THE UNIVERSITY OF OKLAHOMA

Graduate College

ESTIMATION OF PRODUCTION COST VARIANCE USING CHRONOLOGICAL SIMULATION

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

Jia-Yo Chiang

Norman, Oklahoma

1997

UMI Number: 9806321

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

ESTIMATION OF PRODUCTION COST VARIANCE USING CHRONOLOGICAL SIMULATION

A Dissertation APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

By

ACKNOWLEDGMENTS

The author wishes to thank all of his committee members for their time and advice. Especially, the author would like to express his sincere gratitude to Dr. Arthur M. Breipohl for his patience, inspiration and guidance throughout this work. Thanks are also due to Dr. Fred N. Lee for lending his technical programs in accomplishing this task.

The special thank goes to my lovely wife, Ye-Ju, and my adorable daughter, Kelly, for their support and cooperation throughout my course of study. Also, my parents deserve my cordial appreciation for their support and encouragement. Without their understanding and eternal love, to accomplish this long cherished dream would not be possible.

My deep appreciation is extended to my colleague, Andrew Douglas, who shared his ideas and suggestions during these years have been valuable.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS iv

TABLE OF CONTENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

ABSTRACT x

CHAPTER

I. INTRODUCTION 1

I. 1 Production Cost Simulation and Its Uncertainty 2

I. 2 Load Uncertainty Modeling 8

I. 3 Outline of The Dissertation 10

II. CHRONOLOGICAL PRODUCTION COST SIMULATION WITH UNIT AVAILABILITY UNCERTAINTY 12

II. 1 Chronological Production Cost Simulation 13

II. 2 Ordinary Monte Carlo Sampling 15

II. 3 Smart Monte Carlo Sampling 17

III. ANNUAL LOAD VARIATION 21

III. 1 Analyzing Historical Load Data 22

III.2 Normal Sampling And Stratified Sampling 29

III.2.1 Normal Sampling 31

III.2.2 Stratified Sampling 32

III. 3 Chronological Load Projecting Algorithm 36

IV. TWO APPROACHES OF ESTIMATING COST VARIANCES 39

IV.1 Stochastic Approach 39

IV.1.1 Weekly Conditional Sampling 39

IV.1.2 Cost Variances Estimation 44

IV. 2 Probabilistic Approach 46

IV.2.1 Stratified Load Scenarios Sampling 46

IV.2.2 Cost Variances Estimation 55

V. PRODUCTION COST VARIANCES AND COMPARISIONS 57

V.1 Example of the Stochastic Approach 57

V.1.1 Cost Variance Due to Unit Outage Uncertainty 57

V.1.2 Cost Variance Due to Load Variation Uncertainty 59

V.1.3 Cost Variance Due to Both Uncertain Factors 62

V.2 Example of the Probabilistic Approach 65

V.3 Comparisons of the Two Approaches 69

V.3.1 Load Variation 69

V.3.2 Quality of Estimation 70

V.3.3 Effort of Load Variation Modeling 72

V.3.4 Computational Time 72

VI. CONCLUSION 74

BIBLIOGRAPHY 77

APPENDIX A 82

APPENDIX B 83

APPENDIX C.1 86

APPENDIX C.2 94

APPENDIX C.3 99

LIST OF TABLES

TABLE Page

3-1 Adjusted Future Annual Energy and Peak Load 24

5-1 Stochastic Results Due To Unit Outage Uncertainty...................................... 58

5-2 Stochastic Results Due To Load Variation Uncertainty (Case 1) 60

5-3 Estimated Cost Variance Due To Load Variation 62

5-4 Stochastic Results Due To Both Uncertain Factors 64

5-5 Cost Variance and Coefficient of Variation of Stochastic Approach 65

5-6 Annual Cost Summary of Probabilistic Approach 66

5-7 Cost Variance and Coefficient of Variation of Probabilistic Approach 67

5-8 Case Study Summary of Probabilistic Approach 68

5-9 Annual Load Variation Summary 70

5-10 Coefficient of Variation of Cost Summary of Two Approaches 71

A-1 Generation Unit Data 82

B-1 Stochastic Results Due To Load Variation Uncertainty (Case 2) 83

B-2 Stochastic Results Due To Load Variation Uncertainty (Case 3) 84

B-3 Stochastic Results Due To Load Variation Uncertainty (Case 4) 85

LIST OF FIGURES

FIGURE		Page
3-1	Estimated Mean of Weekly Energy With Uncertainty Band	25
3-2	Estimated Mean of Weekly Peak Load With Uncertainty Band	25
3-3	Estimated Weekly Energy Growth	26
3-4	Estimated Weekly Peak Load Growth	27
3-5	Correlation Coefficient Between Weekly Energy and Peak Load	28
3-6	Correlation Coefficient of Weekly Energy (One Week Lag)	29
4-1	Sampled Weekly Energy and Peak Load	44
4-2	Stratified Strata of A Bivariate Normal Distribution	48
4-3	Stratified Strata of A Normal Distribution	52
4-4	Daily Curves of Three Load Scenarios	54

ABSTRACT

This dissertation describes research on the effects of the uncertainty in annual load variation and uncertainty in generation availability on the variance of production cost in an electrical power system. Two different approaches of load uncertainty modeling are developed. The load uncertainty modeling accounts for uncertainty when reliable weather (e.g. temperature) forecasts are not available. Both approaches can be used to estimate the variance of long-term production cost, typically for an annual study.

Production costs are usually simulated on the basis that the availability of generation capacity is subject to random failures of system generating units. In order to estimate the variance of cost, both the random forced outages of units and load uncertainty should be modeled in a production cost simulation. In this dissertation, the effects of uncertainties in generation availabilities will be analyzed using a Monte Carlo approach.

In this long-term (annual) load model, emphasis is placed on modeling the variation in chronological load so that a chronological production cost simulator can efficiently produce an estimate of the variance of annual production cost. A stochastic approach, using a conditional weekly sampling scheme, is proposed to model the annual load variation. Then, a probabilistic approach using stratified sampling is proposed to model the load variation on an annual basis. Furthermore, the stochastic and probabilistic approaches are compared in terms of accuracy and effort. The result of this dissertation is

to provide a model that can be used to estimate the variance of annual production cost using a chronological production cost simulator. Thus, the variance in production cost can be expressed as a function of load uncertainty and uncertainty in generator availability.

CHAPTER I

INTRODUCTION

The objective of an electrical utility has been to provide electrical service to its customers in an economical and reliable manner. As part of this process, system operational scheduling, fuel budgeting, transmission planning and system planning studies play important roles. Unfortunately, these functions are complicated by future uncertainties. Also, the increasingly competitive markets and the onset of de-regulation of electrical utility industry with their concomitant emphasis on prices rather than simply costs have increased the interest in reflecting uncertainty in power system planning studies. Therefore, the need to represent uncertainty of future conditions in system operation and planning is widely recognized in the electrical power industry.

Regardless of how long is the lead time, a generation system needs to be scheduled and operated to satisfy the operational constraints for the planning horizon at a minimum cost. System operational scheduling and system planning both depend on many forecast parameters, including: unit availability, electrical demand, fuel prices, and environmental and regulatory requirements. Unfortunately, the forecast of these parameters is subject to uncertainty. Furthermore, this range of uncertainty tends to become broader as the forecast lead time lengthens. Although uncertainties are broadly considered in long-term operational planning, it is also very clear that uncertainties are also essential in the analysis of short-term operational strategy.

Planning under uncertainty involves identifying the potential uncertain (or risky) events and assigning probabilities to those events. Among the uncertain factors that impact the system production cost, this dissertation's attention is concentrated on the uncertainties of generation units' forced outages and load uncertainty. The problem of modeling the reliability of the generating unit forced outage performance has received considerable attention in the past. However, load uncertainty may play a larger role in production cost uncertainty than uncertainty in generation availability does. Furthermore, not only will the accuracy of load representations impact the estimation of production cost of an electric power system, but also the uncertainty in load will contribute to the variation in production cost. Power system scheduling and planning begins with the development of load representation models. Consequently, to estimate the variance of production cost induced by load variation, emphasis is placed on developing a chronological load model that includes uncertainty. In this dissertation, presuming that the annual load forecasts do not use weather forecasts, the annual load variation representations are modeled to estimate the variance of production cost in system planning studies. In the sequel, uncertainty in load and uncertainty in generation availability are reflected through production cost simulation in order to estimate the uncertainty in operational cost.

I.1 Production Cost Simulation and Its Uncertainty

When planning future operations, questions that naturally arise are how much generating capacity is required to meet the potential demand and how much uncertainty is associated with the plan for committing generation. As power generating units experience unexpected forced outages, the uncertainty in unit's availability must be included in production cost simulation if accurate results are to be obtained. Usually, probabilistic methods can be used to account for the characteristics of the random outage rate. In this research, the unit forced outage rate is implemented in Monte Carlo sampling using a discrete-state Markov process (stochastic process).

On the other hand, many techniques and approaches to short-term load forecasting have been proposed in the last two decades. Among them are: artificial neural networks algorithm [1, 2, 3], time series analysis [4, 5, 6], linear regression methods [7, 8] and fuzzy logic approaches [9, 10]. Most of these load forecast procedures use the weather information to refine the prediction. Without a reliable long-term weather forecast, electrical load forecasts with these long lead times are challenging, and are complicated by the future uncertainties. Since accurately predicting production costs is a vital function for supporting system planning and operational scheduling, production cost simulations with the capability of reflecting annual load uncertainty are becoming an important tool in power system operational planning.

To plan for the future, a production cost simulator is needed to assess and compare the projected production costs of alternative strategies with a view to optimizing the overall system economy. Production cost models are computational models designed to calculate a generation system's production cost. They are widely used throughout the electrical utility industry as a tool in long-range system planning, in fuel budgeting and in system operation. The primary function, i.e., estimating future system energy cost, is accomplished by simulating the operation of the generation system to meet the projected future loads. Since generating units are not perfectly reliable and future loads are forecasted with uncertainties, many production cost programs are used mainly to compute the expected cost based on deterministic or probabilistic models. To provide more meaningful simulation results, a production simulator should be able to evaluate the effects of uncertainties. A production cost simulation with load and generator availability uncertainty has the capability to provide information about cost variation, as well as expected cost, can be useful in decision making.

There are two prevalent types of production cost simulation: a probabilistic (or load duration curve based) approach and a Monte Carlo approach. A probabilistic model (load duration curve method) was suggested by Baleriaux [11] and Booth [12]. The basis of this simulation method is the use of probability distributions to describe the system load and generating units' forced outage. The combination of these distributions are used to obtain the expected energy generated for each unit in the system. In this formulation, the load is described by means of a load duration curve which ignores the chronological variation in the load distribution. It has been shown in Reference [13] that using the concept of unit duty cycle, the Equivalent Load Duration Curve (ELDC) approach, can be extended to approximate unit start-up failure and start-up time considerations. In recent years, the method of moments described in Reference [14] and [15] had been adopted in the computation of the ELDC. This development greatly improved the computational efficiency of the analytical approach. In practice, the probabilistic model is fast compared to chronological simulations and is attractive for planning studies that extend for more than one year. However, an inherent limitation of the analytical approach to production cost calculation is that the chronological information of the electrical load is lost.

The need for generating the distributions of production costs, as opposed to an expected value only, has been emphasized by Rau and Nesculescu [16]. Reference [17] extends the Baleriaux formulation so as to be able to compute the variance of the system production costs. Due to the large quantity of coefficient of variation of cost observed in Reference [17], Mazumdar and Yin suggest that production costing models should be enlarged to include the variance of cost as well. In Reference [18], Rau and Hegazy further pointed out that the variance of the production cost is greatly influenced by the method of modeling the load and the implicit assumption regarding the outage of generating units. They provide a simpler procedure than Mazumdar does to obtain the variance of production cost. Reference [19] presents a stochastic outage capacity state model for evaluating the random error in production cost, which is also estimated via the Baleriaux-Booth approach. Using these probabilistic approaches to estimate the variance of production cost [16, 17, 18, 19], the load uncertainty can be included by the use of probability distributions that describe the system loads. Thus, the effects of load uncertainty on production cost can be modeled using a load duration curve method.

Since production costs are sensitive to chronological effects, Monte Carlo simulations provide more realistic and accurate results than those provided by probabilistic models, especially in the short-term power system operational scheduling. Monte Carlo simulation is a numerical simulation procedure applied to problems involving random variables with known probability distributions. The Monte Carlo simulations described in Reference [20~23] have been developed to estimate the production cost considering the chronological variation of generating units' availabilities. This method involves repeating the simulation process using a particular set of sampled values of the random variables that are generated corresponding to their probability distributions in each iteration. Therefore, the results of Monte Carlo simulation are estimated statistically.

A major concern with regards to the application of Monte Carlo simulations is that a large number of samples is required to obtain meaningful results. The method for estimating approximate confidence bounds of production cost for Monte Carlo simulation is proposed in Reference [23]. This method can be used to determine a sufficient number of simulation samples needed to achieve a specified degree of precision, thus resulting in savings in computing time. In Reference [24], a combined control variable and stratified sampling method has been proposed for Monte Carlo production cost simulation. This method reduces the computational time and enables the chronological simulation to play a much more significant role in medium range planning studies. A structure for stochastic chronological production cost simulations is described in Reference [25] which is useful for handling annual chronological constraints in operational scheduling and planning.

Monte Carlo simulations have been used to model the effects of unit outage characteristics on the variance of annual production costs based on an expected load profile [22, 23]. Most production cost simulators estimate the expected production cost, but do not give information regarding the extent of the fluctuation of these costs.

Information containing only the expected production costs may be insufficient for system planners on many occasions. A decision maker may want to estimate the accuracy of the results to prepare for the unexpected in the future. For example, an increasly important use of production cost estimation is to select a most economical transaction or to set the priorities to the available transaction options. In this case, production cost simulation studies are used to evaluate potential benefits associated with contractual purchases or sales of electricity to another utility or energy marketer. The energy transaction can be modeled as specified megawatt sales or purchases on an hourly basis. As such, they can be considered as a territorial load adder or subtracter for the system load. However, if the actual annual energy turns out to be more than what is expected due to the uncertainty in the forecasted system loads, the increase of the annual energy will cause the system production cost to increase dramatically. Thus, for this unexpected situation, the fluctuation in the estimated production cost will not be captured by a single estimate of the expected production cost. Moreover, when a situation exists such that the difference between the estimated production costs of two transaction options is very small in magnitude, a measure of the variation of the production cost about the expected cost considering the load uncertainty will be useful to a decision maker.

Although Wang proposes a method for estimating confidence bounds of Monte Carlo simulation results in Reference [23], this method only focuses on the effects of unit availability uncertainty on the variance of production cost. Usually, the forecasted system loads are handled by a deterministic approach in Monte Carlo simulations. Thus, the uncertainty in system load is not considered, when the variance of production cost is estimated based on the expected load profile. An expected load profile will not be adequate in estimating the variance of production cost. In order to estimate the variance of production cost caused by annual load variation, a method of modeling the variation in a chronological load profile is required when using a Monte Carlo simulation technique.

I.2 Load Uncertainty Modeling

When the future uncertainty is included, a measure for precision on the projected cost becomes very import. A decision maker would like to make a decision between "higher risk - lower operating cost" and " lower risk - higher operating cost". For most utilities, uncertain load requires higher planning reserve margins than known demand does. Therefore, if this additional information is desired by the decision maker, a method of modeling load uncertainty is essential. In this dissertation, instead of modeling the expected load profile only, a chronological load uncertainty model will be developed to represent the annual load variation. Herein, two approaches to load variation modeling, in conjunction with a chronological production cost simulation, are proposed.

Since the load at a fixed time is a sum of independent variables, i.e., the loads of many customers, and based on the Central Limit Theorem, the random variables of this application have normal distributions. Likewise, it is reasonable to presume that both the total energy and peak load of a particular week are also normal random variables. Furthermore, the characteristics of these random variables change with time.

If a collection of random variables is changed with a parameter such as time or space, such collections are known as stochastic processes. Consider a random variable that represents the weekly peak load in an electrical system, it is intuitive to model the weekly peak load as a stochastic process. Similarly, the weekly energy also should be treated as a stochastic process. Moreover, these two random variables are not independent in each week. In addition, the electric load is highly dependent on the weather conditions, which are changing all the time. Therefore, the correlation between weekly energy and peak load changes as the time progresses as well. For this research, a discrete-state Markov process unit availability model has been adopted in Monte Carlo simulation. Modeling the system load as a Gauss-Markov random process is consistent with that of the unit outage modeling. These representations lead us to model the annual load variation using a stochastic approach.

In the stochastic approach, a conditional sampling scheme for the weekly energy and peak load will be used to model the annual load variation. However, modeling the annual load variation in weekly detail requires the two sets of distributions that describe the stochastic characteristics of the weekly energy and peak load. Accurate estimations of these weekly parameters usually are the bases used to model the annual load variation. In practice, it is too tedious to estimate all the weekly parameters, and it is not efficient to sample a large number of load profiles to reflect the annual variation. A simplified approach to estimate the variance of cost induced by unit outage and load variation uncertainties is necessary.

Ideally, it would be useful to provide the estimate of expected cost when the future load is high or low, in addition to the expected cost and the variance of cost based on the expected load. The probabilistic approach is developed to approximate the variances of cost with reduced number of simulations; and to provide the expected costs of three scenarios (low, mean, and high load). The annual energy and peak load are modeled by a joint normal distribution with the consideration of correlation. Thus, the variance of production cost induced by load uncertainty are estimated according to the simulation results of the three load scenarios, which are represented using stratified sampling. In addition, the results of the probabilistic approach are compared to those of the stochastic approach to ensure that comparable estimations have been obtained.

I.3 Outline of The Dissertation

The reminder of this dissertation consists of five chapters. A review of the production cost simulator and the generation availability uncertainty modeling for Monte Carlo simulation will be provided in Chapter II. The smart Monte Carlo sampling for estimating the expected production cost and ordinary Monte Carlo sampling to evaluate the variances of production cost are also discussed in this chapter. In Chapter III, the annual load variation is analyzed. Also a stratified sampling technique to describe probabilistic load scenarios and a conditional sampling scheme used to sample the stochastic annual load profile are presented. In Chapter IV, the stochastic and probabilistic approaches to modeling annual chronological load variation are described. The mathematical formulation to estimate the variance of production cost caused by unit outages and load variation using two proposed approaches are presented in this chapter as well. The estimation of the expected costs and variances of a medium size utility using both approaches are illustrated in Chapter

V. Also, these two approaches are compared in this chapter. Chapter VI concludes this dissertation.

To complete this dissertation, the programs of both approaches are listed in Appendix C. The stochld.c and probald.c are the main programs for the stochastic and probabilistic approaches respectively. The loadlib.c is a subroutine library for both programs. These two main programs are used to generate the annual load profiles in the format of *Scheduler's* binary load file. All the programs are written in C language.

CHAPTER II

CHRONOLOGICAL PRODUCTION COST SIMULATION WITH UNIT AVAILABILITY UNCERTAINTY

In the electrical power industry, a production cost simulation consists of modeling generation system, system load, a number of constraints including, fuel constraints, transmission constraints, emission constraints, and transactions in order to determine the generation operating expenses that the utility will incur for a future period. In the past, chronological production cost simulation algorithms were not popular because they required higher computational effort and are consequently of limited use in the context of long-term studies. As the electric utility industry is becoming a competitive market, there is an incentive to reduce costs and increase profits. The ability to accurately evaluate the future production cost is of considerable importance to utility planners. Thus, the need to estimate accurately both short-term and long-term production costs has called for the reconsideration of detailed chronological simulation so that chronological events can be included.

In this chapter, in order to estimate the variance of production cost, we begin with a review of some important features concerning the uncertainty in unit availability sampling, which affect the estimation of the production cost. Followed the review are the descriptions of the two different Monte Carlo sampling techniques used to model unit outage uncertainty.

II.1 Chronological Production Cost Simulation

In this dissertation, the precision of the estimate of production cost is studied by the estimated variance. To estimate the variance of annual production cost, a production cost simulator based on Monte Carlo simulation technique is used. A Monte Carlo simulation is a numerical simulation procedure applied to problems involving random variables with known probability distributions. Assuming that the probability distributions of these random variables are given, the first step in Monte Carlo sampling, is to generate random samples based on their distributions. The next step is deterministic, in which the unit commitment and economic dispatch are used to obtain the simulated results. Finally, after the first two steps have been repeated a sufficient number of times, a statistical analysis of the simulated results can be performed.

Computer programs for electrical energy production, based on the Monte Carlo sampling techniques, have long been used in the utility industry. A Monte Carlo type production cost simulator, *Scheduler*, is used in our studies. *Scheduler* is a chronological resource scheduling program developed by Power Costs, Inc.. This software package is capable of both short-term operational scheduling and long-term operation planning, because it simulates the operational costs of the power system for time periods between one hour and several years. This simulator captures the chronological aspects of power system operations including unit commitment constraints such as, minimum up-time, minimum down-time and start-up costs. *Scheduler* uses the sequential bidding method [26] to solve the unit commitment problem. It also models adaptive compliance with annual emission and fuel constraints.

That is, an allocation of the constrained fuel is made for the first week; the simulation results of the first week are monitored at the end of the first week, and an allocation is made for the second week based on the results of the first week's operation. This adaptive fuel allocation is used to assign the upcoming week's target consumption of a constrained fuel in order to satisfy the overall fuel constraints.

Because of random load fluctuations and random outages of generating units, any realistic long-term planning or production cost simulation procedure must take into account the uncertainty of the future. Production costs are usually simulated on the basis that the availability of generation capacity is subject to random failures (i.e., forced outages) as well as scheduled maintenance outages of system generating units. The forced outage of generating units assumes that the generating units are independent of each other and the load. Thus, the availability status of each generating unit is determined by independent Monte Carlo sampling.

