Chapter 1

Introduction

1.1-Prelude

The interest and efforts of the last couple of years to modernize distribution systems have resulted in the development of many new concepts and methods for power distribution systems. Similarly to the transmission systems several years ago, monitoring (to improve knowledge of current system conditions), contingency analysis (to determine the effects of outages of system facilities) and control (to provide real-time guidelines for eliminating undesirable system condition) are gaining an increasing popularity in distribution systems.

In many countries, engineers are trying to find reliable solutions for problems such as load flow or state estimation problems related to distribution systems. Research has, however, usually focused on the application of approaches that have been designed for transmission systems. They did not account for the specific requirements of a distribution circuit-based analysis which are very different from transmission requirements **[2]**. Typically, distribution systems are characterized by unbalanced phases and have very short lines with high R/X ratios. Furthermore, the loads are highly distributed and very diverse. It is becoming apparent that current transmission systems programs are inadequate for distribution systems. New methods involving these characteristics are needed to deal with the problems inherent to these systems.

1.2-Statement of the Problem

The latest papers dealing with load flow and state estimation studies **[5, 10]** have begun to take into consideration the specific characteristics of the distribution systems. Various algorithms have been developed for solving these problems and they differ in their characteristics, as well as in their mathematical foundations. Generally, they account for a detailed modeling of the system. Most of the algorithms, however, assume the availability of a number of measurements large enough so that the system is observable. Unfortunately, distribution systems are characterized by a lack of accessible real-time measurements. As a result, numerous studies can be called into question.

The aim of this thesis is to bring a reliable solution for estimating the confidence intervals containing the missing values of power consumption. The missing consumption values are replaced by confidence intervals that contain the exact values with a given level of confidence. In this way, measured values can be used wherever they are recorded, and estimated confidence intervals can be utilized otherwise. The same basic idea may be adapted to provide missing data required to solve planning problems such as transformer sizing.

1.3-The Proposed Solution

The proposed solution is based on the application of a computationally intensive statistical method, namely the bootstrap method, to estimate confidence intervals containing the exact values of power consumption. Two distinct methods have been developed. They utilize the bootstrap method in its two modes: nonparametric and parametric. The first method estimates confidence intervals for the nodal hourly power consumption. The lower and upper bounds of these intervals can then be used in load flow calculations related to distribution networks. The second method estimates confidence intervals for the nodal maximum power consumption per customer. These intervals can be employed in solving planning problems such as the sizing of transformers.

The bootstrap method is a technique for making inferences about a population’s characteristic. It uses a random sample drawn from that population. This sample is usually the only information known about the corresponding population. Thus, it is very important that this sample be representative of the population. In the first study, the sample consists of nodal hourly power consumption values recorded during several previous years. In the second study, the sample consists of nodal maximum power consumption values per customer.

The nonparametric bootstrap method was chosen because it does not assume that the random sample was selected from a probability distribution. It relieves the analyst from having to make assumptions about the population. This is done by drawing, randomly with replacement, a large number of ‘resamples’ of same size as the original sample. From each resample, the statistic considered is calculated. The method relies on the fact that the empirical distribution function based on the sample is an estimate of the distribution of the population. This estimate is used to calculate a confidence interval that contains, with a given level of confidence, the true parameter value. This level of confidence is fixed to 95%.

In the estimation of confidence intervals for nodal hourly power consumption, the nonparametric bootstrap method is used. In the first step of the calculation, the database has been processed in order to create the original basic sample. This processing is presented in Chapter 3. A nonparametric bootstrap study is then carried out. Two thousand resamples are drawn randomly with replacement from the original sample. For each resample, the estimator of the mode is calculated. The mode was chosen in that case because it is defined as being the value that occurs with the highest frequency. It is then possible to estimate the mode's sampling distribution by building the relative frequency histogram from the values previously calculated. A 95% confidence interval is inferred by rejecting the values of the mode, x, such that $F_X(x) < 0.025$ or $F_X(x) > 0.975$ where F_X is the cumulative distribution function for the mode estimate.

