

Stratified Proportional Designs

One relatively simple way to allocate sample points to the various sampling cells is to do so in proportion to the number of substations in the population that reside in that cell. For example, if 12% of the population falls into the first sampling cell, then a researcher using a proportional design would allocate 12% of the total sample points to that cell or stratum. Equation (A-5) shows the allocation formula for a proportional design.

$$n_h = n * \frac{N_h}{N} \qquad \text{for h=1,...,L} \qquad \textbf{Equation A-5}$$

Stratified Neyman Designs

The Neyman allocation is derived by mathematically solving the optimization problem of allocating the total number of samples among the L cells in a way that maximizes precision. This can be obtained by minimizing the precision as shown in Equation (A-2). The solution to this optimization problem is given in Equation (A-6).

$$n_h = n \cdot \left(\frac{N_h \cdot \sigma_h}{\sum_h N_h \cdot \sigma_h} \right) \quad \text{h=1,...,L} \qquad \textbf{Equation A-6}$$

Stratified Controlled Designs

In controlled allocation, the allocation scheme is dividing by providing equal weights to both equal and proportional allocation. The mathematical expression to this optimization problem is given in Equation (A-7).

$$n_h = (\frac{1}{2}) * (\frac{n}{L} + n * \frac{N_h}{N}) \quad \text{for h=1,...,L} \qquad \textbf{Equation A-7}$$

B
UTILITY SURVEY

Introduction

Since the completion of the Distribution Power Quality (DPQ) project in 1995, several utilities have implemented system-wide PQ monitoring programs both in distribution and transmission systems. The wealth of data collected since the DPQ project provides a unique opportunity to synthesize meaningful information regarding variability of grid power, specifically on voltage sag rates, based on system characteristics. It also presents a unique opportunity to compare the results of the DPQ project with the data from these monitoring programs. One of the major efforts of the Premium Power project for 2001 is to determine the types and sizes of monitoring systems employed since the completion of the DPQ project in 1995 and, based on this information, assess:

- Site selection criteria for the data analysis section on Phase II based on this population of existing sites
- Methodology for how the data will be collected, sanitized, and analyzed in Phase II of this project
- Comparison of voltage sag rate (depth and magnitude) to the DPQ data
- Variability of voltage sag rate based on site characteristics
- The RBM Indices such as $SARFI_{70}$, $SARFI_{ITIC}$, and $SARFI_{SEMI}$ based on site characteristics

In an effort to reduce the amount of data that is required from each utility, the first step of this process is to collect a limited amount of data from each utility for their monitoring sites. This information will be used during the site selection process to develop the stratification variables. Once the site selection process has been completed, more detailed data from the sites that have been selected will be requested from the respective utilities.

Site Descriptors

Section 1 – Utility Specific

This section only needs to be filled out once for each utility.

Primary contact name, address, phone number, and email for the utility personnel:

____

____

____

What type of monitoring data do you have? Check all that apply.

Dranetz-BMI: • PASS • PES
PQView
RPM
PML
Square D
Nexus
GE
Other ____

____

How will you deliver the monitoring data to EPRI PEAC?

CD-ROM
Zip Drive
Other ____

____

Within your utility service area, on a scale of 1-10 (1 being not important, 10 being very important), rate the importance of the following variables on voltage sag rate.[2]

Causes	**Lightning**	**Tree Limb**	**Inadequate Ground Resistance**	**Brush/Forest Fire**	**Equipment Failure (Splice, cable, etc.)**	**Animal Contact**	**Dig-In Vehicle**	**Others**
(1-10)								

---

[2] We understand that this is a subjective question, and even within one service area, the causes could vary from one location to another; the purpose of this question is to assess the variables that affect voltage sag rate in a general area.

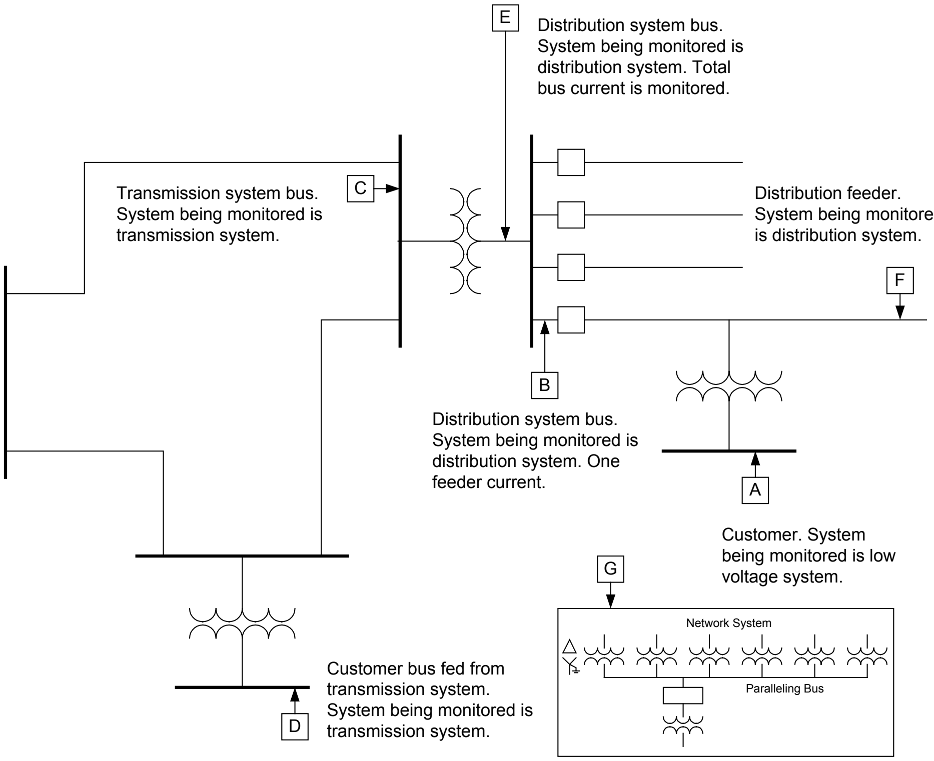
Section 2 - Information on Monitor Location