Scheduler models the effects of uncertainty in generator availability using Monte Carlo sampling. When Monte Carlo sampling is used with a chronological model, the resulting simulation is called stochastic. Stochastic models of production cost have the advantages of accuracy; however, many repetitions of a chronological algorithm are often required for the estimated mean of an observed quantity to be a close approximation of the true mean. Such a procedure makes the simulation very time-consuming. As a result, a conditional sampling technique is introduced to improve the precision of Monte Carlo estimates, and thus reduces the required number of repetitions. By using this "smart" Monte Carlo sampling, the coefficient of variation of cost, which is due to unit forced outages , is significantly reduced so that one annual production cost simulation is sufficient to estimate the expected production cost. However, in order to estimate the variance of the production cost, the smart Monte Carlo sampling must be replaced by ordinary Monte Carlo sampling.

On the other hand, *Scheduler* and all other chronological models, of which the author is aware, model the system load using a deterministic approach. Since the expected chronological load profile is used, the resulting production cost only takes into account the unit availability uncertainty. That means it estimates the production cost based on the expected loads. Thus, a tool which can be effectively reflect the variation in system load to production cost simulation will be useful, and should be considered especially for long-term studies. This is the major contribution of the research reported in this dissertation.

II.2 Ordinary Monte Carlo Sampling

A generating unit has several operating conditions. A planned maintenance outage (often called an "overhaul") occurs when a unit is removed from service to perform routine work that is scheduled in advance. A forced outage occurs when a unit is unexpectedly forced out of service for repair. The forced outage rate is defined as the fraction of time that the generating unit is unavailable for service. Thus, the forced outage rate measures the ratio of the hours when the unit is unavailable due to a forced outage divided by the exposure time, i.e., available hours plus forced outage hours. Both forced outages and maintenance outages are considered in the production cost simulation.

In a Monte Carlo type simulation, a random sampling technique is used. The heart of Monte Carlo sampling is a pseudo random generator. The occurrences of events in the simulated operations are determined by the outcomes of the random draws. In practice, random numbers are usually generated on a weekly basis. Units on outages and units in service remain unchanged for the entire week. In order to recognize the chronology of all simulated events, Monte Carlo sampling is used to describe the availability status of each generating unit in the production cost simulation. Chronological Monte Carlo models represent the system state as a set of resource availabilities. The system state, described by the on/off status of every available unit, can be used to formulate a deterministic model. A simulation for a given time period, e.g., one week, for one sample system state is called an iteration. For each generating unit, the system state is established for each iteration by comparing a randomly generated number with the forced outage rate of that unit. The set of states of each unit typically has just two possibilities, the possibility of that unit being available at the unit's rated capacity and the possibility of it being on outage at zero capacity. The state of a unit can be determined to be 1 – which represents available, if randomly generated number is greater than the forced outage rate of that unit – else the state will be 0, which represents unavailable.

Assuming the generating unit's availabilities are independent of each other, a set of unit states is sampled for each iteration. After establishing a set of system states for each iteration, a chronological simulation model can be applied to simulate the production cost of each iteration. With a sufficient number of repeated iterations, the simulation results are estimated statistically to represent the expected value of production costs. In order to estimate the mean and variance of the production cost with an acceptable precision, a large number of simulations is required. Therefore, a "smart" Monte Carlo sampling technique is introduced to fulfill the need to reduce the simulation time and maintain the accuracy of production cost estimation.

II.3 Smart Monte Carlo Sampling

As the Monte Carlo simulation results are derived by averaging the outcomes of several iterations, the basic Monte Carlo method requires a large number of repetitions to reduce the coefficient of variation of the resulting cost to an acceptable value. These averages will be imprecise if the number of iterations is not large enough. For example, if the number of iterations is not large enough, the frequency of unit outage will not closely approach the unit forced outage rate. Clearly, the number of repetitions will affect the precision of Monte Carlo estimates. In *Scheduler*, a "smart" sampling scheme is applied to improve the precision of the estimates.

In this production cost model, planned maintenance outages are considered first. When the iteration (week) is in the unit's planned maintenance period, the unit is unavailable. These weeks will not be included in the weeks sampled for forced outages. For those weeks which the unit is not on planned maintenance, "smart" Monte Carlo sampling is performed. For each unit, the expected number of iterations (weeks) in which the unit is forced out can be estimated as the total number of iterations (weeks) that the unit is not on planned maintenance schedule times the unit forced outage rate. That is,

$$EOUT = FOR * WAVA \quad (2\text{-}1)$$

where

$EOUT$	Expected number of weeks in which a unit is forced out
FOR	The unit forced outage rate
$WAVA$	Number of week in which a unit is not on planned maintenance

For each unit, the availability status of the first iteration is established by comparing a randomly generated number with its forced outage rate. For the subsequent iterations, the generating unit state can be determined based on the states of the previous iterations. In order to have the number of weeks that the unit is unavailable closely match the expected number of weeks that the unit is unavailable, the forced outage rate for the remaining weeks should be adjusted to reflect the updated conditions. After the first iteration, a randomly generated number is compared to the conditional forced outage rate of that unit, $CFOR_i$ for week i, which is defined as follows:

$$CFOR_i \cong \frac{(EOUT - FOUT_{i-1})}{52 - i + 1} \quad (2\text{-}2)$$

where

$FOUT_{i-1}$: Weeks forced out through week i-1.

Therefore, the sampled unit state is conditional upon the number of weeks in which that unit has been forced out. When the observed number of weeks that the unit has been unavailable reaches the expected number of weeks that the unit is unavailable, the conditional outage rate will be modified to be zero or negative. At this point, the states of the following iterations will be determined to be available. When the observed number of weeks that the unit has been available reaches the expected number of weeks the unit is available, the conditional outage rate will all be modified to be one or greater than one. Then, the states of the following iterations will be determined to be unavailable. Thus the numbers of unavailable and available weeks are both approximately bounded by their expected number of weeks. By using this type of sampling, the simulated number of weeks that the unit is unavailable and the expected number of unavailable weeks will be very close (i.e., the maximum different is one week). After the availability state of each unit is determined by this kind of conditional Monte Carlo sampling, the weekly deterministic chronological production cost simulation can be used.

In this subsection, we present an approximate theoretical analysis of the variance of the cost resulting from the smart Monte Carlo sampling [25]. Since the weekly availability states of a generating unit are dependent on the sampled availabilities of the previous week, the weekly production costs are not independent on each other. However, in order to demonstrate the effect on the variance of production cost using smart Monte Carlo sampling, if the adaptive fuel allocation is ignored ,we assume that the weekly production costs are independent, each with a mean (μ_i) and variance (σ_i^2). The mean (μ_Y) and variance (σ_Y^2) of annual production cost can be estimated as $\mu_Y = \sum_{i=1}^{52} \mu_i$ and $\sigma_Y^2 = \sum_{i=1}^{52} \sigma_i^2$. Assuming $\frac{\sigma_i}{\mu_i} = \frac{\sigma}{\mu}$, the coefficient of variation of annual production cost is $\frac{\sigma_Y}{\mu_Y} = \frac{\sqrt{\sum_{i=1}^{52} \sigma_i^2}}{\sum_{i=1}^{52} \mu_i} = \frac{\sqrt{52}\sigma}{52\mu} = \frac{\sigma}{\sqrt{52}\mu}$, which is less than the coefficient of the variation of weekly cost, σ/μ. However, the weekly samples of this conditional weekly sampling are not independent. This conditional sampling forces the number of weeks that unit is on forced outage to be very close to the expected number of weeks on forced outage. Hence, the variance of the number of weeks that unit is forced out is reduced. Consequently, this smart Monte Carlo sampling can be used to reduce the coefficient of variation of the estimated annual production cost.

The main advantage of this smart Monte Carlo sampling over ordinary Monte Carlo sampling is that smart Monte Carlo sampling can be used to simulate production cost with a small coefficient of variance with one annual simulation. Thus, the estimate of the expected production cost results from this conditional sampling scheme is more accurate. In other words, as compared to ordinary Monte Carlo sampling, the estimated variance of production cost induced by unit outage uncertainty is reduced by using a smart sampling scheme. It will be misleading to estimate the variance of the production cost using the smart Monte Carlo sampling. This is the reason that ordinary Monte Carlo sampling must be used when the estimation of the variance of production cost is important for a decision maker.

CHAPTER III

ANNUAL LOAD VARIATION

The demand of an electric utility varies with time. Although the overall pattern is predictable, there is a large random component that makes the hourly prediction of the electrical load difficult. While an accurate hourly load representation is important, the amount of computer time is almost directly proportional to the number of distinct hourly loads being evaluated. Therefore, there is an incentive to compromise load representation from 8760 hourly loads per year to something less. For this reason, most long-term planning studies tend to use a load duration curve, which gives the description of the load level of a system using a probability distribution. However, the load presentation in *Scheduler* is based on a chronological load profile which will maintain the chronology of time. Thus, our approach to modeling annual load variation is to generate several annual chronological profiles to be used in estimating variance of cost using a chronological production cost simulator.

In order to model annual load variation, it is essential to understand the behavior of electrical load in some detail. In this chapter, the annual variation of the electrical load will be analyzed by using historical load data of the previous five contiguous years of the stimulated system. The main task of this chapter is to present a methodology which can be used to estimate the means and variances of energy and peak load in both annual and weekly time frames. Also, the correlation between energy and peak load, which is critical for the load variation modeling, is estimated. Two sampling techniques for future load variation modeling are implemented based on the characteristics of the system load. With these two sampling techniques, a chronological load projecting algorithm is developed to transform a pair of sampled energy and peak load into a chronological load profile. This chronological load projecting algorithm, in conjunction with different schemes for sampling energy and peak load, can be used to create chronological load profiles in the stochastic and probabilistic approaches of load variation modeling.

III.1 Analyzing Historical Load Data

Electric load varies from day to day, week to week, month to month, and year to year. Among those future uncertainties that affect the electrical demand, the weather conditions and economical growth are two significant factors considered in the modeling of annual load variation. In general, weather forecasts with long lead times are usually either unavailable or unreliable, so we model the load variation without using weather forecasts in an annual horizon. However, the annual economical growth of the service area is usually reflected on the historical load data. Thus, to minimize the influence of economical growth in estimating load variation, the historical load data are adjusted to account for the different growths of different years. The variations in energy and peak load are measured to represent the variation in load.

Even though we model the annual load variation without the information of weather forecasts, the electrical load is still affected by the weather conditions, especially in summer and winter seasons. During the summer, load increases with high temperature. Similarly, during the winter, load increases with low temperature.

Therefore, the influence of weather should be minimized in the process of adjusting historical load data. In practice, the monthly energy of April usually is used to estimate the economical growth with the least amount of weather influence. However, it is possible that abnormal weather may occur in April. An estimate of the non-weather sensitive energy of a year, which is affected the least by the weather, will be a better choice to estimate the annual economical growth. In this dissertation, the average weekly energy of the seven lowest weekly energies of a year is used to represent the non-weather sensitive energy of that year. Hence, the annual load growths of the previous years are estimated using these non-weather sensitive energies as the measure of annual load growth. For example, if the average of the seven lowest weekly energies for year 1 and base year are $Noneng_1$ and $Noneng_b$ respectively. The load growth from year 1 to base year is estimated to be $growth_1 = Noneng_b / Noneng_1$; while the adjusted loads of year 1 are the hourly loads of year 1 multiplied by $growth_1$. Consequently, the historical load data can be adjusted based on the estimated annual load growths to account for the load growths of the previous years. After the annual load growths are adjusted, five adjusted historical load profiles correspondent to the previous five years are used to estimate the means and variances of energy and peak load statistically.

Assuming that energy and peak load are jointly normally distributed, the annual load variation is described by the variations of the total energy and the peak load and correlation between these two variables. A medium size utility, system A, is used in this chapter to illustrate the estimation of annual load variation. Consider year 5 as the base year, we adjust the hourly loads between year 1 and year 4 to base year using the estimated load growths as described previously. If the growths of both annual energy and peak load from the base (fifth) year to the future (sixth) year are not available from the system, they can be estimated according to a linear trend of the estimated load growths of the previous five years. In this case, the growths of the annual energy and the peak load of system A are estimated to be 0.0314 and 0.0228 respectively. Then, the adjusted annual energies and peak loads of these five years are projected to the future year using the estimated growths, and are listed in Table 3-1. The means and standard deviations of annual energies and peak loads are also estimated and provided in Table 3-1.

Table 3-1 Adjusted Future Annual Energy and Peak Load

	Annual Energy (MWH)	Peak Load (MW)
Year 1	23,351,532	5,065
Year 2	22,835,826	4,821
Year 3	23,332,126	5,049
Year 4	23,234,975	4,845
Year 5	22,322,742	4,617
Mean	23,015,440	4,880
STDEV	439,754	185

To model load variation in weekly detail, the means and standard deviations of weekly energy and peak load are estimated as well. According to the five sets of adjusted hourly load data, the mean and standard deviation of weekly energies and peak load are estimated statistically. The smoothed results of the estimated means of weekly energies and peak loads with a one standard deviation envelope are shown in Figure 3-1 and Figure 3-2 respectively.

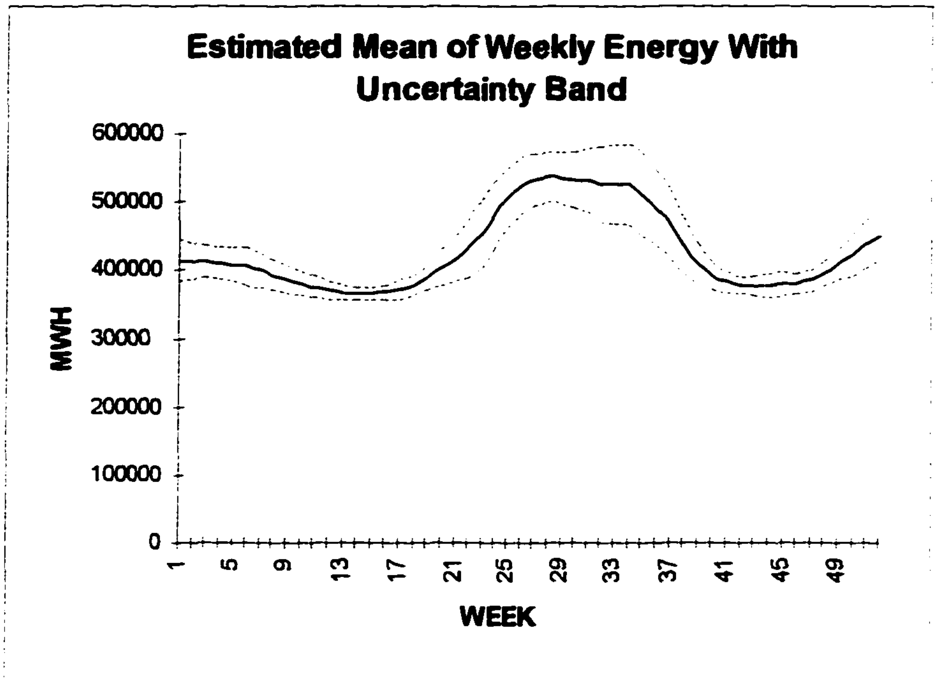

Figure 3-1

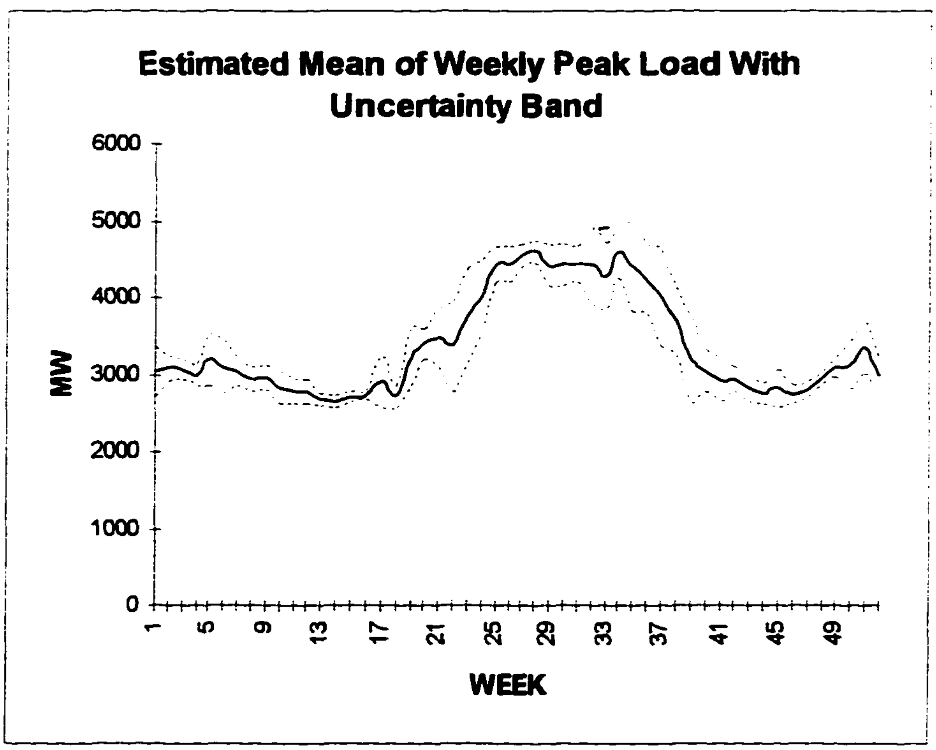

Figure 3-2

Comparing Figure 3-1 and Figure 3-2, there is a consistent weekly pattern in both graphs. Obviously, the expected value of the weekly peak load tracks the expected value of weekly energy quite well. It is on double that there is a correlation between weekly energy and peak load observed from these two graphs.

To simulate the load variation of the future year, the growths of the weekly energies and peak loads from base (fifth) year to future (sixth) year are estimated and smoothed by a moving average routine with a five-week window. The growths of weekly energies and peak loads with the weekly smoothed values are shown in Figure 3-3 and Figure 3-4 respectively. Once again, both the weekly energy growth and weekly peak load growth follow the same pattern. This indicates that there is correlation between weekly energy and peak load.

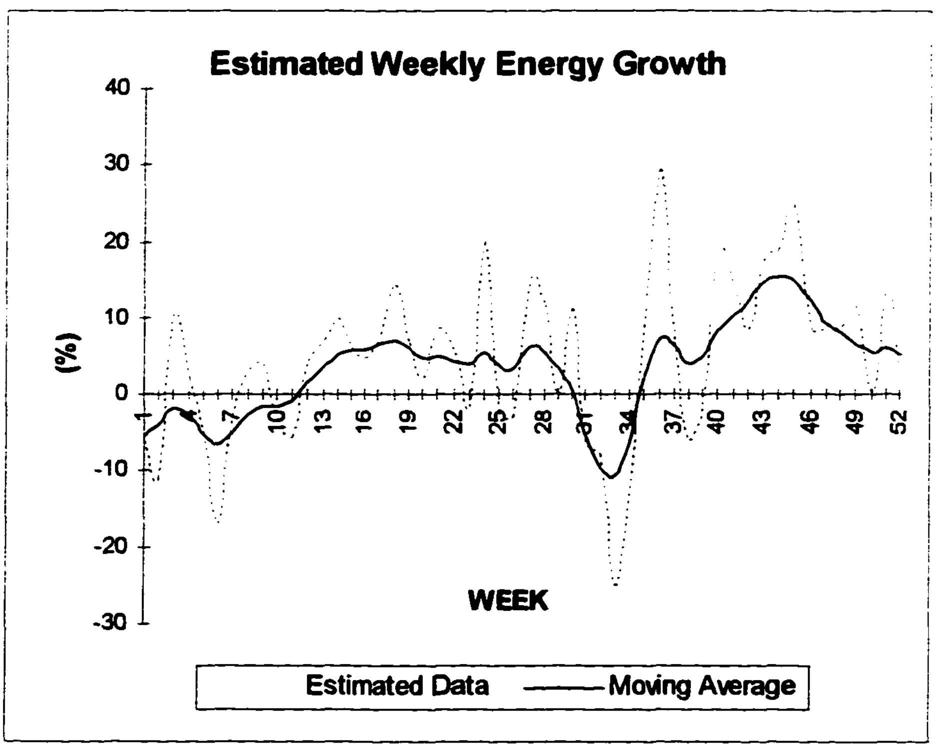

Figure 3-3

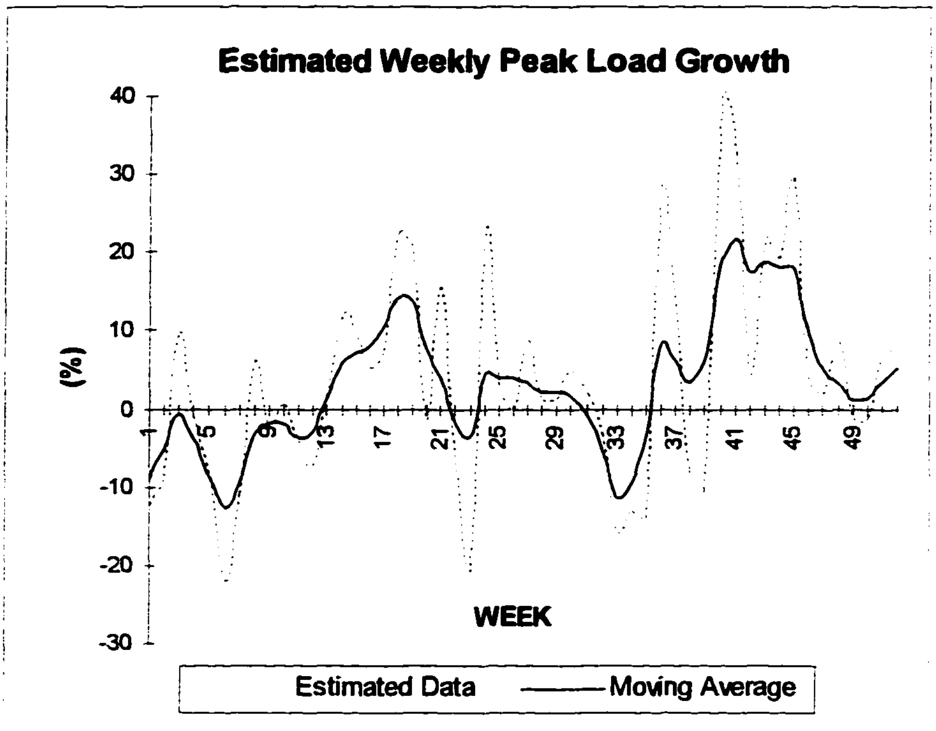

Figure 3-4

We observed that there are two major dips that occurs on week 6 and week 33 in the smoothed values of weekly energy growth in Figure 3-3. These negative weekly growths occur when there is a decreasing trend in the adjusted weekly energies of previous five years. The same situation is also found in the smoothed values of weekly peak load growth during the same weeks in Figure 3-4. Again, this implies the correlation between weekly energy and peak load is significant.