The nonparametric bootstrap method is a very powerful method which can be utilized to solve many different problems. There are, however, some examples for which this method may provide inaccurate results. The estimation of confidence intervals for nodal maximum power consumption per customer is one of them. In Chapter 6, it is explained why the nonparametric method is not accurate in that case. The parametric bootstrap is then introduced as a more appropriate method to calculate these intervals. This method is different from the nonparametric bootstrap method in that it utilizes a parametric model to represent the original sample's distribution. Why it is preferred to the use of a textbook formula? The basic idea is that the parametric bootstrap method provides more accurate answers than textbook formulas. Furthermore, it can provide answers to problems when no textbook formulas exist, which is precisely the case in distribution systems. Two examples are proposed in Chapter 6 in which this method is applied. A data processing presented in Chapter 3 allows us first to create the basic original sample. A study using this sample is then carried out in order to define a more realistic parametric model for the data distribution. It is chosen between the three following well-known distributions: the beta, the lognormal and the chi-square distribution. For each one, the associated parameters are estimated. A measure of distance, namely the Kullback-Leibler's measure, is presented and then applied to select the closest parametric distribution to the sample's distribution. Next, instead of sampling with replacement from the data sample, two thousand resamples are drawn from the parametric estimate of the population. For each one, the maximum likelihood estimate, which is the largest sample's value, is calculated. This statistic is chosen because in the case of transformer sizing, the maximum of consumption is the important value to know. The way to infer 95% confidence intervals is the same as in the previous study.

1.4-Outline of the Thesis

Chapter 2 gives an overview of the American distribution networks. The aim here is to highlight the main characteristics of these networks. In particular, the differences between urban and country distribution networks are outlined. At this point, it is possible to enumerate the main differences between distribution and transmission systems. A discussion of some solutions brought to the state estimation and load flow problems related to distribution networks concludes this chapter.

Chapter 3 describes and analyzes a large collection of real data. It is used in the confidence interval estimation presented in Chapters 5 and 6. This dataset was recorded on a distribution network for different demand points. First, the database is processed to allow a faster and more reliable data analysis. Second, two other processing procedures are performed in order to create the basic random samples for the calculations of the confidence intervals.

Chapter 4 presents the original method involved to calculate confidence intervals for nodal hourly power consumption: the nonparametric bootstrap method **[15]**. A comparison with the classical procedures based on parametric distribution assumptions allows us to highlight its reliability and its efficiency. The computational translation of the method is also explained. The random generator, a vital part of this program, is described **[19, 20]**.

In *Chapter 5*, a confidence interval for the nodal hourly power consumption at a particular demand point is estimated for two specific examples. For a specified hour of the year, the calculation uses the nodal hourly power consumption values recorded during the previous years together with the nonparametric bootstrap method to infer a confidence interval for the mode estimator. This result is needed in load flow calculations related to distribution networks.

In *Chapter 6*, a confidence interval for the nodal maximum power consumption per customer is estimated through examples. But first, it is explained why the nonparametric bootstrap method is not appropriate in that case and may provide inaccurate results. The parametric bootstrap is then presented as an accurate method to solve this particular problem. The typical characteristics of the parametric bootstrap method compared to the nonparametric bootstrap method are described. The method is then used to calculate confidence intervals for two examples. These intervals are needed in planning problems such as transformer sizing.

In *Chapter 7*, the conclusions drawn from the previous chapters are summarized. Ideas for future research are presented.

The appendices contain supplementary material. Appendices A, C, D include the programs *PROCESS 1, 2, 3* written in C and FORTRAN languages. They are used in the data processing presented in Chapter 3. Appendix E includes the FORTRAN language source code of the random generator presented in Chapter 4. Appendix F gives the program needed to calculate the confidence intervals for nodal hourly power consumption. Finally, Appendices G and H describe the two parts of the program utilized to estimate the 95% confidence intervals for nodal maximum power consumption per customer. These programs are written in FORTRAN language.