Utilities have an option of either filling out Section 2A or Section 2B. Section 2A requires information for all monitoring sites to be filled in on one form. Section 2B requires one form to be filled out for each monitor location.

Section 2A – All Monitor Sites

Please fill out this survey for all monitors on your system.

Please review the figure below carefully and fill out your information based on the monitor locations shown in the figure. This figure in general captures all possible monitor locations. However, if you have locations that do not correspond to any one of the options in the figure, please sketch a one-line diagram for your monitor location and fill out the information. *(For simplicity, the distribution system is shown to be radial; however, monitor location A, B, F, or E may represent systems that are loop, primary selective, secondary selective, or spot network.)*



Please identify the number of monitors for each location category.

A – Customer Bus Distribution System

B – Feeder Monitor Distribution System

C – Transmission Bus Transmission System

D – Customer Bus Transmission System

E – Distribution Bus Distribution System

F – Distribution Feeder Distribution System

G – Underground Network

H – Other (explain):

For all monitor locations (except for G), please identify the distribution and transmission voltage levels and the number of monitors at each level. ***(For network system G, please identify the secondary network voltage and primary voltage.)***

# of Monitors	Distribution Voltage Level (kV) Line-Line	Transmission Voltage Level (kV) Line-Line

For all monitor locations (except for C, D, or G), please identify the number of monitors based on the distribution feeder characteristics.

# of Monitors	Distribution Feeder Characteristics
	Mostly Overhead
	Mostly Underground
	Mixed

For all monitor locations (except for C, D, or G), please identify the type of service systems and the number of monitors for each system.

# of Monitors	Type of Service Systems
	Radial
	Primary Loop
	Primary Selective
	Secondary Selective
	Spot Network

For all monitor locations (except for C, D, or G), please identify the number of monitors based on the distribution feeder type and load density of MVA/Mile.

# of Monitors	Distribution Feeder Type	Range of MVA/Mile
	Mostly Urban	
	Mostly Rural	
	Mixed	

For all monitor locations (except for A, C, D, or G), please identify the number of monitors based on customer load characteristics.

# of Monitors	Customer Load Characteristics
	Mostly Industrial
	Mostly Residential
	Mostly Commercial
	Mixed

For all your monitor locations, please identify the number of monitors based on lightning flash density of the area of vulnerability. ***(If you do not have this data, we will assume all of your monitor locations have the general lightning density based on the geographical area of your service territory.)***

# of Monitors	Customer Load Characteristics
	High Lightning Density (≥8 flashes/sq km/year)
	Medium Lightning Density (>3 and <8 flashes/sq km/year)
	Low Lightning Density (≤3 flashes/sq km/year)

Please identify approximately how long you have been collecting data at the monitoring locations.

# of Monitors	# of Years of Data
	<2 years
	2-4 years
	>4 years

Please return to:

Chris Melhorn
EPRI PEAC Corporation
942 Corridor Park Boulevard
Knoxville, Tennessee 37932
(865) 218-8013
fax (865) 218-8001
cmelhorn@epri-peac.com

Section 2B – Site Specific for Each Monitor Location

Please fill out this survey for each monitor on your system.