To examine how significant the correlation between weekly energy and peak load is, the correlation coefficients between energy and peak load of the same week are estimated. In Figure 3-5, the estimated weekly correlation coefficients between energy and peak load are smoothed by a moving average routine and plotted.

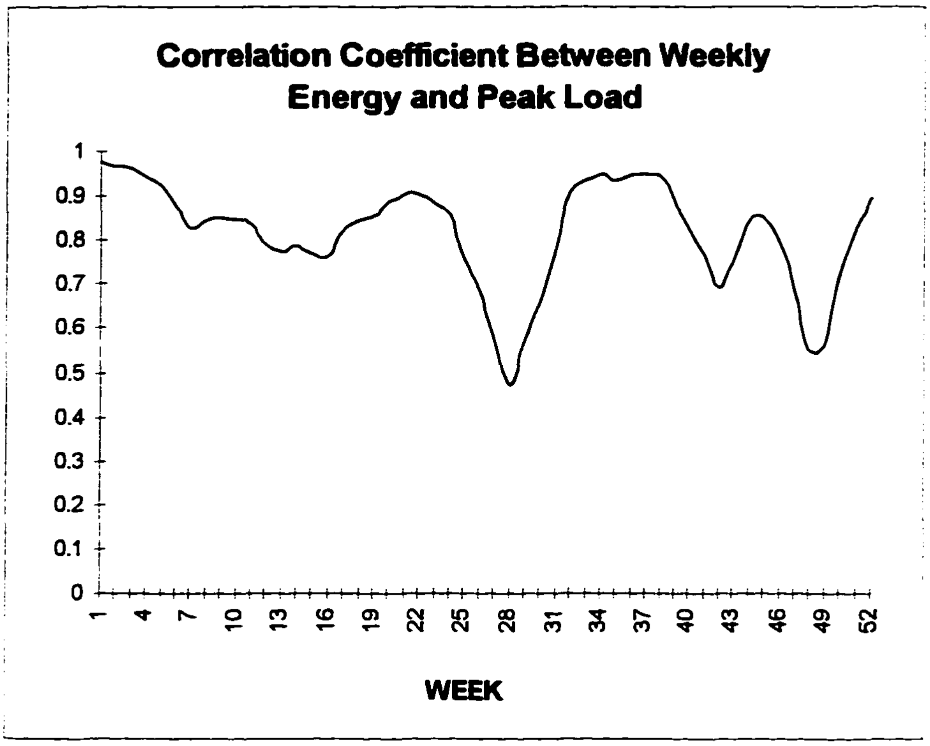

Figure 3-5

Figure 3-5 verifies that the weekly correlation coefficients between energy and peak load are significant during most of the weeks. Therefore, it is necessary to estimate the weekly correlation coefficients between energy and peak load when modeling the dependency between weekly energy and peak load.

In addition, we observed that the estimated annual energy standard deviation from the previous five years of load data is 439,754 MWH. However, the square root of the summation of the weekly energy variances is estimated to be 259,018 MWH which only contributes about half of the annual energy standard deviation. This indicates that the energy correlation between weeks also contributes a significant part of the annual energy variance. This analysis of the historical load data provides strong evidence to support modeling the correlation between the weekly energies. It also provides an estimation of the variance of annual energy which we expect the weekly sampling scheme to capture for the simulated system. Thus, the weekly energy correlation coefficients with one week lag are also estimated and shown in Figure 3-6.

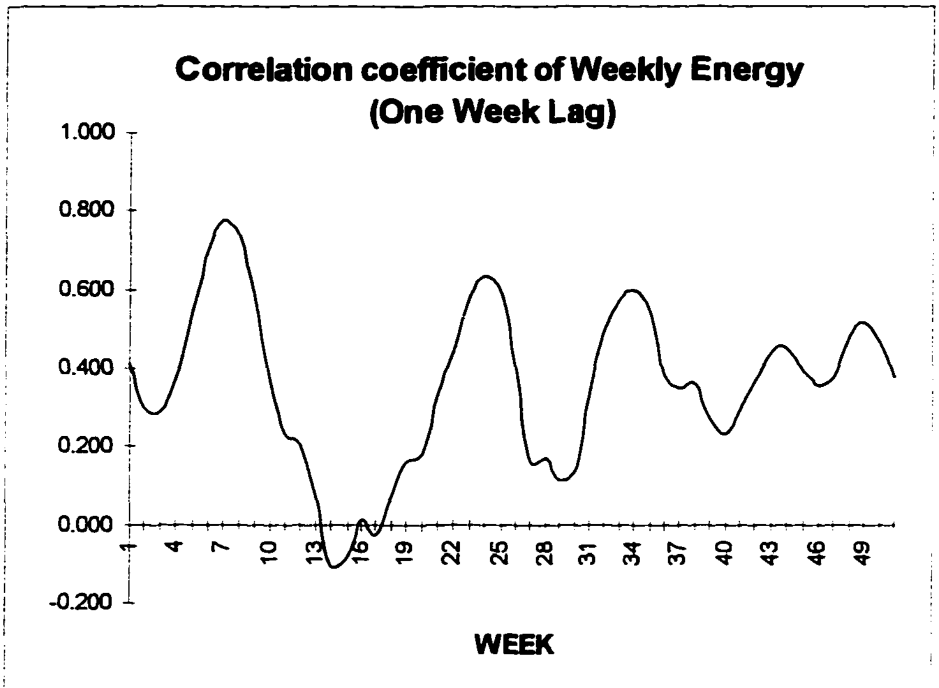

Figure 3-6

However, the variation of the weekly peak loads is less stable than that of weekly energy, because the weekly peak load is more likely to be affected by temperature. In the sampling of weekly peak loads, the weekly peak load correlation coefficients with one week lag are not considered.

III.2 Normal Sampling And Stratified Sampling

To evaluate the impact of the uncertainty of a random variable, which has infinite possible outcome values, is to carry out the "what-if" studies. Since sampling is the basis to perform hundreds or thousands of "what-if" scenarios, sampling techniques are essential tools for us to simulate the uncertainty of a random variable. Sampling is the process of drawing values randomly from a probability distribution associated with that random variable. The selection of a value from a probability distribution is called sampling and each sample represents a possible outcome. When enough samples have been drawn, the sampled values of a probability distribution become distributed in a manner which approximates the known distribution. Sampling is usually used in a simulation to generate possible values from probability distribution functions which are used to represent the characteristics of the random variables. Thus, sampling can be used to provide information about the population from which the sample has been selected. As a result, the behavior of a random variable can be simulated.

Since the range of outcomes is often associated with levels of probability of future occurrence, most people are interested in predicting the expected value of outcome variable and are also concerned about those situations which turn out to be not what they expected. In general case, we might want the analysis to be extended to include several special cases, such as "worst case" and "best case", in addition to the expected value. In this dissertation, a normal sampling technique is used to simulate the possible future events of a random variable which has a normal distribution. Stratified sampling is used to sample random variables which have pre-defined subgroups. This stratified sampling scheme provides a way to evaluate the extreme cases, such as: worst case and best case and to reduced the sample size with reduced sampling error.

III.2.1 Normal Sampling

Sampling of random variables has a variety of applications in statistical simulations. Specifically, we might want to use samples from a particular distribution to represent a random variable rather than the entire population, when the population is too large to process. For example, we might want to use normal distribution to characterize the population of weekly energies or weekly peak loads. Instead of using the entire population in simulations, normal samples are used to represent the possible outcomes of the normal random variable.

A normal random number generator is used to generate independent samples from a normal distribution. Normal sampling is a tool to draw independent samples from a normal distribution to represent the population of a random variable. Normal sampling can be accomplished for each variable by transforming a uniformly distributed random number generator into a random normal distribution. An example of the formulae to generate normally distributed random numbers is listed in Equation (3-4) [27].

$$X_1 = \sqrt{-2\ln(u_1)} * \sin 2\pi * u_2 \qquad (3\text{-}4)$$

where

u_1 and u_2 are two random numbers from a standard uniform distribution.

X_1 is a random number which has a standard normal distribution $N(0,1)$.

III.2.2 Stratified Sampling

Simple random sampling does not take into consideration any information that is known about the elements of a population. Under this sampling scheme, it is possible that a particular subgroup of the population would not be represented by a small number of samples. Therefore, stratified sampling [28, 29, 30] is used to avoid this kind of situation. Stratified sampling is a technique of sampling that draws independent samples from a set of mutually exclusive subgroups called strata. In stratified sampling, the whole population is divided into strata, and a simple random sampling can be used to draw samples from each stratum. This method ensures that each stratum is represented in the overall samples.

To use the stratified sampling, it requires not only to divide a distribution into a finite number of strata, but also to know the probability of each stratum. Therefore, to apply stratified sampling in drawing samples from the distribution of a random variable, it is essential that the subgroups of the whole population are pre-defined; this also implies that the sample size of the strata is known. To benefit most from stratified sampling, stratified sampling should be used when the population of a random variable to be sampled is similar within each stratum; however, relatively different among the strata. Thus, a more precise estimate of a stratum mean can be obtained from a small amount of samples within each stratum. These estimates can then be combined to form a more precise estimate for the entire population.

In our application, the implementation of stratified sampling is used to estimate the mean (μ) and variance (σ^2) of the overall population with finite samples. Suppose there are m different subgroups within the entire population and each subgroup has mean (μ_i) and variance (σ_i^2). We further define the ω_i as the fraction of the entire population in subgroup i, then $\sum_{i=1}^{m} \omega_i = 1$. In stratified sampling, we have the flexibility to select a sample rate for each subgroup and to obtain the estimations of the overall population by weighting the estimations of each subgroup. In the following sub-section, the estimations of the population mean and variance is illustrated [29].

We now consider the discrete case for simplicity, although the argument can also be made for the continuous case. Suppose there are only a total number of N possible outcome values of x in the whole population, and n_i values of x are within subgroup i. In addition, these subgroups are not overlapping, and together they comprise the whole population. Thus, we have $N = \sum_{i=1}^{m} n_i$ and $n_i/N = \omega_i$. To distinguish the difference between the true value and the estimated value of a variable, the estimated value is indicated by the hat symbol (^). Therefore, the estimation of population mean can be derived by the follow:

$$\begin{aligned}\hat{\mu} &= \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}\sum_{i=1}^{m}\sum_{l=1}^{n_i} x_l \\ &= \frac{1}{N}\sum_{i=1}^{m} n_i \sum_{l=1}^{n_i} \frac{x_l}{n_i} = \frac{1}{N}\sum_{i=1}^{m} n_i \hat{\mu}_i \\ &= \sum_{i=1}^{m} \frac{n_i}{N}\hat{\mu}_i = \sum_{i=1}^{m} \omega_i \hat{\mu}_i\end{aligned} \tag{3-6}$$

Note that the estimated means of all subgroups, $\hat{\mu}_i$, are weighted by the fraction, ω_i, to estimate the population mean. Also, the estimation of population variance is derived by:

$$\hat{\sigma}^2 = \sum_{i=1}^{N}(x_i - \mu)^2 p(x_i) \tag{3-7}$$

where $p(x_i)$ is the probability of the value x_i. The probability also can be expressed in terms of conditional probabilities:

$$p(x_i) = \sum_{j=1}^{m} p(x_i|j)p(j) = \sum_{j=1}^{m} p(x_i|j)\omega_j \tag{3-8}$$

where $p(x_i|j)$ is the conditional probability of the value x_i, given that x_i belongs to subgroup (or stratum) j. $p(j)$ is the probability that x_i belongs to subgroup j, so $p(j)$ is estimated by ω_j. Thus, we have $\sum_{i=1}^{N} p(x_i|j) = 1$ and $\sum_{i=1}^{N} x_i p(x_i|j) = \mu_j$ for each subgroup j. Hence, it follows from Equation (3-8) that Equation (3-7) can be expressed as:

$$\begin{aligned}
\hat{\sigma}^2 &= \sum_{i=1}^{N}(x_i - \mu)^2 p(x_i) \\
&= \sum_{i=1}^{N}(x_i - \mu)^2 \sum_{j=1}^{m} p(x_i|j)\omega_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{m} \omega_j (x_i - \mu)^2 p(x_i|j) \\
&= \sum_{j=1}^{m} \omega_j \left[\sum_{i=1}^{N}(x_i - \mu)^2 p(x_i|j)\right] \\
&= \sum_{j=1}^{m} \omega_j \left[\sum_{i=1}^{N}[(x_i - \mu_j) - (\mu_j - \mu)]^2 p(x_i|j)\right] \\
&= \sum_{j=1}^{m} \omega_j \left[\sum_{i=1}^{N}(x_i - \mu_j)^2 p(x_i|j) + 2(\mu_j - u)\sum_{i=1}^{N}(x_i - \mu_j)p(x_i|j) + (\mu_j - u)^2 \sum_{i=1}^{N} p(x_i|j)\right] \\
&= \sum_{j=1}^{m} \omega_j \left[\sum_{i=1}^{N}(x_i - \mu_j)^2 p(x_i|j) + 2(\mu_j - u)\{\sum_{i=1}^{N} x_i p(x_i|j) - \mu_j \sum_{i=1}^{N} p(x_i|j)\} + (\mu_j - \mu)^2\right] \\
&= \sum_{j=1}^{m} \omega_j \left[\sum_{i=1}^{N}(x_i - \mu_j)^2 p(x_i|j) + 2(\mu_j - u)\{\mu_j - \mu_j\} + (\mu_j - \mu)^2\right]
\end{aligned}$$

$$
\begin{aligned}
\hat{\sigma}^2 &= \sum_{j=1}^{m} \omega_j \left[\sum_{i=1}^{N} (x_i - \mu_j)^2 p(x_i|j) + (\mu_j - \mu)^2 \right] \\
&= \sum_{j=1}^{m} \omega_j \left[\sigma_j^2 + (\mu_j - \mu)^2 \right] \\
&= \sum_{j=1}^{m} \omega_j \sigma_j^2 + \sum_{j=1}^{m} \omega_j (\mu_j - \mu)^2
\end{aligned} \tag{3-9}
$$

As a result, the estimation of population variance can be separated into two parts. The first summand is the weighted average of the variances which is contributed by each subgroup. The second summand is the weighted average of the variance which is caused by the errors between the population mean and the mean of each subgroup. The aforementioned formulations suggest that the mean and variance of the whole population can be estimated based on a weighted average of those estimations of each subgroup.

How to divide the whole population into subgroups is dependent on different applications. Reference [31] states that a better choice of strata is to define the strata such that the variation of samples in each stratum of a random variable is small. In order to provide the expected cost at high load and low load scenarios in addition to mean load scenario, three subgroups are defined in the probabilistic approach. Based on a standard normal distribution and its symmetrical properties, a solution to allocate three subgroups can be found such that the conditional variances of each stratum are equal. An approximate solution is found in Reference [32] using a bisection method. The probabilities for these subgroups under a standard normal distribution are approximately $\{0.2, 0.6, 0.2\}$.

III.3 Chronological Load Projecting Algorithm

To estimate the cost variance induced by load variation using a chronological production cost simulator, it is necessary to reflect the load variation in the chronological load curves. To reflect the load uncertainty in the chronological load profiles, the annual (or weekly) energy and peak load are sampled according to their distributions in which the uncertainty is included. Then, the annual (weekly) load profiles are generated using a load projecting algorithm based on the sampled annual (weekly) energy and peak load. Consequently, the load variation represented by the estimated variations of energy and peak load is embedded in the resulting load profiles.

As we observed, the system load follows a quite consistent pattern from day to day, week to week and year to year. Although, the energy and peak load are sampled based on their distribution, the chronological patterns are also important to simulate the characteristics of an electrical power system. In this dissertation, an annual typical load profile is used to carry out the chronological patterns in the system load. A load projecting algorithm which converts this typical load curve to a new chronological load profile with the specified energy and peak load of a specified period (week or year) has been developed. Thus, the generated load scenarios will inherit the characteristic of the chronological pattern of the system load and the variation in system load.

We first generate a typical load curve by averaging the adjusted hourly loads of the previous five years to obtain an annual pattern of the system load profile. To preserve the continuity of chronological load at the first hour of March 1, 24 hourly load values are deleted at the end of the year, if a leap-year historical load profile is used in generating a typical load profile of a non-leap-year. On the other hand, 24 hourly loads are added at the end of the year, if a historical load profile of a non-leap year is used in generating a typical load profile of a leap year. Thus, the effect of leap year on creating a typical load profile is taken into account.

To average the adjusted hourly loads of the same weekday type is extremely important to maintain the weekly pattern in the process of creating a typical hourly load profile. Since the weekday type of January 1 of each year may not be the same, the weekday type of each day also should be considered when averaging the adjusted loads. Therefore, the loads of the adjusted load profiles are shifted to the appropriate weekday type dependent on the year of the historical load data and the year we want to study before averaging. Using these steps, a typical load profile with the historical chronological pattern can be generated.

In this load projecting algorithm, we presume there is a linear relation between the new load, $L_n(i)$, and typical load, $L_o(i)$, at hour i. The linear equation used to transform the typical load profile to a new load curve is defined as follow:

$$L_n(i) = a + b * L_o(i) \tag{3-10}$$

By providing the sampled energy and peak load of the defined horizon and the energy and peak load of the typical load profile of the same horizon, the coefficients, a and b, can be calculated using the following two equations

$$Peak_n = a + b * Peak_o \tag{3-11}$$

$$Energy_n = T * a + b * Energy_o \tag{3-12}$$

where

T	Number of hours in the defined period.
$Peak_o, Peak_n$	Typical and new peak loads in the defined period, respectively.
$Energy_o, Energy_n$	Typical and new energies in the defined period, respectively.

Consequently, a new hourly load profile, with the provided total energy and peak load, can be constructed using the calculated coefficients, a and b.

CHAPTER IV

TWO APPROACHES OF ESTIMATING COST VARIANCES

IV.1 Stochastic Approach

Load forecasting is a fundamental function in electric power system operation and planning. Many techniques and approaches on short-term load forecasting have been proposed in the last two decades [1-9]. However, to forecast an electrical load profile accurately for the long-term, such as an annual horizon, is almost impractical, because the most important forecast indicator, temperature forecast, is not reliable for such lengths of lead time. Another way to deal with the long-term load forecast is to forecast the weekly energy and peak load and reflect these forecast values into a weekly hourly load profile. Using this approach, we can either to forecast the expected values or to sample the values based on their distributions. Thus, incorporating the uncertainty in a chronological load profile is possible. In the stochastic approach, we model future loads by sampling weekly energies and weekly peak loads according to their distributions.

IV.1.1 Weekly Conditional Sampling

Suppose that both weekly energy and peak load are normally distributed. To sample the weekly energies and peak loads, the means and variances of the weekly energies and peak loads are required to describe their distributions. Also, the correlation coefficients between weekly energies with one week lag time and the weekly correlation coefficients between energy and peak load of each week are estimated and implemented in this conditional sampling scheme. Since 168 load values are included in weekly energy to provide an averaging and smoothing effect, the weekly energy data have much less variability than that of the weekly peak load. We will sample the weekly energy first, the weekly peak load will then be sampled conditionally on the sampled weekly energy during the same week. As we have observed, the correlation between weekly energies with one week lag is more significant than the correlation between weekly peak loads with one week lag. Even after adjusting the influences of weather, the weekly peak load series still has much variability. This is the reason that the correlation between weekly peak loads is not modeled in the weekly peak load modeling.

As described previously, the loads between year 1 and year 5 (i.e., the previous five years) were adjusted to a future (sixth) year based on the estimated weekly energy growths and peak load growths to account for the differences in the load growth. Thus, we have the adjusted future loads, which produce five future weekly energies and five future weekly peak loads for each week. The means and variances of the weekly energy and peak load are estimated using standard statistical procedures. Also, the correlation coefficient ($\hat{\rho}_{E_{i-1},E_i}$) between weekly energies of week i and week i-1 and the weekly correlation coefficient ($\hat{\rho}_{E_i,P_i}$) between energy and peak load of week i are estimated as follows:

$$\hat{\rho}_{E_{i-1},E_i} = \frac{\hat{\sigma}_{E_{i-1},E_i}}{(\hat{\sigma}^2_{E_i} * \hat{\sigma}^2_{E_{i-1}})^{1/2}} \tag{4-1}$$

$$\hat{\rho}_{E_i,P_i} = \frac{\hat{\sigma}_{E_i,P_i}}{(\hat{\sigma}^2_{E_i} * \hat{\sigma}^2_{P_i})^{1/2}} \tag{4-2}$$

where

$\hat{\sigma}_{E_{i-1},E_i}$ The estimated covariance between weekly energies of week i-1 and week i.

$\hat{\sigma}_{E_i,P_i}$ The estimated weekly covariance between energy and peak load of week i.

$\hat{\sigma}^2_{E_i}, \hat{\sigma}^2_{P_i}$ The estimated variances of energy and peak load of week i, respectively.

All the estimated values described above are also smoothed by applying a moving average routine.

The historical load data reveal that there is a significant weekly energy correlation that contributes to the annual energy variance, and the variation of weekly energy is more stable than that of the weekly peak load. Using a normal sampling procedure, the weekly energies and peak loads are sampled to represent the future loads in this stochastic load variation modeling. Based on the Gauss-Mark property of the weekly energy, the energy correlation with one week lag is modeled in the weekly energy sampling [33]. This conditional sampling scheme for weekly energy is presented as follows [33]:

Week 1:

$$E_1 = \hat{\mu}_{E_1} + \hat{\sigma}_{E_1} N_1 \tag{4-3}$$

Week 2~52

$$E_i = \hat{\mu}_{E_i} + \hat{\rho}_{E_{i-1}.E_i} \frac{\hat{\sigma}_{E_i}}{\hat{\sigma}_{E_{i-1}}}(E_{i-1} - \hat{\mu}_{E_{i-1}}) + \hat{\sigma}_{E_i}(1 - \hat{\rho}^2_{E_{i-1}.E_i})^{1/2} N_2 \quad (4\text{-}4)$$

where

E_i, P_i The sampled energy and peak load of week i, respectively.

$\hat{\mu}_{E_i}, \hat{\mu}_{P_i}$ The estimated weekly mean energy and mean peak load of week i, respectively.

N_1, N_2, N_3 The independent samples of standard normal distribution.

The energy for week one is sampled based on its mean and variance only. For week 2 to week 52, the weekly energy correlation with the previous week is also modeled. The first and second summand in Equation (4-4) represent the conditional (upon E_{i-1}) mean, while the third summand is the random conditional variation from the conditional mean.

The weekly peak load is also modeled as a normal distribution with the estimated mean and variance. In addition, if the weekly correlation coefficient between energy and peak load is significant (i.e., greater than 0.5), the correlation between weekly energy and peak load is included in the weekly peak load sampling. The weekly peak load is sampled as follows:

if $\hat{\rho}_{E_i.P_i} \geq 0.5$

$$P_i = \hat{\mu}_{P_i} + \hat{\rho}_{E_i.P_i} \frac{\hat{\sigma}_{P_i}}{\hat{\sigma}_{E_i}}(E_i - \hat{\mu}_{E_i}) + \hat{\sigma}_{P_i}(1 - \hat{\rho}^2_{E_i.P_i})^{1/2} N_3 \quad (4\text{-}5)$$

else

$$P_i = \hat{\mu}_{P_i} + \hat{\sigma}_{P_i} N_3 \tag{4-6}$$

The weekly energies and peak loads of the entire year are sampled using the proposed conditional sampling scheme. Then, given the sampled weekly energies and peak loads, an associated weekly chronological load curve is generated using the linear load projecting algorithm. Notice that weekly energies and peak loads are randomly sampled based on their normal distributions, which may cause large variability in the sampled values from week to week. As a result, the change from a week with a high sample value to a week with a low sample value will create a huge discontinuities during the transition between weeks, which is not likely to happen in real situations. To use the annual chronological load curve composed of these weekly curves, the discontinuities that occur at the beginning of each week are adjusted by applying a moving average routine to the loads of the first three hours of each week. One example of the sampled energies and peak loads is plotted in Figure 4-1, included are the expected weekly energies and peak loads for the purpose of comparison.

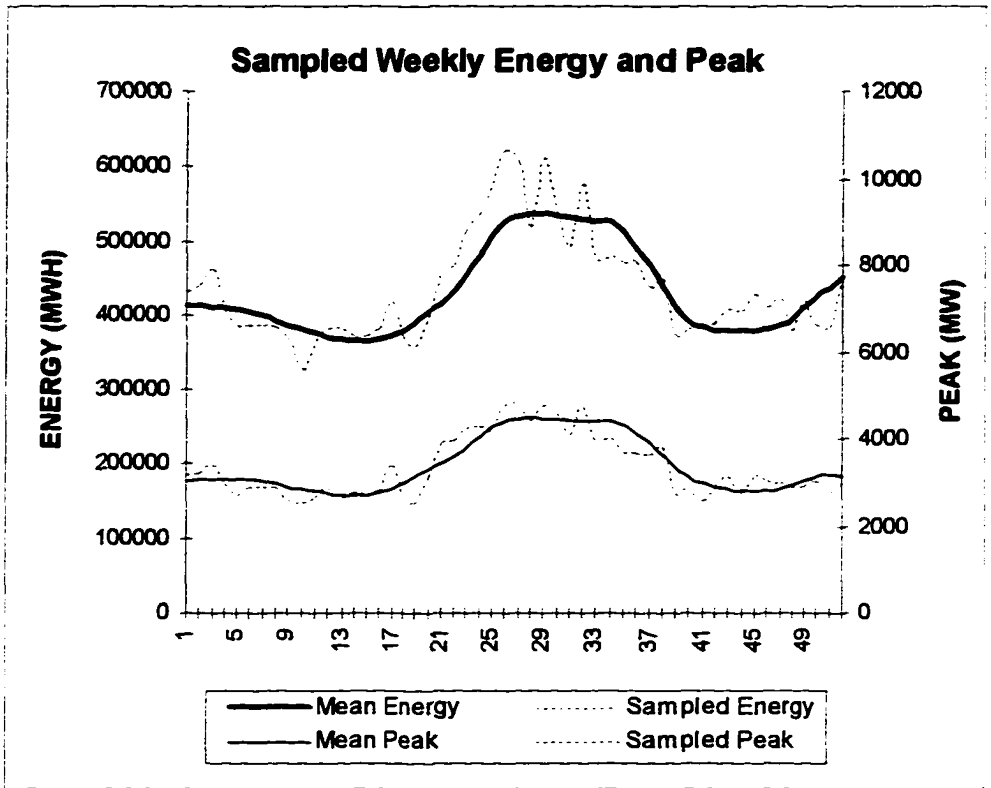

Figure 4-1

IV.1.2 Cost Variances Estimation

Using this stochastic approach to sample the load profiles, the variance of the annual production cost caused by load variation can be simulated based on the estimated load variation of the simulated system. The generated load profiles including the expected annual load curve are simulated using a chronological production cost simulator. The number of simulations in which the estimated variance of annual cost will converge is usually dependent on the simulated system. Instead of specifying a fixed number of simulations, the simulations are continued until the estimate of the variance of cost becomes stable. The variance of cost will be re-estimated as an indication of convergence when an additional annual simulation is completed. If the change of the variance of cost is within the specified tolerance criteria, say 5%, for five consecutive simulations, then no more simulation is required. The mean and variance of production cost will be estimated at this point.

Since the annual load variation is modeled by a weekly sampling scheme, several load profiles should be sampled and simulated until the sampled annual load variation is comparable to the system annual load variation. The convergence criterion, which requires the change of estimated cost variance within 5% error bounds for five consecutive runs, does not guarantee that the simulated annual load variation is comparable with the system annual load variation. Thus, we need to impose an additional criterion on annual energy to ensure the agreement of annual load variation ,because the annual energy has a dominant effect on the production cost. The simulations should be continued until the mean and standard deviation of the sampled annual energy are both within 5% error bounds of the mean and standard deviation of system annual energy as well.

To estimate the variance of cost induced by unit availability uncertainty using a chronological production cost simulator, an expected annual load profile is used. In the simulations, different random number seeds are selected to generate different unit outage patterns. In this case, the estimated mean and standard deviation of annual production cost are caused by the only uncertain factor, unit outage. On the other hand, the variance of cost induced by load variation uncertainty is simulated using different sampled load profiles with a fixed random number seed. In this way, the unit outage patterns remain the same in each simulation; the load variation is the only uncertain factor that impacts the production cost. In this case, we assume the effects of different fixed random number seed is not significant by presuming the cost variance induced by unit outage uncertainty is much smaller than that caused by load variation uncertainty.

In this dissertation, the uncertainties in load variation and unit outage are the only two uncertain factors that influence the variance of production cost. Thus, the total cost variance, which is caused by both load variation and unit outage, can be roughly approximated based on the estimated cost variances caused by each individual uncertain factor. The expected costs estimated in the previous two study cases are close, so we can estimate the total cost variance to be the summation of the two cost variances induced by each individual uncertain factor. However, to estimate the cost variance induced by both uncertain factors more accurately, a set of simulations was made by varying both the random number seed and sampled load curve. Then, the mean and variance of cost induced by both uncertain factors can be obtained.

IV.2 Probabilistic Approach

IV.2.1 Stratified Load Scenarios Sampling

The idea of the probabilistic approach is to simulate the production cost based on the loads of the three pre-defined load scenarios and to estimate the variance of cost according to the probability of each load scenario. Usually, enough samples are required to be sampled from each subgroup to describe the distribution of the entire population when using a simple random sampling. Thus, a large number of total samples will result in long simulation time. However, if intelligently used, a stratification sampling always can result in a smaller variance for the estimated mean and variance than that is given by a comparable simple random sampling. This encourages us to implement stratified sampling in the load variation modeling, and thus reduce the simulation time. In the probabilistic approach, the annual energies and peak loads of the pre-defined load scenarios are sampled using a stratified sampling scheme.

Implied by the Central Limit Theorem is that annual energy and peak load, which are determined by a large number of independent causes, tend to have a normal probability distribution function. In addition, the dependency of these two random variables suggests that a bivariate normal distribution is an ideal distribution to model the two normal random variables with correlation. A joint normal distribution is used to describe the annual energy and peak load of the high, mean, and low load scenarios; this is accomplished by partitioning the joint normal distribution into three strata based on stratified sampling techniques [28, 29, 30]. The bivariate joint normal density function describes a bell-shape surface above the $(x_1,\ x_2)$ plane. Given a joint normal distribution $N(0,0;\sigma_1,\sigma_2,\rho_{1,2})$, a contour plot of equal density of the bivariate normal distribution with the three pre-defined strata is shown in Figure 4-2. They are high load (A1), mean load (A2), and low load (A3) scenarios. The major axis of the ellipse make an angle of $\theta=\frac{1}{2}\tan^{-1}\left[2\rho_{1,2}\sigma_1\sigma_2/(\sigma_1^2-\sigma_2^2)\right]$ with respect to the x_1 axis. Note that this angle is 45°, if $\sigma_1=\sigma_2$ and $\rho_{1,2}>0$, regardless of the numerical value of $\rho_{1,2}$. Figure 4-2 also shows that the ellipse is symmetric about $\sigma_2x_1-\sigma_1x_2=c$, the major axis.

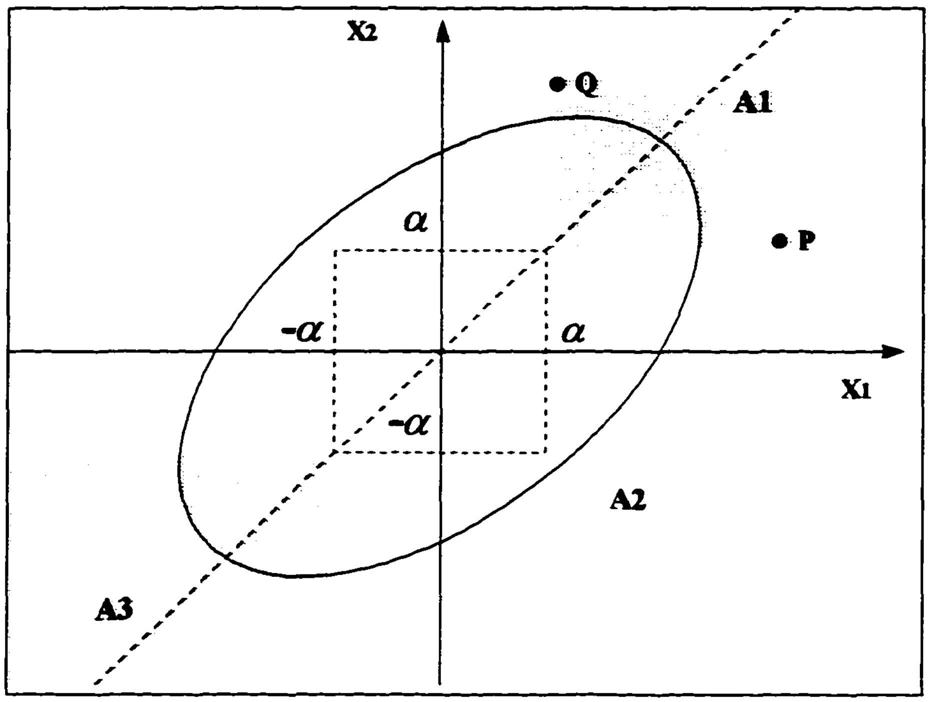

Figure 4-2 Stratified Strata of A Bivariate Normal Distribution

When sampling a random variable based on its distribution, a stratified sampling technique is introduced to reduce the sample size of each stratum without diminishing the precision of the estimate. The high load scenario is pre-defined such that the annual energy and peak load within this stratum are both high. Using a small number of samples to represent a stratum, the points on the edges of area A1, such as P and Q, will not be appropriate to stand for that stratum. A typical sample which is selected from those points on the major axis within a stratum is a good candidate to represent the population of that stratum. To simplify the analysis, a special case of interest is the bivariate normal distribution with mean 0 and variances 1 and a positive correlation coefficient $\rho_{1,2}$. Given a joint normal distribution $X \sim N(0,0;1,1,\rho_{1,2})$ described as in Equation (4-7), the major axis of the ellipse is $x_1 = x_2$.

$$f_{x_1,x_2}=\frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}}e^{\frac{-1}{2(1-\rho_{1,2}^2)}(x_1^2-2\rho_{1,2}x_1x_2+x_2^2)} \tag{4-7}$$

The goal of our formulation is to find the expected value with respect to x_1 and x_2, conditional on the major axis. This can be accomplished by formulating the conditional density $f_{x_1|z}$, where we express the statistics of Z in terms of the function $z = x_1 - x_2$. By defining $z = x_1 - x_2$, the system $x_1 = x_1$ and $z = x_1 - x_2$ has a single solution: $x_1 = x_1$, $x_2 = x_1 - z$. Since the jacobian, $J(x_1, x_2) = \begin{vmatrix} 1 & 0 \\ 1 & -1 \end{vmatrix}$, of the transformation equals -1, the transformation [34] yields:

$$\begin{aligned}
f_{x_1,z} &= |J^{-1}| f_{x_1,x_2}(x_1, x_1 - z) \\
&= \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}}e^{-\frac{1}{2(1-\rho_{1,2}^2)}[x_1^2-2\rho_{1,2}x_1(x_1-z)+(x_1-z)^2]} \\
&= \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}}e^{-\frac{1}{2(1-\rho_{1,2}^2)}[2(1-\rho_{1,2})x_1^2-2(1-\rho_{1,2})x_1z+z^2]} \\
&= \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}}e^{-\frac{1}{(1-\rho_{1,2}^2)}[(1-\rho_{1,2})x_1^2-(1-\rho_{1,2})x_1z+\frac{(1-\rho_{1,2})z^2}{4}]} * e^{-\frac{1}{2(1-\rho_{1,2}^2)}(z^2-\frac{(1-\rho_{1,2})z^2}{4})} \\
&= \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}}e^{-\frac{1}{(1+\rho_{1,2})}(x_1-\frac{z}{2})^2} * e^{-\frac{z^2+\rho_{1,2}z^2}{4(1-\rho_{1,2}^2)}} \\
&= \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}}e^{-\frac{1}{(1+\rho_{1,2})}(x_1-\frac{z}{2})^2} * e^{-\frac{z^2}{4(1-\rho_{1,2})}}
\end{aligned} \tag{4-8}$$

Hence, the density function of z can be determined by integrating (4-8) respect to x_1 from $-\infty$ to ∞ as follows:

$$
\begin{aligned}
f_z(z) &= \int_{-\infty}^{\infty} f_{x_1,z}(x,z)\,dx_1 \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}} e^{-\frac{1}{(1+\rho_{1,2})}(x_1-\frac{z}{2})^2} * e^{-\frac{z^2}{4(1-\rho_{1,2})}} dx_1 \\
&= e^{-\frac{z^2}{4(1-\rho_{1,2})}} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}} e^{-\frac{1}{(1+\rho_{1,2})}(x_1-\frac{z}{2})^2} dx_1
\end{aligned}
\tag{4-9}
$$

Define $u = \sqrt{\frac{2}{1+\rho_{1,2}}}(x_1 - \frac{z}{2})$ and $du = \sqrt{\frac{2}{1+\rho_{1,2}}}dx_1$; Equation (4-9) yields

$$
\begin{aligned}
f_z(z) &= e^{-\frac{z^2}{4(1-\rho_{1,2})}} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}} e^{-\frac{u^2}{2}} \sqrt{\frac{1+\rho_{1,2}}{2}}\,du \\
&= e^{-\frac{z^2}{4(1-\rho_{1,2})}} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{2(1-\rho_{1,2})}} e^{-\frac{u^2}{2}} du \\
&= \frac{1}{2\sqrt{\pi(1-\rho_{1,2})}} e^{-\frac{z^2}{4(1-\rho_{1,2})}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\
&= \frac{1}{2\sqrt{\pi(1-\rho_{1,2})}} e^{-\frac{z^2}{4(1-\rho_{1,2})}}
\end{aligned}
\tag{4-10}
$$

Therefore, the conditional density function, $f_{x_1|z}$, can be derived as follows:

$$
\begin{aligned}
f_{x_1|z}(x_1|z) &= f_{x_1,z}(x_1,z) \Big/ f_z(z) \\
&= \frac{\frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}} e^{\frac{-(x_1-z)^2}{(1+\rho_{1,2})}} * e^{\frac{-z^2}{4(1-\rho_{1,2})}}}{\frac{1}{2\sqrt{\pi(1-\rho_{1,2})}} e^{\frac{-z^2}{4(1-\rho_{1,2})}}} \\
&= \frac{1}{\sqrt{\pi(1+\rho_{1,2})}} e^{\frac{-(x_1-z)^2}{(1+\rho_{1,2})}}
\end{aligned}
\tag{4-11}
$$

Thus, the bivariate normal distribution is transformed into a scalar normal distribution based on the given condition. So far, we have obtained a conditional density function from which the samples of each scenario can be sampled. Thus, we assign 20% of the distribution $f_{x_1|z=0}$ to the high load scenario, such that $\int_{\alpha}^{\infty} f_{x_1|z=0} dx_1 = 0.2$. This particular α can be look up from a standard normal distribution table after applying a scaling transformation on x_1. Thus, if we define $t = \sqrt{\frac{2}{1+\rho_{1,2}}} x_1$ and $dt = \sqrt{\frac{2}{1+\rho_{1,2}}} dx_1$, the resulting transformation is described as follows:

$$\begin{aligned} &\int_{\alpha}^{\infty} \frac{1}{\sqrt{\pi(1+\rho_{1,2})}} e^{\frac{-x_1^2}{(1+\rho_{1,2})}} dx_1 \\ &= \int_{\alpha\sqrt{\frac{2}{1+\rho_{1,2}}}}^{\infty} \frac{1}{\sqrt{\pi(1+\rho_{1,2})}} e^{\frac{-t^2}{2}} \sqrt{\frac{1+\rho_{1,2}}{2}} dt \\ &= \int_{\alpha\sqrt{\frac{2}{1+\rho_{1,2}}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt = 0.2 \end{aligned} \tag{4-12}$$

Using one of the tables of the standard normal distribution, we found that $\alpha\sqrt{\frac{2}{1+\rho_{1,2}}} \cong 0.85$. Thus $\alpha \cong 0.85\sqrt{\frac{1+\rho_{1,2}}{2}}$.

The conditional mean respect to x_1 can be calculated according to Equation (4-13).

$$\begin{aligned} E[x_1 | x_1 \geq \alpha, x_1 - x_2 = 0] &= \frac{1}{0.2}\int_{\alpha}^{\infty} \frac{x_1}{\sqrt{(1+\rho_{1,2})\pi}} e^{\frac{-x_1^2}{(1+\rho_{1,2})}} dx_1 \\ &= \frac{1}{0.2} * \frac{-1}{2}\sqrt{\frac{1+\rho_{1,2}}{\pi}} e^{\frac{-x_1^2}{(1+\rho_{1,2})}} \Big|_{\alpha}^{\infty} \\ &= 2.5 * \sqrt{\frac{1+\rho_{1,2}}{\pi}} (-e^{-\infty} + e^{\frac{-\alpha^2}{(1+\rho_{1,2})}}) \\ &= 2.5\sqrt{\frac{1+\rho_{1,2}}{\pi}} e^{\frac{-\alpha^2}{(1+\rho_{1,2})}} \end{aligned} \tag{4-13}$$

For the case $\rho_{1,2} = 0.9$, the conditional mean can be calculated as follows:

$$
\begin{aligned}
E\left[x_1 \middle| x_1 \geq \alpha, x_1 - x_2 = 0\right] &= 2.5\sqrt{\frac{1+\rho_{1,2}}{\pi}}e^{\frac{-\alpha^2}{(1+\rho_{1,2})}} \\
&= 2.5\sqrt{\frac{1+0.9}{\pi}}e^{-\frac{0.85^2(1+0.9)}{2(1+0.9)}} \\
&= 1.35
\end{aligned}
\tag{4-14}
$$

Now, we use the estimated conditional expected values to describe three scenarios. 20% of the area under the conditional density, $f_{x_1|z=0}$, is dedicated to the high load scenario, 60% to the mean load scenario and 20% to the low load scenario. This allocation of the conditional distribution is shown by Figure 4-3.