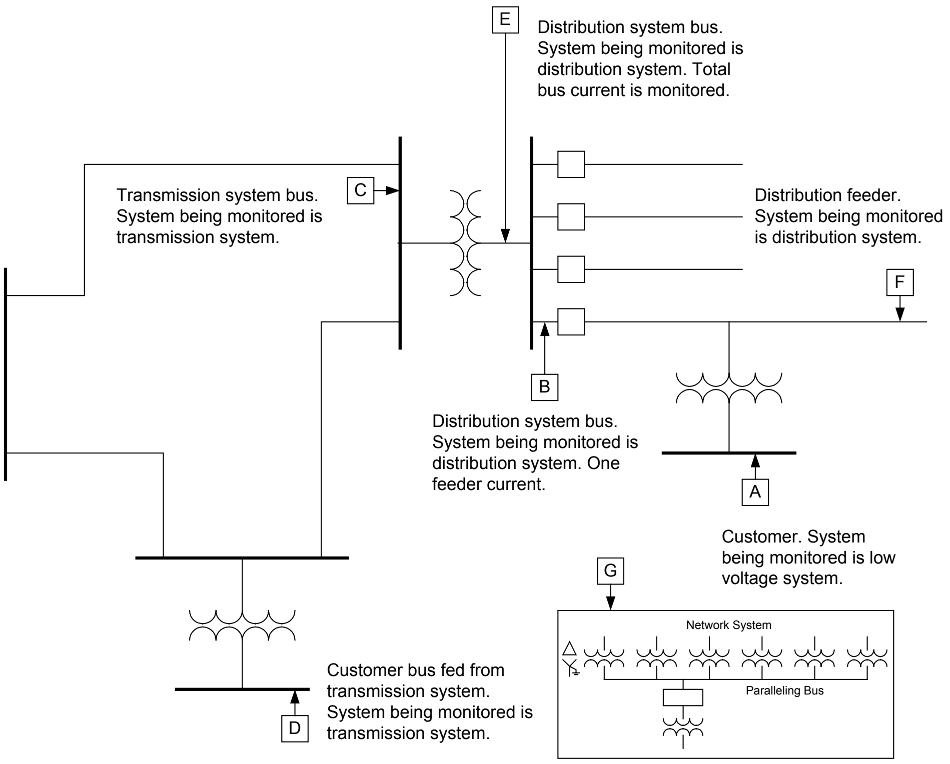
Unique Site Name: (e.g., Transformer #1 at Boulder Substation)
*Please include the zip code for each monitor. The zip code will be used for determining the lightning flash density for each monitoring site.

Type of monitor at this site: (e.g., Dranetz-BMI 7100)

System level that instrument is monitoring.
(See figure below. For simplicity, the distribution system is shown to be radial. However, monitor location A, B, F, or E may represent systems that are loop, primary selective, secondary selective, or spot network.)

A – Customer Bus Distribution System
B – Feeder Monitor Distribution System
C – Transmission Bus Transmission System
D – Customer Bus Transmission System
E – Distribution Bus Distribution System
F – Distribution Feeder Distribution System
G – Underground Network
H – Other (explain): ______________________________

____



Characteristics of the distribution/transmission system that impact the voltage sag rate at each monitor location.

(For location G - Underground Secondary Network- please specify the network secondary voltage and the primary voltage.)

4.1 Distribution System Voltage (Line-Line) ______________________________ **kV**
(Not Applicable for Locations D or C)

4.2 Transmission System Voltage (Line-Line) ______________________________ **kV**

4.3 Distribution Feeder Characteristics
(Not Applicable for Locations D or C)

A – Mostly Overhead
B – Mostly Underground
C – Mostly Mixed

Approximate Total Circuit Length in Miles (for all 3-ph feeders from the substation)

____ M

4.4 Distribution Feeder Type Based on Load Density in MVA/Mile

(Not Applicable for Locations D or C)

A – Mostly Rural; Load Density in MVA/Mile ______________________
B – Mostly Urban; Load Density in MVA/Mile ______________________
C – Mostly Mixed; Load Density in MVA/Mile ______________________

4.5 Provide an approximate mix of customer class fed by the distribution substation

(Not Applicable for Locations D or C)

Residential ____________%
Industrial ____________%
Commercial ____________%

For all monitor locations (except for C, D, or G), please identify the type of service systems and the number of monitors for each system.

A – Radial
B – Primary Loop
C – Primary Selective
D – Secondary Selective
E – Spot Network

Monitoring Data

How much monitoring data do you have for this site?

Less than one monitor year
One to two monitor years
More than two monitor years. Provide actual number. ______________________

Please return to:

Chris Melhorn
EPRI PEAC Corporation
942 Corridor Park Boulevard
Knoxville, Tennessee 37932
(865) 218-8013
fax (865) 218-8001
cmelhorn@epri-peac.com

Target:

Power Quality for Improved Energy Delivery and Distribution

About EPRI

EPRI creates science and technology solutions for the global energy and energy services industry. U.S. electric utilities established the Electric Power Research Institute in 1973 as a nonprofit research consortium for the benefit of utility members, their customers, and society. Now known simply as EPRI, the company provides a wide range of innovative products and services to more than 1000 energy-related organizations in 40 countries. EPRI's multidisciplinary team of scientists and engineers draws on a worldwide network of technical and business expertise to help solve today's toughest energy and environmental problems.

EPRI. Electrify the World

Printed on recycled paper in the United States of America

1005921

EPRI • 3412 Hillview Avenue, Palo Alto, California 94304 • PO Box 10412, Palo Alto, California 94303 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com