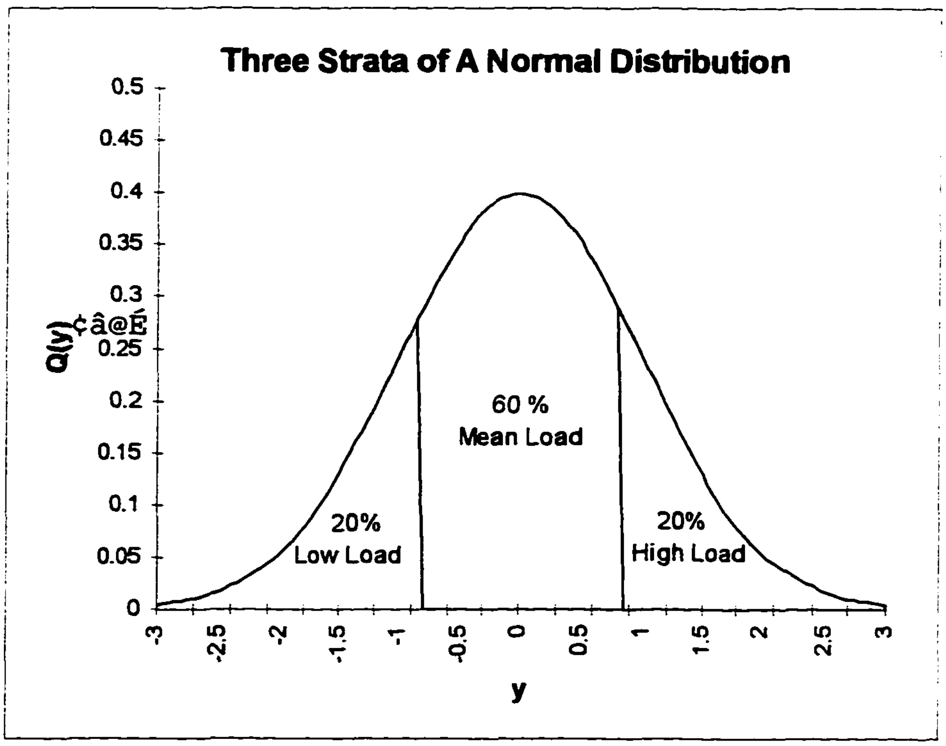

Figure 4-3

Then, the conditional (upon $\int_{\alpha}^{\infty} f_{x_1|z=0} dx_1 = 0.2$) mean with respect to x_1 is used to describe the annual energy of the high load scenario. When $\int_{\alpha}^{\infty} f_{x_1|z=0} dx_1 = 0.2$ and $\rho_{1,2} = 0.9$ are defined, α is found to be 0.83. Therefore, the conditional mean with respect to x_1 of the high load scenario can be calculated as follows:

$$E\left[x_1 \middle| x_1 \geq 0.83, x_1 - x_2 = 0\right] = \frac{1}{0.2} \int_{0.83}^{\infty} \frac{x_1}{\sqrt{1.9\pi}} e^{\frac{-x_1^2}{1.9}} dx_1 = 1.35 \qquad (4\text{-}16)$$

Similarly, the conditional means of the mean load and low load scenarios are estimated to be 0 and -1.35 respectively. Since the joint normal density function is symmetric about $x_1 = x_2$, the conditional mean with respect to x_1 is equal to the conditional mean with respect to x_2. Therefore, the conditional means of the annual peak load (x_2) of the three scenarios are also calculated to be 1.35, 0 and -1.35 respectively. Consequently, these conditional means of each stratum can be used to describe how far away the annual energy and peak load should deviate from their means of the annual energy and peak load. As a result, the annual energy and peak load of the high load scenario are sampled as follows:

$$H_E = \hat{\mu}_E + 1.35 * \hat{\sigma}_E \qquad (4\text{-}17)$$

$$H_p = \hat{\mu}_p + 1.35 * \hat{\sigma}_P \qquad (4\text{-}18)$$

where

$\hat{\mu}_E, \hat{\mu}_P$ The estimated means of the adjusted future annual energy and annual peak.

$\hat{\sigma}_E, \hat{\sigma}_P$ The estimated standard deviations of the adjusted future annual energy and annual peak.

Similarly, the annual energies and peak loads of the "mean load" and "low load" cases are

$$M_E = \hat{\mu}_E \quad (4\text{-}19)$$

$$M_P = \hat{\mu}_P \quad (4\text{-}20)$$

$$L_E = \hat{\mu}_E - 1.35 * \hat{\sigma}_E \quad (4\text{-}21)$$

$$L_p = \hat{\mu}_p - 1.35 * \hat{\sigma}_P \quad (4\text{-}22)$$

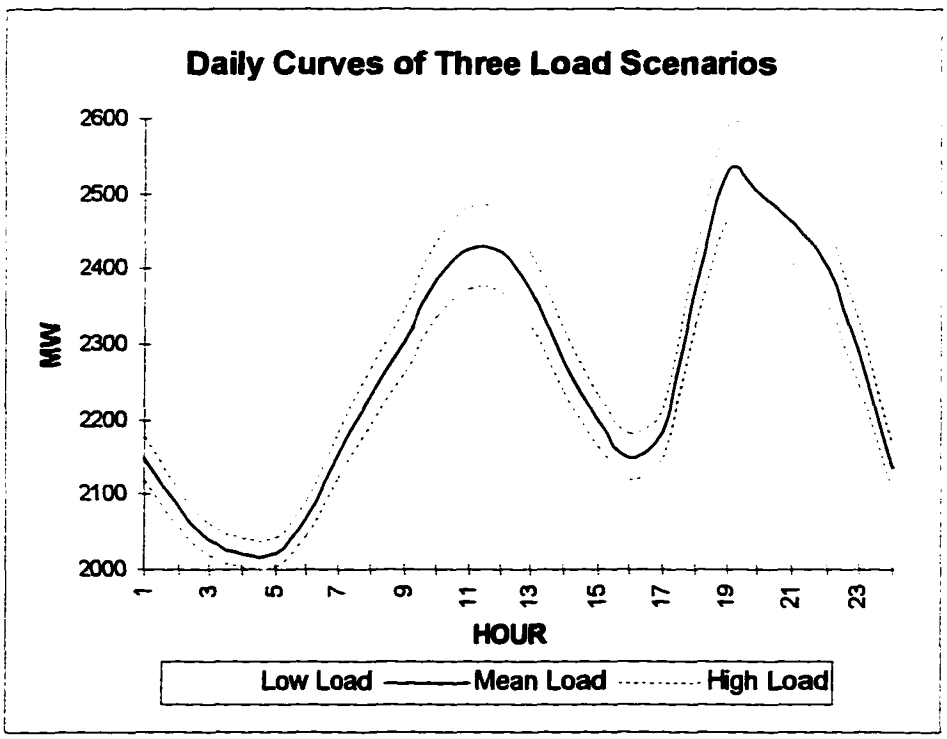

Figure 4-4

According to a typical load profile and the sampled annual energies and peak loads, three load scenarios (high load, mean load, and low load) are created by using the load projecting algorithm described in Section III.3 and are used in a chronological production cost simulator. A set of winter daily load curves of these three scenarios is illustrated in Figure 4-4.

IV.2.2 Cost Variances Estimation

We assume that fifteen annual simulations, five runs for each load scenario, will be required to estimate the cost variation due to uncertainty in unit outages and load. In each load scenario, the variance is induced by different random forced outage patterns. The population variance due to unit outage uncertainty is the weighted average of the individual variances of each stratum. Thus, the variance of cost due to the unit random forced outages can be estimated as follows:

$$\hat{\sigma}_u^2 = 0.2 * \frac{\sum_{i=1}^{n}(H_i - \overline{H})^2}{n-1} + 0.6 * \frac{\sum_{i=1}^{n}(M_i - \overline{M})^2}{n-1} + 0.2 * \frac{\sum_{i=1}^{n}(L_i - \overline{L})^2}{n-1} \tag{4-23}$$

where

n	The sample size. (n=5 in this case)
H_i, M_i, L_i	The costs of the high load, mean load, and low load for run i respectively.
$\overline{H}, \overline{M}, \overline{L}$	The expected costs of the high load, mean load, and low load respectively.

In each stratum, compared to cost variation due to load uncertainty, we assume that the cost variation caused by the uncertainty in unit outage is small. That is the estimated data vary little within a stratum. The variance of cost due to load uncertainty is evaluated according to the same fifteen simulated production costs previously obtained. The population mean of the production cost is first calculated as follows:

$$\hat{\mu} = 0.2 * \overline{H} + 0.6 * \overline{M} + 0.2 * \overline{L} \qquad (4\text{-}24)$$

We assume the unit uncertainty is averaged out in calculating the expected cost of the three load scenarios. The expected cost of each load scenario varied with different load profile. Thus, the variance of cost caused by load variation can be estimated by the variance of the estimated means of all the scenarios within the whole population. This is formulated as the following equation:

$$\hat{\sigma}_l^2 = 0.2 * (\hat{\mu} - \overline{H})^2 + 0.6 * (\hat{\mu} - \overline{M})^2 + 0.2 * (\hat{\mu} - \overline{L})^2 \qquad (4\text{-}25)$$

By implementing stratified sampling, the estimation of population variance ($\hat{\sigma}_{Total}^2$) can be estimated based on Equation (3-9), which is the sum of the weighted average of the variance of each stratum (i.e. $\hat{\sigma}_u^2$) and the variance of the means of the three pre-defined strata (i.e. $\hat{\sigma}_l^2$).

$$\hat{\sigma}_{Total}^2 = \hat{\sigma}_u^2 + \hat{\sigma}_l^2 \qquad (4\text{-}26)$$

CHAPTER V

PRODUCTION COST VARIANCES AND COMPARISIONS

The effects of load variation on annual production cost are measured in terms of the variance of cost. In this chapter, both the stochastic and the probabilistic approaches are applied using the chronological production cost simulator, *Scheduler*, to estimate the variance of annual production cost of a medium size utility. System A is used, the capacity of which is roughly 6,000 MW; the system spinning reserve is set to 500 MW. The generation unit data of system A is summarized in Appendix A. In addition to the generation unit data of this system, historical load profiles of the latest five contiguous years are used as input for analyzing the variation in system load. The two uncertain factors, which are considered in this research, are the unit outage uncertainty and load variation uncertainty. First, the variance of the production cost caused by each individual factor is estimated. Then, the variance of the production cost induced by both uncertain factors is estimated. To conclude this chapter, the simulation results of both approaches are compared.

V.1 Example of the Stochastic Approach

V.1.1 Cost Variance Due to Unit Outage Uncertainty

A common use for the production cost simulation is the estimation of the expected cost of the power system over an extended period of time. In the stochastic approach, an expected load profile is used to estimate the expected value and the variance of annual production cost caused by the uncertainty in unit outages alone. The effects of unit outage uncertainty are simulated by varying the random number seed in the Monte Carlo simulations. Since the required number of simulations is not specified in the stochastic approach, the simulations are continued until the estimate of the variance of cost converges in order to obtain a reasonable estimation. That means the variance of cost is obtained when the change of the estimated variance of cost is within a 5% tolerance criterion for five consecutive simulations. The simulated results, based on the expected load profile with different random number seeds, are summarized in Table 5-1.

Table 5-1 Stochastic Results Due To Unit Outage Uncertainty

Random Seed	Annual Production Cost ($)	STDEV of Cost	Cost Variance Change (%)
1	255210784	-	-
2	253439648	1252382	-
3	254414432	887065	-99.33
4	252656976	1115962	36.82
5	251390912	1491271	44.00
6	252298304	1410591	-11.77
7	251303808	1480210	9.15
8	250352976	1651397	19.66
9	254189136	1629454	-2.72
10	251585216	1584051	-5.81
11	252350816	1506122	-10.62
12	251346912	1484768	-2.89
13	251882144	1433396	-7.30
14	249827152	1550672	14.55
15	254227472	1574675	3.03
16	251561328	1536766	-4.99
17	251420112	1505971	-4.13
18	251478816	1474432	-4.32
19	251609200	1440991	-4.69

In this case, the estimation of the variance of cost reaches a stable state at run #19. From Table 5-1, the mean and standard deviation of the production cost induced by the unit outages uncertainty alone are estimated to be $252,239,271 and $1,440,991, respectively. The coefficient of variation of cost is approximately 0.57 %.

V.1.2 Cost Variance Due to Load Variation Uncertainty

If a decision maker wants to know how significant the load variation uncertainty will impact the production cost, the load variation should be the uncertain factor that varies in the simulations. For the purposes of estimating the variance of cost caused by load variation alone, a fixed random number seed, "1", is arbitrarily selected and different load profiles are used in the simulations of Table 5-2. Thus, assuming that the estimated cost variance due to load variation uncertainty is more significant than that due to unit outage uncertainty, we can estimate the variance of cost induced by load variation based alone on an arbitrarily selected random number seed. However, the convergence criterion, which requires the change of estimated variance of cost to be within 5% for five consecutive runs, does not guarantee that the simulated annual load variation is comparable with the system annual load variation. Therefore, an additional criterion was imposed on the annual energy to ensure the comparability of annual load variation, because the annual energy has a more dominant effect on the production cost than annual peak load does. The simulations should be continued until the mean and standard deviation of the estimated annual energy of the simulated load profiles are both within the 5% error bounds of the system annual energy as well. The necessity of this additional criterion is also demonstrated using the results of the stochastic approach in Table 5-2.

Table 5-2 Stochastic Results Due To Load Variation Uncertainty

Load Curve ID	Annual Energy (MWH)	Cost ($)	Cost Variance Change %	Coefficient of Variation of Cost %
1	22705996	246194656		
2	23813079	264214160	-	-
3	23101752	251892112	-91.4	3.62
4	22950312	251264544	-44.9	3.02
5	23000461	251976800	-32.1	2.63
6	23621917	260457104	0.3	2.62
7	22997089	251411984	-16.2	2.44
8	22814032	248996736	-6.8	2.63
9	23178872	254237536	-13.9	2.21
10	22926674	251213392	-10.6	2.11
11	23230723	255083136	-9.7	2.01
12	23173320	254464896	-9.5	1.92
13	22909122	249743728	-4.0	1.88
14	23888278	253981625	24.9	2.17
15	23101319	252666928	-7.3	2.09
16	22673799	246056224	6.5	2.17
17	22923532	249988032	-4.1	2.13
18	23297418	256056736	-4.5	2.08
19	23087990	251238480	-4.9	2.03
20	23056694	251437264	-4.9	1.98
21	24097126	249322736	-2.3	1.96
22	23742451	261582352	8.2	2.05
23	22901736	268460832	24.4	2.34
24	23348134	257339328	0.0	2.34
25	22492190	244708240	2.9	2.38
26	23862874	263074848	4.8	2.44
27	23135635	254813200	-4.0	2.39
28	22506714	248215968	-0.3	2.39

In Table 5-2, even though the change of cost variance satisfies the convergence criteria of cost variance at run # 21, the standard deviation of the estimated annual energy ($\hat{\sigma}_E$ =321508) of the first 21 simulated load profiles is not within the 5% error bounds of the standard deviation (σ_E=439,754) of the system annual energy. This small amount of annual energy variation results a small coefficient of variation of cost. The simulations should be continued until a set of comparable mean and standard deviation of annual energy ($\hat{\mu}_E$=23162117, $\hat{\sigma}_E$ =418800) are obtained at run #28. Then, the mean and standard deviation considering load variation alone are statistically estimated according to the costs summarized in Table 5-2. They are $252,976,626 and $4,967,378 respectively. The estimated coefficient of variation of cost is 2.39%.

In the simulations of Table 5-2, the cost variance caused by load variation uncertainty is estimated based on an arbitrary selected random number seed. There is no doubt that different unit outage patterns will result in different estimations of the variance of production cost, even though the unit outage pattern of each simulation is unchanged in each case. We expect to estimate the variance of cost, which is induced by load variation alone, by using a fixed unit outage pattern. This estimation will make sense only when the estimated variance of cost due to load variation uncertainty is more significant than due to the unit outage uncertainty. The simulated results of Table 5-1 and Table 5-2 confirm the previous assumption that cost variance due to load variation is more significant than that due to unit outage uncertainty. A related question is how significant the difference is in estimating cost variance based on different pre-selected unit outage patterns. The following case study illustrates how significant the estimated variance of cost induced by load variation alone will be affected by different pre-selected unit outage patterns.

In the following study, different random number seeds are selected for different cases. However, the same random number seed is used for all simulations in each case. Thus, consider the effects of load variation uncertainty alone, the difference in the estimated coefficient of variation of annual production cost cased by different random number seed can be observed. The estimated coefficients of variation of cost are summarized in Table 5-3. Also, the more detailed simulation results of each case are listed in Appendix B.

Table 5-3 Estimated Cost Variance Due To Load Variation

Case	Random Number Seed	Total Number of Simulations	Coefficient of Variation of Cost
1	1	28	2.39 %
2	2	30	2.43 %
3	3	30	2.46 %
4	6	30	2.45 %

The results of Table 5-3 show that the estimated coefficients of variation of cost based on different unit outage patterns are fairly close. The effects of different unit outage patterns on the estimation of the variance of cost caused by load variation uncertainty are not significant.

V.1.3 Cost Variance Due to Both Uncertain Factors

We assume that the uncertainties in load and unit outages are the only two factors that influence the cost variance in the production cost simulation. We also notice that the estimated expected costs of the previous two sub-sections are fairly close, \$252,239,271 vs. \$252,976,626. Suppose that the variances of cost induced by the two different factors are independent, the total variance of cost, which is caused by load variation and unit outages uncertainty, can be roughly approximated based on the estimated cost variances caused by each individual uncertain factor. Thus, the total variance of cost can be approximated to be the summation of the two estimated cost variances induced by each individual uncertain factor. That is $\$1,440,991^2 + \$4,967,378^2 = \$5,172,166^2$.

However, to estimate the total variance of cost with better precision, another set of simulations can be made by varying both the random number seed (unit outage pattern) and the sampled load curve in the same study. Again, the convergence criteria for the change of the variance of cost and the mean and standard deviation of annual energy are all set to 5%. The simulation results, which can be used to estimate the variance of cost induced by both uncertain factors at the same time, are summarized in Table 5-4. Once again, the estimated standard deviation of the annual energy of the first 21 simulations are not comparable with the estimated value using the historical load data. The estimated variance of cost becomes stable at run #31, and the simulated load variation is comparable with the system load variation as well. Hence, the mean and standard deviation of the cost are estimated to be \$253,320,131 and \$5,025,573 respectively. The estimated cost variance, \$5,025,573, of the simulation results is approximately equal to the approximated total variance of cost, \$5,172,166.

Table 5-4 Stochastic Results Due To Both Uncertain Factors

Load Curve ID	Annual Energy (MWH)	Cost ($)	Cost Variance Change	Coefficient of Variation of Cost
1	22705996	249847392	%	%
2	23813079	266051808	-	-
3	23101752	254785008	-90.3	3.23
4	22950312	252217152	-34.1	2.81
5	23000461	251674624	-22.9	2.54
6	23621917	260457104	-8.4	2.43
7	22997089	250521216	-6.6	2.36
8	22814032	248137520	2.3	2.40
9	23178872	255814000	-13.3	2.49
10	22926674	251039328	-8.3	2.16
11	23230723	255729184	-10.1	2.06
12	23173320	253697264	-9.9	1.97
13	22909122	249233648	-0.9	1.96
14	23888278	262954672	14.2	2.11
15	23101319	254512608	-7.7	2.03
16	22673799	245575684	10.5	2.15
17	22923532	249325088	-2.2	2.13
18	23297418	256277536	-4.8	2.08
19	23087990	251660928	-5.0	2.03
20	23056694	251300400	-4.4	1.99
21	24097126	248910592	-1.0	1.98
22	23742451	262770976	10.2	2.09
23	22901736	269097536	24.2	2.39
24	23348134	256395248	-4.1	2.35
25	22492190	244725984	6.2	2.42
26	23862874	265686976	8.7	2.53
27	23135635	256219296	-3.7	2.49
28	22506714	246093680	2.7	2.52
29	23524232	257944224	-2.5	2.49
30	23756729	259793920	-1.1	2.48
31	22952411	249120352	-0.9	2.47

To complete this sub-section, the estimated variances of annual cost using the stochastic approach are summarized in Table 5-5 for the purpose of comparison.

Table 5-5 Cost Variance and Coefficient of Variation of Stochastic Approach

	Variance	Coefficient of Variation
Due to Unit	$\$1,440,991^2$	0.57 (%)
Due to Load	$\$ 4,976,378^2$	2.39 (%)
Due to Both	$\$ 5,172,166^2$	2.47 (%)

V.2 Example of the Probabilistic Approach

Modeling the annual load variation in weekly detail requires a large number of chronological load profiles in order to simulate the annual load variation. Consequently, to estimate the variances of cost due to different uncertain factors requires long computational time. A simplified approach to deal with the annual load variation is to sample the energy and peak load in an annual horizon instead of a weekly horizon. Thus, by enlarging the time interval of the random variable, the stochastic process is simplified to a probabilistic model. The annual energy and peak load of each load scenario are sampled so that they can represent each load scenario with an associated probability.

The same generation system and historical load profiles are used in the probabilistic approach to illustrate the estimation of the variance of cost. Based on the adjusted annual load data of Table 3-1, the means and standard deviations of annual energy and peak load are $\hat{\mu}_E$=23,015,440 (MWH), $\hat{\sigma}_E$=439,754 (MWH), $\hat{\mu}_P$=4880 (MW), and $\hat{\sigma}_P$=185 (MW). Also, the correlation coefficient ($\hat{\rho}_{E,P}$) between the annual energy and peak load is estimated to be 0.9 for system A. According to the stratified sampling scheme, the annual energy and peak load of each scenario are calculated according to Equation (4-17) ~ (4-21) described in Chapter IV.2.1.

H_E = 23015440 + 1.35*439754 = 23,609,108 (MWH)

H_P = 4880 + 1.35*185 = 5,130 (MW)

M_E = 23,015,440 (MWH)

M_P = 4,880 (MW)

L_E = 23015440 - 1.35*439754 = 22,421,772 (MWH)

L_P = 4880 - 1.35*185 = 4,630 (MW)

Given the annual energy and peak load of each load scenario, a chronological load profile can be created using the load projecting algorithm described in Chapter III.3. These three hourly load curves are linearly transformed based on a typical load profile of the simulated system. In the following simulations, we also use the ordinary Monte Carlo sampling scheme in the production cost simulation to estimate the variances of cost. Fifteen annual production cost simulations are made, five for the high load, five for the mean load, and five for the low load, to evaluate the variance of cost. The resulting production costs are summarized in Table 5-6, and no unserved energy has occurred in these simulations.

Table 5-6 Annual Cost Summary of Probabilistic Approach

($)	High Load	Mean Load	Low Load
Run No. 1	263,899,664	254,288,096	244,491,504
Run No. 2	263,065,696	253,302,656	243,149,984
Run No. 3	263,806,208	254,035,280	244,225,888
Run No. 4	261,779,104	252,147,728	241,900,384
Run No. 5	260,838,048	251,144,192	241,284,256
Mean	262,677,744	252,983,590	243,010,403

The expected annual production cost is represented by the population mean which is the weighted average of the means of all scenarios. It is estimated to be $252,927,784 using Equation (4-24). In each load scenario, the variance of cost is induced by different unit outage patterns which are used to simulate the unit availability uncertainty. Given the probability of each load scenario based on stratified sampling, the variance of cost caused by unit outage uncertainty is calculated to be $\$1,341,385^2$ based on Equation (4-23). It is the weighted average of the variances of all scenarios. The results of Table 5-6 show that the variation of cost in each load scenario is very small when compared to the variance of the expected costs of the three scenarios. Based on Equation (4-25), the variance of cost induced by load variation alone can be estimated to be $\$6,219,735^2$ which is the variance of the expected costs of all scenarios. According to Equation (4-26), the variance of cost induced by both uncertain factors is also estimated to be $\$6,362,736^2$ which is the summation of variances induced by the two uncertain factors. These estimations of the variance of cost are summarized in Table 5-7, and the estimated coefficient of variation of cost are included as well.

Table 5-7 Cost Variance and Coefficient of Variation of Probabilistic Approach

	Variance of Cost	Coefficient of Variation of Cost
Due to Unit	$\$1,341,385^2$	0.53 (%)
Due to Load	$\$ 6,219,735^2$	2.46 (%)
Due to Both	$\$ 6,362,736^2$	2.52 (%)

In this probabilistic load model, we assumed that only five simulations for each scenario are required to estimate the variances of cost. This assumption is justified according to the following study using the probabilistic approach. In this study, the total number of simulations of case 1 is nine, and the number of simulations in each load scenario is increased to observe the stability of the estimated variances of cost. The results of this study are summarized in Table 5-8.

Table 5-8 Case Study Summary of Probabilistic Approach

Case	Total Number of Simulations	Cost Variation Due to Unit Outage	Cost Variation Due to Load Uncertainty
1	9	0.22%	2.45%
2	12	0.40%	2.45%
3	15	0.53%	2.46%
4	18	0.53%	2.46%
5	21	0.54%	2.46%

The results show that the coefficient of variation of cost due to load variation uncertainty is more stable than that due to unit outages, because the load scenarios are not random samples. Instead they are stratified points from a probability distribution. The slight variation in cost that occurs is due to the difference in the random sampling of unit outages that also slightly affects the cost variation due to the load uncertainty. The results imply that three runs are adequate to estimate the cost variance due to load variation uncertainty alone, but the estimated variance of cost due to unit outages will be more accurate when the number of simulations is more than four. Thus the number of simulations for the probabilistic approach is verified to be 5 in each scenario (total

15) in order to estimate the variance of cost caused by unit outage uncertainty and load variation at the same time.

V.3 Comparisons of the Two Approaches

The stochastic and probabilistic approaches are compared in this sub-section. Depending on the quality of the chronological production cost simulator that is used, both approaches may result in different estimates of coefficient of variation of cost and computational time. However, no matter which chronological production cost simulator is used, the results of the two proposed approaches should be consistent based on the same simulator. Included in the results will be the comparisons of the two approaches based on the quality of estimation, the efforts of load variation modeling, and computational time. To compare the estimated variance of cost, it is important to ensure that the annual load variation modeled by both approaches is comparable. The comparability of the simulated annual load variations of both approaches is addressed first.

V.3.1 Load Variation

In this dissertation, the annual load variation is measured by the variations in both annual energy and peak load. In the stochastic approach, the mean and standard deviation of annual energy and peak load are modeled by weekly samples, while the probabilistic approach uses the estimated mean and standard deviation of annual system load to represent the annual load variation. The annual load variation modeled by the two approaches is summarized in Table 5-9. In this table, the statistical data of annual load variation for the stochastic approach are estimated based on the 28 sampled load profiles. The results of Table 5-9 indicate that the annual load variation modeling of the two approaches are reasonable close.

Table 5-9 Annual Load Variation Summary

	Stochastic Load Variation	Probabilistic Load variation
Energy Mean	23,162,117	23,015,440
Energy STDEV	418,800	439,754
Peak Load Mean	5,092	4,880
Peak Load STDEV	265	185

V.3.2 Quality of Estimation

The primary goal of this research is to quantify the effects of both load uncertainty and unit availability uncertainty on annual production cost. The quality of the estimated results is the most important when compare the two proposed approaches. However, the philosophy of developing these two approaches is first to model the annual load variation according to the stochastic property of load variation over time in order to accurately estimate the variance of cost caused by load uncertainty. A simplified approach is then developed to reduce the efforts with an acceptable accuracy of estimation. Consequently, we already presume that the estimation of the stochastic approach should be more accurate than that of the probabilistic approach, because the stochastic behavior of the annual load variation is modeled. In addition, the stochastic approach models the annual load variation in a weekly horizon which is more detailed than an annual horizon in the probabilistic approach. Moreover, enough annual load profiles are sampled to ensure that annual load variation is well represented in the stochastic approach. On the other hand, only three load scenarios are used in the probabilistic approach to probabilistically describe the annual load variation according to the variation of system load.

The simulation results of the previous two examples are used to compare the quality of the estimated variances of cost. As the two approaches result in slightly different expected costs, the variation of annual production cost is compared in terms of the coefficient of variation of cost. The coefficients of variation of cost caused by load uncertainty, unit outage uncertainty, and both are estimated by the two approaches and are summarized in Table 5-10.

Table 5-10 Coefficient of Variation of Cost Summary of Two Approaches

Coefficient of Variation of Annual Cost	Stochastic Approach	Probabilistic Approach
Due to Unit Outage Uncertainty	0.57 %	0.53 %
Due to Load Uncertainty	2.39 %	2.46 %
Due to Both Uncertain Factors	2.45 %	2.52 %

The simulation results of Table 5-10 show that the estimated coefficients of variation of cost are fairly close. The difference in the estimated coefficients of variation of cost using different approach are not significant. Thus, we conclude the two approaches are comparable in the quality of estimation. Table 5-10 also reveals the important result that the cost variation due to load variation is significantly larger than that induced by the unit outages. This result is quite consistently estimated by both proposed approaches.

V.3.3 Effort of Load Variation Modeling

Since both approaches can result in a comparable estimation, it is important to know how much effort is required for each approach. In terms of the efforts of load variation modeling, the probabilistic annual load representation is much simpler than the weekly sampling scheme of the stochastic approach. To model the load variation stochastically, the weekly energies and peak loads are defined as the random variables which are described by their estimated distributions. Compared to the estimation of the distributions of annual energy and peak load in the probabilistic approach, it is more tedious to estimate all 52 pairs of distributions of random variables in the stochastic approach. In order to implement the conditional sampling, the estimations of the weekly correlation between energy and peak load and the correlation between weekly energies with one week lag are also required in the stochastic approach. Obviously, to model the annual load variation in a weekly horizon is more complicated than that in an annual horizon. Considering the efforts of modeling, the probabilistic approach is preferred. However, in order to use the probabilistic approach in estimating the variances of cots, it is necessary to verify the comparability of the estimated results first for different power system.

V.3.4 Computational Time

Certainly, the computational time of either approach will depend on what kind of computer and the production cost simulator that are used. However, it is fair to compare the computational time of two approaches based on the same production cost simulator and computer. Since the computational time required for each single annual simulation is fairly close, the total computational time of each approach is directly proportional to the total number of simulations. The simulation results obtained in this chapter show that the total number of simulations of the stochastic approach requires at least 28 simulations when the purpose is to estimate the cost variance caused by each individual uncertain factor or by both uncertain factors together. On the other hand, only 15 simulations are required to estimate the variances of cost in the probabilistic approach. Thus, the probabilistic approach can almost reduce the computational time required to estimate the variance of cost by one half.

For reference purposes, the total computational time required to estimate the variance of annual production cost using *Scheduler* for both approaches are provided. The computational time required for an annual simulation is about 7 minutes on a Pentium-90 computer with 24 MB RAM. Hence, it requires at least 3 hours (28 annual simulations) to estimate the variance of cost induced by load variation alone using a stochastic approach. It takes approximately one hour and 45 minutes (15 annual simulations) to estimate the variance of cost caused by unit outage uncertainty, load variation, and both factors using a probabilistic approach on the same computer.

CHAPTER VI

CONCLUSION

This dissertation reports on the investigation of the impact of the uncertainty in load and unit availability uncertainty on annual production cost using a chronological production cost simulation. The effects of these uncertainties on the annual production cost is measured in terms of the variance of cost. In addition to estimating the expected value of annual production cost, the variances of cost induced by load variation, unit outage uncertainty, and that induced by both factors are estimated. Thus, the estimated expected cost becomes much more meaningful when a measure of the variation of the annual production cost about the expected value is also included.

In this dissertation, both the stochastic and the probabilistic approaches of the annual load variation are proposed and demonstrated by applying each approach to estimate the variances of annual production cost of a medium size utility. In both cases, a chronological load projecting algorithm is used with different sampling schemes to generate hourly load profiles. The stochastic approach used a weekly conditional sampling scheme to model the load variation over time, while the probabilistic approach used a stratified sampling technique to describe the annual load variation probabilistically. The results herein have presented two approaches which simulate load variation over time in the format of a chronological load profile using either one of the two approaches.

This dissertation reports that the use of either one of these two approaches together with Monte Carlo simulations can be used to quantify the variances of cost due to unit outage uncertainty and load variation. In addition, the illustrative examples of both approaches reveal the import result that the estimated cost variance due to load uncertainty is more significant than that due to unit availability uncertainty, and therefore the annual load variation should be emphasized in estimating the variance of annual production cost.

In comparing the probabilistic approach to the stochastic approach, it is clear that the annual load variation modeling of the stochastic approach is more complicated, but more accurate than that of the probabilistic approach. The results of the stochastic approach are more accurate because the weekly load variation modeling does capture the stochastic behavior of the system load. However, the accuracy of the estimated variances of cost using the probabilistic approach is verified by the results of the stochastic approach. We recommend that the probabilistic approach should be compared with the stochastic approach for each system. Given comparable results, the simulations according to the probabilistic load variation modeling can be performed with significantly less time.

To estimate the cost variation due to unit availability, load variation and due to both factors, the probabilistic model can provide reasonable accuracy with less computational time. In addition, the probabilistic approach provides extra information, such as the expected cost of both the high and low load scenarios. Based on its performance, we conclude that the probabilistic approach is an efficient approach to approximate the variances of cost.

When estimating uncertainty in operational cost, uncertainty in load and uncertainty in generator availability are reflected through the production cost simulation in this dissertation. The uncertainty in unit outages is modeled by Monte Carlo sampling in a chronological production cost simulation, while the load variation can be modeled by either a stochastic approach or a probabilistic approach. Thus, a tool to reflect the variation in system load into the estimated variance of annual production cost will be useful and should be considered, especially for long-term studies. Consequently, the variance of annual production cost which is induced by the load uncertainty can be estimated. This is the major contribution of the research reported in this dissertation.

BIBLIOGRAPHY

[1] Alex D. Papalexopoulos, Ahangyou Hao and Tie-Mao Peng, "An Implementation of A Neural Network Based Load Forecasting Model for the EMS," *IEEE Trans. on Power Systems*, Vol. 9, No. 4, November 1994.

[2] Bih-Yuan Ku, Robert J. Thomas, Chiew-Yann Chiou, Chia-Jen Lin, "Power System Dynamic Load Modeling Using Artificial Neural Networks," *IEEE Trans. on Power Systems*, Vol. 9, No. 4, November 1994.

[3] A. Asar and J.R. McDonald, "A Specification of Neural Network Application in the Load Forecasting Problem," *IEEE Trans. on Control Systems Technology*, Vol. 2, No. 2, 1994.

[4] W. Christaanse, "Short-Term Load Forecasting Using General Exponential Smoothing," *IEEE Trans. on PAS*, Vol. PAS-90, No. 4, April 1971.

[5] M.T. Hagan, S.M. Behr, "The Time Series Approach to Short-Term Load Forecasting," *IEEE Trans. on PWRS*, Vol. PWRS-2, No. 3, August 1987.

[6] Andy Pole, Mike West and Jeff Harrison, *Applied Bayesian Forecasting and Time Series Analysis*, Chapman & Hall, New York, 1994.

[7] Alex D. Papalexopoulos, Timothy C. Hesterberg, "A Regression-Based Approach to Short-Term Load Forecasting," *IEEE Trans. on Power Systems*, Vol. 5, No. 4, November 1990.

[8] Takeshi Haida and Shoichi Muto, "Regression Based Peak Load Fortecasting Using A Transformation Technique," *IEEE Trans. on Power Systems*, Vol. 9, No. 4, November 1994.

[9] K. Liu and F.L. Lewis, "Some Issues about Fuzzy Logic Comtrol," *Proc. 23nd IEEE Conference on Decision and Control, 1993*

[10] K. Liu and I. H. Yu, "Adaptive Fuzzy Logic Approximation of Unknown Nonlinear System: State-Variable Feedback Tuning, "*Proc. 1994 American Control Conference.*

[11] H. Baleriaux, E. Jamoulle and Fr. Linard de Guertchin, "Simulation de L'exploitation d'unparc de machines thermiques de production d'electricite couples a des stations de pumpage," Revu E. (edition S.R.B.E.), Vol. 5, No. 7, 1967.

[12] Booth, "Power System Simulation Model based on Probability Analysis," *IEEE Trans. on PAS*, Vol. PAS-91, No. 1, 1972.

[13] A.D. Patton, C.Singh, M. Sahinoglu, "Operating Consideration in Generating Capacity Reliability Modeling -- An Analytical Approach," *IEEE Trans. on PAS*, May 1981.

[14] Rau, P. Toy and K.F. Schenk, "Expected Energy Production Costs by the method of Moments," *IEEE Trans. on PAS*, Vol. PAS-99, September/October 1980.

[15] J. P. Stremel, R.T. Jenkins, R.A. Rabb and W.D. Bayless, "Production Costing Using The Cumulant Method of Representing the Equivalent Load Curve," *IEEE Trans. on PAS*, Vol. PAS-99, September/October 1980.

[16] N. S. Rau and C. Necsulescu, "Probability Distributions of Incremental Cost of Production and Production Cost," *IEEE Trans. on Vol. PAS-104*, No. 12, December 1985.

[17] Mazumdar, C.K. Yin, "Variance of Power Generating System Production Costs," SP 88-68, Summer Power Meeting of the IEEE Power Engineering Society, Portland, July 1988.

[18] Narayan S. Rau, SM, and Youssef Hegazy, "The Variance of Production Cost", *IEEE Trans. on Power System*. Vol. PWRS-5, No. 2, May 1990.

[19] F.N. Lee, M. Lin, A.M. Breipohl, "Evaluation of the Variance of Production Cost Using a Stochastic Outage Capacity State Model," *IEEE Trans. on Power System*. Vol. 5, No. 4, November 1990.

[20] B. Manhire and R. T. Jenkins, "BenchMark: A Monye Carlo Hourly Chronological Simulation Model Which Includes Effects of Ramp-rates and Reservoir Constraints", Conference Procedings on Generation Planning: Modeling and Decisionmaking, August 10-12, 1982.

[21] D. T. Lee and L. M. Szeto, " A system of Energt Models for Operational Planning", Proceedings of the Canadian Electrical Association Spring Meeting, March, 1984.

[22] L. Wang, J. Endrenyi and L. D. Wilson, " The Application of Monte Carlo Simulation in Power System Reliability Evaluation", EPRI Workshop Proceedings: Power System Reliability-Research Needs and Priorities, EPRI WS-77-60, October 1978.

[23] L. Wang, "Approximate Confidence Bounds on Monte Carlo Simulation Results for Energy Production," *IEEE Trans. on Power System*, Vol. 4, No. 1, February 1989.

[24] A. Breipohl, F.N. Lee, J. Huang and Q. Feng, "Sample Size Reduction in Stochastic Production Simulation," *IEEE Trans. on Power System*, Vol. 5, No. 3, August 1990.

[25] A.M. Breipohl, F.N. Lee, Min-Ywan Lin, Hui-Dong Wang, and Jia-Yo Chiang, "An Adaptive Production Simulation Structure for Operational Planning with Annual Emission/Fuel Constraints," *IEEE Trans. on PWRS*, 1994.

[26] F.N. Lee, "Thermal Unit Commitment by Sequential Methods," IEEE Tutorial Text 90EH0328-5-PWRS, Application of Optimization Methods for Economy/Security Functions in Power System Operations.

[27] George J. Anders, "Probability Concepts In Electric Power Systems", John Wiley & Sons, New York, 1990.

[28] William G. Cochran, *Sampling Techniques*, Wiley, New York, 1977

[29] Stuart L. Meyer, *Data Analysis For Scientists and Engineers*, Wiley, New York, 1975.

[30] Chester H. McCall, *Sampling and Statistics Handbook for Research*, The Iowa State University Press, 1982.

[31] Hammersley, J.M., and Handscomb, D. C., *Monte Carlo Methods*, John Wiely & Suns, New York. 1964.

[32] Andrew P. Douglas, Arthur M. Breipohl, Fred N. Lee, and Rambabu Adapa, "Risk Due to Load Forecast Uncertainty in Short Term Power System Planning", *Submitted for publication in IEEE Trans. On PWRS.*

[33] A. M. Breipohl, F. N. Lee, D. Zhai, and R. Adapa, "A Gauss-Markov Load Model for Application in Risk Evaluation and Production Simulation", *IEEE Trans. on Power Systems*, Vol.7, NO. 4, Nov. 1992.

[34] Athanasios Papoulis, *Probability, Random Variables and Stochastic Process*, McGraw-Hill, , 1984.

APPENDIX A

Table A-1 Generation Unit Data

Unit #	Type	Pmax (MW)	Pmin (MW)	Min. Down (Hour)	Min. Up (Hour)
1	Gas	74	22	24	24
2	Gas	26	26	48	48
3	Gas	26	26	48	48
4	Gas	48	12	4	4
5	Gas	178	45	48	48
6	Gas	239	68	48	48
7	Gas	394	80	48	48
8	Gas	184	45	48	48
9	Coal	500	15	48	120
10	Coal	500	150	48	120
11	Coal	515	150	48	120
12	Gas	58	15	48	48
13	Gas	57	15	48	48
14	Gas	122	30	48	48
15	Gas	260	60	48	48
16	Gas	64	16	4	4
17	Gas	530	100	48	120
18	Gas	507	100	48	120
19	Gas	500	100	48	120
20	Gas	19	19	4	4
21	Coal	505	150	48	120
22	Coal	510	150	48	120
23	Gas	11	11	4	4
24	Gas	160	80	48	48
25	Gas	160	80	48	48
26	Gas	110	60	48	48

APPENDIX B

Table B-1 Costs Due To Load Variation Case 2
(Random Number Seed = 2)

Load Curve ID	Annual Energy (MWH)	Cost ($)	Cost Variance Change	Coefficient of Variation of Cost
1	22705996	247875896	%	%
2	23813079	266051808	-	-
3	23101752	255813248	-90.8	5.00
4	22950312	253574224	-46.8	3.64
5	23000461	253971456	-32.4	3.01
6	23621917	262237056	-0.4	2.60
7	22997089	253600768	-16.9	2.41
8	22814032	250678224	-5.9	2.35
9	23178872	255663952	-14.2	2.20
10	22926674	253471488	-11.3	2.08
11	23230723	257351408	-9.1	1.99
12	23173320	255693168	-9.9	1.90
13	22909122	250985360	-2.3	1.88
14	23888278	266878992	26.6	2.19
15	23101319	254499952	-7.3	2.11
16	22673799	247608160	6.9	2.20
17	22923532	251630384	-4.0	2.16
18	23297418	257006208	-5.4	2.10
19	23087990	254891616	-5.9	2.04
20	23056694	253259152	-4.9	1.99
21	24097126	270102528	23.6	2.27
22	23742451	263325456	17.6	2.19
23	22901736	251104016	2.9	2.22
24	23348134	259308736	0.8	2.23
25	22492190	246163472	13.0	2.39
26	23862874	264557824	4.7	2.45
27	23135635	256831360	-0.7	2.44
28	22506714	251819456	-1.5	2.42
29	23524232	258448848	0.7	2.43
30	23656729	260405328	1.0	2.43

Table B-2 Costs Due To Load Variation Case 3
(Random Number Seed = 3)

Load Curve ID	Annual Energy (MWH)	Cost ($)	Cost Variance Change	Coefficient of Variation of Cost
1	22705996	248853376	%	%
2	23813079	267246176	-	-
3	23101752	256961520	-91.9	3.65
4	22950312	254126304	-45.0	3.04
5	23000461	254412144	-31.9	2.66
6	23621917	263645792	1.7	2.67
7	22997089	254134304	-16.1	2.48
8	22814032	250979072	-4.2	2.44
9	23178872	256994336	-13.9	2.28
10	22926674	254097040	-11.0	2.17
11	23230723	258003712	-9.6	2.07
12	23173320	257334656	-9.5	1.98
13	22909122	252463360	-4.3	1.94
14	23888278	267394960	23.4	2.21
15	23101319	254795744	-6.8	2.14
16	22673799	248731296	5.9	2.21
17	22923532	252497568	-4.0	2.17
18	23297418	258625088	-4.8	2.11
19	23087990	255768816	-5.9	2.06
20	23056694	254023968	-4.7	2.01
21	24097126	270788832	22.5	2.27
22	23742451	264234128	15.6	2.18
23	22901736	252195600	2.8	2.21
24	23348134	260425568	0.7	2.22
25	22492190	247247760	12.4	2.37
26	23862874	265499328	4.0	2.42
27	23135635	257885552	-0.9	2.41
28	22506714	249929776	1.9	2.43
29	23524232	260797440	1.9	2.45
30	23656729	262024768	1.5	2.46

Table B-3 Costs Due To Load Variation Case 4
(Random Number Seed = 6)

Load Curve ID	Annual Energy (MWH)	Cost ($)	Cost Variance Change	Coefficient of Variation of Cost
1	22705996	249847392	%	%
2	23813079	267469872	-	-
3	23101752	254911360	-88.6	3.52
4	22950312	254657424	-45.0	2.93
5	23000461	254967152	-31.4	2.56
6	23621917	263872160	1.7	2.57
7	22997089	254860704	-16.5	2.39
8	22814032	252168464	-6.2	2.32
9	23178872	257228512	-14.1	2.17
10	22926674	254858208	-11.2	2.06
11	23230723	258796196	-9.0	1.97
12	23173320	257119824	-9.9	1.88
13	22909122	252569776	-2.7	1.86
14	23888278	268565856	27.9	2.18
15	23101319	255724912	-7.1	2.11
16	22673799	249352624	5.9	2.18
17	22923532	252895232	-3.7	2.14
18	23297418	258837784	-5.1	2.09
19	23087990	256933760	-5.9	2.03
20	23056694	253928688	-4.1	1.99
21	24097126	271197504	23.2	2.26
22	23742451	264581904	16.9	2.18
23	22901736	252576656	2.7	2.21
24	23348134	260732864	0.5	2.21
25	22492190	247626976	12.7	2.37
26	23862874	266075936	4.6	2.42
27	23135635	258312512	-0.8	2.41
28	22506714	250808352	2.2	2.44
29	23524232	260596288	1.2	2.45
30	23656729	261003904	0.3	2.45

APPENDEX C. 1

```
/* ========================================================================*/
/*      "stochld.h"    stochastic approach of load variation modeling        */
/*                     main program of weekly sampling program               */
/* ========================================================================*/
#include <stdio.h>
#include <string.h>
#include <math.h>
#include "loadlib.h"

extern  double      wke[52][10], wkp[52][10], grow[10];
extern  double      old[366][24], ld[366][24], fld[366][24];
extern  int         byear, fyear;
FILENAME            fname[10];

void main()
{
FILE          *dataf, *fb;
int           option, yearnum, i, j, k, l, lastweek, weeknum, ld[25], samplenum;
double        *p_point, *p_wk1, *p_wk2;
double        *p_ave, *p_cov, *p_cov1, *p_coep;
double        wgrowe[52], wgrowp[52], sgrowe[52], sgrowp[52];
double        oavee[52],  oavep[52],  savee[52],  savep[52];
double        ocove[52],  ocovp[52],  scove[52],  scovp[52];
double        ocov1e[52], ocov1p[52], tempe[52], tempp[52];
double        ocoe[52], scoe[52], ocop[52], scop[52];
double        coep[52], ocoepr[52], scoepr[52];
double        weeke[52], weekp[52], Nordat[52][100], N1, N2, N3;
double        engsum, peaksum, sumvar, aaa, bbb, gaaa, gbbb;
char          buf[10], NOR[50], LDB[50];

/*      load normal samples into array          */
dataf=fopen("C:\\CPROA\\STOCH\\NORMAL","r+b");
if (dataf != NULL){
        for (i=0; i<100; i++)
                {
                for (j=0; j<52; j++)
                        {
```

```
                fseek(dataf, (i*416+j*8), SEEK_SET);
                fread(&Nordat[j][i], 8, 1, dataf);
                }
            }
        }
fclose(dataf);

/*      reset typical load array to all 0           */
for (i=0; i<366; i++)
        {
        for(j=0; j<24; j++)
                ald[i][j] = 0.0;
        }

/*      convert historical load files to binary files           */
dataf=fopen("C:\\CPROA\\DATAF.TXT", "r+b");
fyear =0;
lastweek=8;
if (dataf != NULL){
        fscanf(dataf, "%3d", &option);
        fscanf(dataf, "%3d", &yearnum);

        /*      read load file name             */
        for (i=0; i<yearnum; i++)
                {
                fname[i].name[0]='\0';
                fseek(dataf, (i*11)+6, SEEK_SET);
                fread(&fname[i].name,10, 1, dataf);
                fname[i].name[9]='\0';
                if (option == 1) convert_binary(fname[i].name);

                /*      replace the text file name with a binary file name      */
                strncpy(buf, fname[i].name, 10);
                fname[i].name[0]='\0';
                strncpy(fname[i].name, buf, 6);
                fname[i].name[6]='\0';
                strcat(fname[i].name, "LDB");
                fname[i].name[9]='\0';
                /*      calculate weekly energy and peak                */
                week_eng_peak(fname[i].name, i);
                arrange_daytype(byear, fyear);
```

```
            }
        }
fclose(dataf);

/*      average the load curve added by week_eng_peak subroutine          */
for(j=0; j<366; j++)
        {
        for (i=0; i<24; i++)
                ald[j][i]=ald[j][i]/yearnum;
        }

/*      estimate the weekly energy and peak growth of the future year     */
/*      according to a linear trend                                       */
for(i=0;i <52; i++)
        {
        p_point = &wke[i][0];
        wgrowe[i] = trend(yearnum, p_point)/wke[i][4];
        p_point = &wkp[i][0];
        wgrowp[i] = trend(yearnum, p_point)/wkp[i][4];
        }

/*      calculate 7 lowest weekly energy to estimate annual growth        */
/*      and adjust loads to base year (year 5)                            */
low_set(yearnum);

/*      calculate the weekly mean and variance of the adjusted five years energy      */
p_wk1 =         &wke[0][0];
p_ave =         &oavee[0];
p_cov =         &ocove[0];
covar(5, 0, p_wk1, p_ave, p_cov);

p_wk1 =         &oavee[0];
p_ave =         &savee[0];
move_average(52, 5, p_wk1, p_ave);

for(i=0; i<52; i++)
        tempe[i] = sqrt(ocove[i]);

p_wk1 =         &tempe[0];
p_ave =         &scove[0];
move_average(52, 5, p_wk1, p_ave);
```

```
for(i=0; i<52; i++)
        scove[i] = scove[i]*scove[i];

/*      calculate the weekly mean and variance of the adjusted five years peak      */
p_wk1 =         &wkp[0][0];
p_ave  =        &oavep[0];
p_cov  =        &ocovp[0];
covar(5, 0, p_wk1, p_ave, p_cov);

p_wk1 =         &oavep[0];
p_ave  =        &savep[0];
move_average(52, 5, p_wk1, p_ave);

for(i=0; i<52; i++)
        tempp[i] = sqrt(ocovp[i]);

p_wk1 =         &tempp[0];
p_ave  =        &scovp[0];
move_average(52, 5, p_wk1, p_ave);
for(i=0; i<52; i++)
        scovp[i] = scovp[i]*scovp[i];

/*      calculate the correlation coefficient of weekly energy with one week lag      */
/*      and smooth the weekly coefficient                                             */
p_wk1 =         &wke[0][0];
p_ave  =        &oavee[0];
p_cov1=         &ocov1e[0];
covar(5, 1, p_wk1, p_ave, p_cov1);

for (i=0; i<51; i++)
        ocoe[i] = ocov1e[i] / sqrt(ocove[i] * ocove[i+1]);
p_wk1 =         &ocoe[0];
p_ave  =        &scoe[0];
move_average(51, 5, p_wk1, p_ave);

/*      calculate the correlation coefficient of weekly peak load         */
/*      with one week lag and smooth the weekly coefficient               */
p_wk1 =         &wkp[0][0];
p_ave  =        &oavep[0];
p_cov1=         &ocov1p[0];
covar(5, 1, p_wk1, p_ave, p_cov1);
```

```
for (i=0; i<51; i++)
        ocop[i] = ocov1p[i] / sqrt(ocovp[i] * ocovp[i+1]);
p_wk1 =       &ocop[0];
p_ave  =      &scop[0];
move_average(51, 5, p_wk1, p_ave);

/*      calculate the correlation coefficient between weekly energy and peak          */
p_wk1 =       &wke[0][0];
p_wk2 =       &wkp[0][0];
p_coep =      &coep[0];
covep(5, 0, p_wk1, p_wk2, p_coep);

for (i=0; i<52; i++)
        ocoepr[i] = coep[i] / sqrt(ocove[i] * ocovp[i]);
p_wk1 =       &ocoepr[0];
p_ave  =      &scoepr[0];
move_average(52, 5, p_wk1, p_ave);

/*      smooth the growths of  weekly energy and peak load          */
p_wk1 =       &wgrowe[0];
p_ave  =      &sgrowe[0];
move_average(52, 5, p_wk1, p_ave);

p_wk1 =       &wgrowp[0];
p_ave  =      &sgrowp[0];
move_average(52, 5, p_wk1, p_ave);

/*      adjust weekly growth based on the estimated annual growths of annual       */
/*      energy and peak load such that the annual energy and peak load are         */
/*      comparable with the results of annual growth; The annual energy and        */
/*      peak load growths from 1992 to 1993 is 1.0304 and 1.0228                   */
aaa = 0.0;
bbb = 0.0;
gaaa = 0.0;
gbbb = 0.0;
for (i=0; i<52; i++)
        {
        aaa = aaa + savee[i];
        if (savep[i] > bbb) bbb = savep[i];
        gaaa = gaaa + savee[i] * sgrowe[i];
```

```
        if (savep[i] * sgrowp[i] > gbbb) gbbb = savep[i] * sgrowp[i];
        }
sumvar = 0.0;
for(i=0; i<52; i++)
        {
        sgrowe[i] = sgrowe[i] * 1.0314 / (gaaa / aaa);
        sgrowp[i] = sgrowp[i] * 1.0228 / (gbbb / bbb);
        sumvar += ocove[i];
        }

/*      start sampling load scenarios           */
samplenum =  41;
for(l=0; l < samplenum; l++)
        {
        engsum = 0.0;
        peaksum = 0.0;
        weeknum = 7;
        sprintf(buf, "%1d",l);
        NOR[0]='\0';
        for(i=0; i<52; i++)
                {
              ' if (l != 0)
                        {
                        if (i == 51) weeknum = lastweek;
                        if (i==0)
                          {
                          N1 = Nordat[i][l-1];
                          weeke[i] = sgrowe[i] * (savee[i] + sqrt(scove[i]) * N1);
                          }
                        else
                          {
                          N2 = Nordat[i][l-1];
                          weeke[i] = sgrowe[i] * savee[i] + scoe[i-1] * sqrt(scove[i] /
                          scove[i-1]) * (weeke[i-1] - sgrowe[i-1] * savee[i-1]) +
                          sgrowe[i] * sqrt(scove[i] * (1-scoe[i-1] * scoe[i-1])) * N2 ;
                          }
                          N3 = Nordat[i][100-l];
                          if (scoepr[i] > 0.5)
                                weekp[i] = sgrowp[i] * savep[i] + scoepr[i] *
                                sqrt(scovp[i]/scove[i])*(weeke[i]- sgrowe[i]*savee[i]) +
```

```
                        sgrowp[i]* sqrt(scovp[i]* (1 - scoepr[i] * coepr[i])) *
                        N3;
                else
                        weekp[i] = sgrowp[i] * (savep[i] + sqrt(scovp[i]) * N3);
                }
        else
                {
                weeke[i] = 1.0 * sgrowe[i] * savee[i];
                weekp[i] = 1.0 * sgrowp[i] * savep[i];
                }
        }

/*      use the sampling weekly energy and peak to fit a hourly load curve      */
/*      based on the generated system typical load profile                      */
        weeknum = 7;
        for(i=0; i<52; i++)
                {
                if (i==51) weeknum = lastweek;
                wfittwo( i+1, weeke[i], weekp[i]);
                }
        lsmooth(200);
        weeknum = 7;
        LDB[0]='\0';
        strcpy(LDB, "C:\\CPROA\\STOCH\\LOAD");
        strcat(LDB, buf);
        strcat(LDB, ".LDB");
        fb=fopen(LDB, "w+b");
        ld[0]=1;
        ld[1]=1995;
        for(j=2; j<25; j++)
                ld[j]=0;
        fseek(fb, (long)(0), SEEK_SET);
        fwrite(&ld, sizeof(ld), 1, fb);
        for(i=0; i<52; i++)
                {
                if (i==51) weeknum = lastweek;
                for(k=0; k<weeknum; k++)
                        {
                        ld[0]=i*7+k+2;
                        for(j=1; j<25; j++)
                                {
```

```
ld[j] = (int)fld[i*7 + k][j];
if (ld[j] > peaksum) peaksum = ld[j];
engsum = engsum + ld[j];
}
fseek(fb, (long)(100 * (i*7+k+1)), SEEK_SET);
fwrite(&ld, sizeof(ld), 1, fb);
}
}
fclose(fb);
}
}
```

APPENDEX C. 2

```
/* ================================================================================*/
/*      "probald.h"    probabilistic approach of load variation modeling           */
/*                     main program ofannual sampling program                      */
/* ================================================================================*/
#include <stdio.h>
#include <string.h>
#include <math.h>
#include "loadlib.h"

extern double        wke[52][10], wkp[52][10], grow[10];
extern double        old[366][24], ld[366][24], fld[366][24];
extern int           byear, fyear;
FILENAME             fname[6];

void main()
{
FILE    *dataf, *fb, *ft;
char    buf[10];
int     option, yearnum, i, j, k, lastweek, ld[25];
double annualE[6], annualP[6], meanvarE[2], meanvarP[2], adde, addp;
double *p_eng, *p_peak, *p_meanvarE, *p_meanvarP;

/*      reset typical load array           */
for (i=0; i<366; i++)
        {
        for(j=0; j<24; j++)
                ald[i][j] = 0.0;
        }

/*      convert historical load files to binary files      */
dataf=fopen("C:\\CPROA\\DATAF.TXT", "r+b");
fyear = 0;
lastweek=8;
if (dataf != NULL){
        fscanf(dataf, "%3d", &option);
        fscanf(dataf, "%3d", &yearnum);
```

```
	/*	read load file name	*/
	for (i=0; i<yearnum; i++)
		{
		fname[i].name[0] = '\0';
		fseek(dataf, (i*11)+6, SEEK_SET);
		fread(&fname[i].name, 10, 1, dataf);
		fname[i].name[9] = '\0';
		fseek(ft, i*11, SEEK_SET);
		fwrite(fname[i].name, sizeof(fname[i].name), 1, ft);
		if (option == 1) convert_binary(fname[i].name);

		/*	replace the text file name with a binary file name	*/
		strncpy(buf, fname[i].name, 10);
		fname[i].name[0] = '\0';
		strncpy(fname[i].name, buf, 6);
		fname[i].name[6] = '\0';
		strcat(fname[i].name, "LDB");
		fname[i].name[9] = '\0';
		/*	calculate weekly energy and peak	*/
		week_eng_peak(fname[i].name, i);
		arrange_daytype(byear, fyear);
		}
	}
	fclose(dataf);

/*	average the load curve added by week_eng_peak subroutine*/
for(j=0; j<366; j++)
	{
	for (i=0; i<24; i++)
		ald[j][i] = ald[j][i] / yearnum;
	}

/*	calculate 7 lowest weekly energy to estimate annual growth	*/
/*	and adjust loads to base year (year 5)	*/
low_set(yearnum);

/*	calculate adjusted annual energy and peak and project to future year	*/
for (i=0; i<5; i++)
	{
	annualE[i] = 0.0;
	annualP[i] = 0.0;
```

```
        for(j=0; j< 52; j++)
                {
                annualE[i] += wke[j][i];
                if (wkp[j][i] > annualP[i]) annualP[i] = wkp[j][i];
                }
        /*      annual growth from 1992 to 1993 for energy and peak
                are 1.0304 and 1.0228                                   */
        annualE[i] = annualE[i] * 1.0304;
        annualP[i] = annualP[i] * 1.0228;
        }

/*      the alpha values of annual energy and peak result from stratified sampling  */
adde = 1.345;
addp = 1.345;

/*      calculate mean and variance of annual energy and peak       */
p_eng = &annualE[0];
p_meanvarE = &meanvarE[0];
mean(5, p_eng, p_meanvarE);

p_peak = &annualP[0];
p_meanvarP = &meanvarP[0];
mean(5, p_peak, p_meanvarP);

/*      create high load scenario profile file in Scheduler load format         */
yfittwo(meanvarE[0] + adde * sqrt(meanvarE[1]), meanvarP[0] + addp *
        sqrt(meanvarP[1]) );
fb = fopen("C:\\CPROA\\PROBA\\HILOAD.LDB", "w+b");
ld[0] = 1;
ld[1] = 1995;
for(j=2; j<25; j++)
        ld[j] = 0;
fseek(fb, (long)(0), SEEK_SET);
fwrite(&ld, sizeof(ld), 1, fb);
for(i=1; I <= 365; i++)
        {
        ld[0] = i+1;
        for(j=1; j< 25; j++)
                {
                if (fld[i-1][j-1] - (int)fld[i-1][j-1] > 0.5 )
                        ld[j] = (int)fld[i-1][j-1] + 1;
```

```
            else
                ld[j] = (int)fld[i-1][j-1];
            }
        fseek(fb, (long)(100 * i), SEEK_SET);
        fwrite(&ld, sizeof(ld), 1, fb);
        }
fclose(fb);

/*      create mean load scenario profile file in Scheduler load format         */
yfittwo(meanvarE[0], meanvarP[0]);
fb = fopen("C:\\CPROA\\PROBA\\MELOAD.LDB", "w+b");
ld[0] = 1;
ld[1] = 1995;
for(j=2 ; j< 25; j++)
        ld[j] = 0;
fseek(fb, (long)(0), SEEK_SET);
fwrite(&ld, sizeof(ld), 1, fb);
for(i=1; i<= 365; i++)
        {
        ld[0] = i+1;
        for(j = 1; j< 25; j++)
            {
            if (fld[i-1][j-1] - (int)fld[i-1][j-1] > 0.5 )
                ld[j] = (int)fld[i-1][j-1] + 1;
            else
                ld[j] = (int)fld[i-1][j-1];
            }
        fseek(fb, (long)(100 * i), SEEK_SET);
        fwrite(&ld, sizeof(ld), 1, fb);
        }
fclose(fb);

/*      create low load scenario profile file in Scheduler load format          */
yfittwo(meanvarE[0] - adde * sqrt(meanvarE[1]), meanvarP[0] - addp *
        sqrt(meanvarP[1]));
fb = fopen("C:\\CPROA\\PROBA\\LOLOAD.LDB", "w+b");
ld[0] = 1;
ld[1] = 1995;
for(j = 2; j<25; j++)
        ld[j] = 0;
fseek(fb, (long)(0), SEEK_SET);
```

```
fwrite(&ld, sizeof(ld), 1, fb);
for(i = 1; i<= 365; i++)
        {
        ld[0] = i+1;
        for(j=1; j<25; j++)
                {
                if (fld[i-1][j-1] - (int)fld[i-1][j-1] > 0.5 )
                        ld[j] = (int)fld[i-1][j-1]+1;
                else
                        ld[j] = (int)fld[i-1][j-1];
                }
        fseek(fb, (long)(100 * i), SEEK_SET);
        fwrite(&ld, sizeof(ld), 1, fb);
        }
fclose(fb);
}
```

APPENDEX C. 3

```
/* ==================================================================== */
/*     "global.h"     load uncertainty study global vaiable file         */
/* ==================================================================== */
double wke[52][10], wkp[52][10], grow[10];
double old[366][24], ald[366][24], fld[366][24];
int     byear, fyear;


/* ==================================================================== */
/*     "loadlib.h"     load uncertainty study subroutines & functions file   */
/* ==================================================================== */
#include <math.h>
#include <stdio.h>
#include <string.h>
#include <stdlib.h>
#include "global.h"

extern  double       wke[52][10], wkp[52][10], grow[10];
extern  double       old[366][24], ald[366][24], fld[366][24];
extern  int          fyear;

/*******************************************************/
/*        check leap year of a given year              */
/*******************************************************/
int leapyear(int year)
{
double          a, b, c;
int             leapyear;

a = fmod((double)year, 4.0);
b = fmod((double)year, 100.0);
c = fmod((double)year, 400.0);
if ((a == 0.0 && b != 0.0) || (c == 0.0))          /* leap year */
        leapyear = 1;
else
        leapyear = 0;
return (leapyear);
```

```
}

/ **************************************************************************/
/*      determine the total number of days in a given month of a year      */
/ **************************************************************************/
int monthday(int month, int year)
{
int     ndy;

if (month==1 || month==3 || month==5 || month==7 ||
                        month==8 || month==10 || month==12)
        ndy = 31;
else if (month==2 && leapyear(year) == 1)
        ndy = 29;
else if (month == 2)
        ndy = 28;
else
        ndy = 30;           /* all other months */
return (ndy);
}

/*****************************************************************/
/*      determine the day type of a given date                    */
/*****************************************************************/
int findday(int day, int month, int year)
{
int     dy_of_wk, ndy, mon, yr;
int     year1, year2, dlt;
double rem_day;

if (year < 1993)
        {
        year1 = year;
        year2 = 1992;
        }
else
        {
        year1 = 1993;
        year2 = year - 1;
        }
dy_of_wk = 6;                   /*      January 1, 1993 is Friday       */
```

```
ndy = 0;
for(yr=year1; yr<=year2; yr++)
        {
        if(leapyear(yr) == 1)           /*      a leap year             */
                ndy += 366;
        else
                ndy += 365;             /*      not a leap year         */
        }
rem_day = fmod((double)ndy, 7.0);

if (year < 1993)
        dy_of_wk = dy_of_wk - (int)rem_day;
else
        dy_of_wk = dy_of_wk + (int)rem_day;

if (dy_of_wk > 7)
        dy_of_wk = dy_of_wk - 7;
else if  (dy_of_wk < 1)
        dy_of_wk = dy_of_wk + 7;
ndy = 0;
for (mon=1; mon<month; mon++)
        {
        dlt = monthday(mon, year);
        ndy = ndy + dlt;
        }
ndy = ndy + (day - 1);

dy_of_wk = dy_of_wk + (int)fmod((double)ndy, 7.0);
if(dy_of_wk > 7)  dy_of_wk = dy_of_wk - 7;

return(dy_of_wk);
}

/***************************************************************/
/*          convert a text file into a scheduler binary format            */
/***************************************************************/
int convert_binary(char *filename)
{
FILE   *fp, *fout;
int     j, k, numday, dum, month, day, yearid, ldhr[25];
char    out_file[10];
```

```
/*      og&e load data file      */
out_file[0] = '\0';
strncpy(out_file, filename, 6);
strcat(out_file, "LDB");
out_file[9] = '\0';

/*      read year, check record number and write first year record           */
if ((fp = fopen(filename, "r+b")) == NULL) return (0);
if ((fout = fopen(out_file, "w+b")) == NULL) return (0);

fscanf(fp, "%2d", &month);
fscanf(fp, "%2d", &day);
fscanf(fp, "%2d", &yearid);
yearid = yearid + 1900;

/*      we only have the system data of 1995 and historical load profile only      */
/*      available from 1988 ~ 1992. we will create load file to be used in 1995    */
if (yearid == 1993) yearid=1995;
ldhr[0] = 1;
ldhr[1] = yearid;
for (j=2; j<25; j++)
        ldhr[j] = 0;
fseek(fout, 0, SEEK_SET);
fwrite(&ldhr, sizeof(ldhr), 1, fout);

if (leapyear(yearid) == 1)
        numday = 366;
else
        numday = 365;

/*      write numday records in (*.ldb) file           */
for (j=1; j<= numday; j++)
        {
        ldhr[0] = (int)j+1;
        for (k=1; k<=24; k++)
                fscanf(fp, "%5d", &ldhr[k]);
        if (j != numday)
                {
                for (k=1; k<=3; k++)
                        fscanf(fp, "%2d", &dum);
```

```
                }
        fseek(fout, (long)(sizeof(ldhr)*(j)), SEEK_SET);
        fwrite(&ldhr, sizeof(ldhr), 1, fout);
        }
fclose(fp);
fclose(fout);
return (1);
}

/**************************************************************/
/*      calculate weekly energy and peak load of a set of binary load files     */
/*      52 records:     each record includes 5 year weekly energy & peak        */
/**************************************************************/
int week_eng_peak(char *filename, int yearorder)
{
FILE    *fin;
int     lastweek, numday, i, j , k, ldhr[25];

/*      read year, check record number and write first year record   */
if ((fin = fopen(filename, "r+b")) == NULL) return (0);
fseek(fin, 0, SEEK_SET);
fread(&ldhr, sizeof(ldhr), 1, fin);

if (fyear == 0) fyear = ldhr[1]+5;
byear = ldhr[1];
if (leapyear(ldhr[1]) == 1)
        lastweek = 9;
else
        lastweek = 8;
numday=7;
for (i=0; i<52; i++)
        {
        if (i==51) numday = lastweek;
        wke[i][yearorder] = 0.0;
        wkp[i][yearorder] = 0.0;
        for(j=0; j<numday; j++)
                {
                fseek(fin, (i*7+j+1)*100, SEEK_SET);
                fread(&ldhr, sizeof(ldhr), 1, fin);
                for(k=1; k<=24; k++)
                        {
```

```
                    old[i*7+j][k-1] = (double)(ldhr[k]);
                    wke[i][yearorder] += (double)ldhr[k];
                    if ((double)ldhr[k] > wkp[i][yearorder])
                            wkp[i][yearorder] = (double)ldhr[k];
                    }
            }
    }
fclose(fin);
return (1);
}

/ ************************************************************ /
/*      fit the hourly loads of base year to day type of the future year      */
/*      old[366][24]:  the original hourly load form binary file              */
/*       ald[366][24]: the fitting hourly load for forecasting year           */
/*       byear        :  the base year id                                     */
/*       fyear        :  the forecasting year id                              */
/ ************************************************************ /
void arrange_daytype(int byear, int fyear)
{
int     bnumday, fnumday, i, j, a, b, dif;
double temp[366][24], fw, lw, l2w;

if (leapyear(byear) == 1;
        bnumday = 366;
else
        bnumday = 365;

if (leapyear(fyear) ==1)
        fnumday = 366;
else
        fnumday = 365;
fw = 0.0;
lw = 0.0;
l2w = 0.0;
for(i = 0; i<7; i++)
        {
        for(j=0; j<24; j++)
                {
                fw += old[i][j];
                lw += old[i+357][j];
```

```
            l2w += old[i+350][j];
            }
        }
if (findday(1,1, byear) == findday(1, 1, fyear))
        {
        for(i=0; i<fnumday; i++)
            {
            for(j=0; j<24; j++)
                ald[i][j] += old[i][j];
            }
        if (leapyear(fyear) == 1 && leapyear(byear) !=1)
            {
            for(j=0; j<24; j++)
                ald[365][j] += old[358][j] * lw/l2w;
            }
        }
else
        {
        a = findday(1, 1, byear);
        b = findday(1, 1, fyear);
        dif = b -a;
        if (dif < 0) dif += 7;

        /*      loads of the first day remain the same          */
        for(j=0; j<24; j++)
            temp[0][j] = old[0][j];

        /*      shift day 2 ~ 364-dif forward to the right weekday type      */
        for (i=1; i<bnumday-dif; i++)
            {
            for(j=0; j<24; j++)
                temp[i][j] = old[i+dif][j];
            }
        /*      average the connection point            */
        temp[1][0] =  temp[0][23]/2.0 + temp[1][1]/2.0;

        /*      copy day (364-dif+1 ~ numday) from day (2 ~ 2+dif)          */
        for (i=bnumday-dif; i<fnumday-1; i++)
            {
            for(j=0; j<24; j++)
                temp[i][j] = temp[i-7][j] * lw/l2w;
```

```
                }
        temp[bnumday-dif][0] = temp[bnumday-dif-1][23]/2.0 +
                                temp[bnumday-dif][1]/2.0;

        /*      last day remain the same (new year eve)         */
        for(j=0; j<24; j++)
                temp[fnumday-1][j] = old[bnumday-1][j];
        temp[fnumday-1][0] = temp[fnumday-2][23]/2.0 + temp[fnumday-1][1]/2.0;

        /*      add to average array ald[366][24]               */
        for (i=0; i<fnumday; i++)
                {
                for(j=0; j<24; j++)
                        ald[i][j] += temp[i][j];
                }
        }
}

/*****************************************************************/
/*      find the seven lowest weekly energies and adjust hourly loads       */
/*      based on load growth                                                */
/*      total[5]:       the total energy of the seven energy weeks          */
/*      adjust wk[52][12]  to the first year using total[0]/total[i]        */
/*****************************************************************/
void low_set(int yearnum)
{
double          low[10][10], total[10];
int             i, j, k, l, count;

for (l=0; l<yearnum; l++)
        {
        for (i=0; i<7; i++)
                low[i][l] = 9999999.9;
        for (i=0; i<52; i++)
                {
                count = 0;
                for(j=0; j<7; j++)
                        if (wke[i][l] < low[j][l]) count = count+1;
                if (count > 1)
                        {
                        for(k=6; k >= 8-count; k--)
```

```
                        low[k][l] = low[k-1][l];
                    }
                if (count > 0) low[7-count][l] = wke[i][l];
                }
        }
 for(l=yearnum-1; l>=0; l--)
        {
        total[l] = 0.0;
        for(i=0; i<7; i++)
                total[l] = total[l] + low[i][l];
        grow[l] = total[4]/total[l];
        }
for (i=0; i<52; i++)
        {
        for(j=0; j<yearnum-1; j++)
                {
                wke[i][j] = wke[i][j] * total[4]/total[j];
                wkp[i][j] = wkp[i][j] * total[4]/total[j];
                }
        }
}

/ ************************************************************/
/*      calculate the trend of a series of data                 */
/*      n:      number of the data series                       */
/*      *y:     pointer of the first data element address ·     */
/*      trend:  return the n+1 number according to y=a+bx      */
/ ************************************************************/
double trend(int n, double *y)
{
int     i;
double sumx, sumy, sumxy, sumxx, a, b;

sumx = 0.0;
sumy = 0.0;
sumxy = 0.0;
sumxx = 0.0;
a = 0.0;
b = 0.0;
for(i=0; i<n; i++)
        {
```

```
        sumx += (double)(i+1);
        sumy += *(y+i);
        sumxy += *(y+i) * (double)(i+1);
        sumxx += (double)((i+1)*(i+1));
        }
b = ((double)n*sumxy-sumx*sumy) / ((double)n*sumxx-sumx*sumx);
a = (sumy - b*sumx) / (double)n;
return ( a+ b*(double)(n + 1) );
}

/ *****************************************************************/
/*      calculate the mean and variance of a data set                         */
/*      order:          dimension of the data array                           */
/*      *vector:        pointer of the first array element address            */
/*      *result :       pointer of the result array element address           */
/ *****************************************************************/
void  mean(int order, double *vector, double *result)
{
int     k;

*result = 0.0;
*(result+1) = 0.0;
for(k=0; k<order; k++)
        *result += *(vector+k)/(double)(order) ;
for (k=0; k<order; k++)
        *(result+1) += (*(vector+k) - *result) * (*(vector + k) - *result) /
                (double)(order-1);
}

/ *****************************************************************/
/*      calculate the mean and covariance of energy and peak                  */
/*      lap:            lag number of the two dataset                         */
/*      *dataset:       pointer of the first dataset array element address    */
/*      *average:       pointer of the mean value array element address       */
/*      *result :       pointer of the covariance value array element address */
/ *****************************************************************/
void  covar(int order, int lap, double *dataset, double *average, double *result)
{
int     i, k, num;

num = 52 - lap;
```

```
for (i=0; i<52; i++)
        {
        *(average+i) = 0.0;
        *(result+i)= 0.0;
        for(k=0; k<order; k++)
                *(average+i) += *(dataset + i * 10 + k) / order ;
        }
for (i=0; i<num; i++)
        {
        for (k=0; k<order; k++)
            *(result+i) += ( (*(dataset+i*10+k) - *(average+i))  *
                  (*(dataset + (i+lap)*10 + k) - *(average+i+lap) ) )/(double)(order-1);
        *(result+i) = *(result+i) ;
        }
}

/ ******************************************************************/
/*      calculate covariance between energy and peak                */
/*      lap:            lag number of the two dataset               */
/*      *dataset:       pointer of the first dataset array element address  */
/*      *result :       pointer of the covariance value array element address  */
/ ******************************************************************/
void  covep(int order, int iap, double *dataset1, double *dataset2, double *result)
{
int             i, j, k, num;
double          aver[52][2];
double          *p_aver;

p_aver = &aver[0][0];
num = 52 - lap;
for (i=0; i<52; i++)
        {
        for (j=0; j<2; j++)
                {
                *(p_aver+i*2+j) = 0.0;
                *(result+i) = 0.0;
                }
        for(k=0; k<order; k++)
                {
                *(p_aver+i*2) += *(dataset1+i*10+k) / (double)(order) ;
                *(p_aver+i*2+1) += *(dataset2+i*10+k) / (double)(order) ;
```

```
                }
        }
 for (i=0; i<num; i++)
        {
        for (j=0; j<order; j++)
                *(result+i) += (*(dataset1+i*10+j) - *(p_aver+i*2)) *
                        (*(dataset2+(i+lap)*10+j) - *(p_aver+(i+lap)*2+1)) ;
        *(result+i) = *(result+i)/(double)(order-1) ;
        }
}

/ *****************************************************************/
/*      generate a normal distribution sample based on mean and variance      */
/*      mu:     mean                                                          */
/*      sigma:  standard deviation                                            */
/ *****************************************************************/
double Nor1()
{
int     i;
double sum = 0;

for(i=0; i<12; i++)
        sum += (double)rand()/(double)32767.0;
sum -= (double)6.0;
return sum;
}


double Nor2()
{
double sum = 0;

sum =  sqrt(-2.0* log((double)rand()/(double)32767.0)) *
        sin(2*3.141592654*(double)rand()/(double)32767.0);
return sum;
}

double Nor3()
{
double sum = 0;
```

```
sum =  sqrt(-2.0 * log((double)rand()/(double)32767.0)) *
        cos(2*3.141592654*(double)rand()/(double)32767.0);
return sum;
}

/*******************************************************************/
/*      weekly load projecting algorithm                           */
/*      ald[366][24]:  base load array                             */
/*      fld[366][24]:  fitted load array                           */
/*******************************************************************/
int wfittwo(int weekid, double Neng, double Npeak)
{
double Opeak, Oeng, sumxx, Ofirst,  a, b;
int     j, k, numday, lastweek, n;

Opeak = 0.0;
Oeng  = 0.0;
sumxx = 0.0;
lastweek = 8;
numday = 7;
if (weekid == 52) numday = lastweek;
n=24*numday;
for(j=0; j< numday; j++)
        {
        for(k=0; k<24; k++)
                {
                if (j==0 && k==0) Ofirst = ald[0][0];
                sumxx += ald[(weekid-1)*7+j][k] * ald[(weekid-1)*7+j][k];
                Oeng +=  ald[(weekid-1)*7+j][k];
                if ( ald[(weekid-1)*7+j][k] > Opeak ) Opeak = ald[(weekid-1)*7+j][k];
                }
        }
/*      two points linear filt    */
b = (double) ((Neng- n*Npeak)/(Oeng-n*Opeak));
a = Npeak - b*Opeak;
for(j=0; j<numday; j++)
        {
        for (k=0; k<24; k++)
                fld[(weekid-1)*7+j][k] = a+ b*ald[(weekid-1)*7+j][k];
        }
return (1);
```

```
}

/ ******************************************************************/
/*      annual load projecting algorithm                            */
/*      ald[366][24]:  base load array                              */
/*      fld[366][24]:  fitted load array                            */
/ ******************************************************************/
int yfittwo(double Neng, double Npeak)
{
double Opeak, Oeng,  a, b;
int     j, k, numday;

Opeak = 0.0;
Oeng  = 0.0;
if (leapyear(fyear) ==1)
        numday = 366;
else
        numday = 365;

for(j=0; j<numday; j++)
        {
        for(k=0; k<24; k++)
                {
                Oeng +=  ald[j][k];
                if ( ald[j][k] > Opeak ) Opeak = ald[j][k];
                }
        }
b = (double) ((Neng- numday*24*Npeak)/(Oeng-numday*24*Opeak));
a = Npeak - b*Opeak;
for(j=0; j<numday; j++)
        {
        for (k=0; k<24; k++)
                fld[j][k] = a + b*ald[j][k];
        }
return (1);
}

/ ******************************************************************/
/*      smooth the first three hourly loads of the projected weekly load curve     */
/*      diff:    threshold to be smooth                             */
/ ******************************************************************/
```

```
int lsmooth(int diff)
{
double mv[5];
int     i, j;

for(i=1; i<52; i++)
        {
        mv[0] = 0.0;
        mv[1] = 0.0;
        mv[2] = 0.0;
        mv[3] = 0.0;
        mv[4] = 0.0;
        if (fabs(fld[i*7][0] - fld[i*7-1][23]) > diff)
                {
                if ( fabs(fld[i*7][3]-fld[i*7-1][23])/4.0 > 300.0  && fabs(fld[i*7][5]-
                        fld[i*7-1][23])/6.0  < fabs(fld[i*7][3]-fld[i*7-1][23])/4.0 )
                        {
                        for(j=0;j<5; j++)
                                mv[j] = fld[i*7-1][23]+((fld[i*7][4]- fld[i*7-1][23])/6.0)
                                        * (float)(j+1);
                        for(j=0; j<5; j++)
                                fld[i*7][j] = mv[j];
                        }
                else
                        {
                        for(j=0;j<3; j++)
                                mv[j] = fld[i*7-1][23] +((fld[i*7][3]-fld[i*7-1][23])/4.0)
                                        * (float)(j+1);
                        for(j=0; j<3; j++)
                                fld[i*7][j] = mv[j];
                        }
                }
        }
return (1);
}

/ *******************************************************************/
/*      apply moving average to a series to smooth the variability           */
/*      size:           size of time series                                    */
/*      window:         moving average window size                             */
/*      odata:          original data                                          */
```

```
/*      ndata:          smoothed data                                   */
/***********************************************************************/
void  move_average(int size, int window, double *odata, double *ndata)
{
int     i;

for (i=0; i<size; i++)
        *(ndata + i) = 0.0;
if (window == 5)
        {
        *(ndata) += ( *(odata) * 0.4 + *(odata+1) * 0.2 + *(odata+2) * 0.1 )/0.7;
        *(ndata+1) += ( *(odata) * 0.2 + *(odata+1) * 0.4 + *(odata+2) * 0.2 +
                        *(odata+3) * 0.1 )/0.9;
        for (i=2; i<size-2; i++)
                *(ndata+i) += *(odata+i-2) * 0.1 + *(odata+i-1) * 0.2 + *(odata+i) *
                        0.4 + *(odata+i+1)* 0.2 + *(odata+i+2)* 0.1;
        *(ndata+size-2) += ( *(odata+size-4) * 0.1 + *(odata+size-3) * 0.2 +
                *(odata+size-2) * 0.4 + *(odata+size-1) * 0.2 )/0.9;
        *(ndata+size-1) += ( *(odata+size-3) * 0.1 + *(odata+size-2) * 0.2 +
                *(odata+size-1) * 0.4)/0.7;
        }
if (window == 7)
        {
        *(ndata) += ( *(odata) * 0.3 + *(odata+1) * 0.2 + *(odata+2) * 0.1 +
                *(odata+3) * 0.05 )/0.65;
        *(ndata+1) += ( *(odata) * 0.2 + *(odata+1) * 0.3 + *(odata+2) * 0.2 +
                *(odata+3) * 0.10 + *(odata+4) * 0.05)/0.85;
        *(ndata+2) += ( *(odata) * 0.1 + *(odata+1) * 0.2 + *(odata+2) * 0.3 +
                *(odata+3) * 0.20 + *(odata+4) * 0.10 + *(odata+5) * 0.05)/0.95;
        for (i=3; i<size-3; i++)
                *(ndata+i) += *(odata+i-3) * 0.05 + *(odata+i-2) * 0.1 + *(odata+i-1)
                        * 0.2 + *(odata+i) * 0.3 + *(odata+i+1)* 0.2 +
                        *(odata+i+2)* 0.1 + *(odata+i+3)* 0.05;
        *(ndata+size-3) += ( *(odata+size-6) * 0.05 + *(odata+size-5) * 0.1 +
                        *(odata+size-4) * 0.2 + *(odata+size-3) * 0.3 +
                        *(odata+size-2) * 0.2 + *(odata+size-1) * 0.1)/0.95;
        *(ndata+size-2) += ( *(odata+size-5) * 0.05 + *(odata+size-4) * 0.1 +
                        *(odata+size-3) * 0.2 + *(odata+size-2) * 0.3 +
                        *(odata+size-1) * 0.2)/0.85;
        *(ndata+size-1) += ( *(odata+size-4) * 0.05 + *(odata+size-3) * 0.1 +
                        *(odata+size-2) * 0.2 + *(odata+size-1) * 0.3)/0.65;
```

```
        }
if (window == 9)
        {
        * (ndata)+=  ( *(odata) * 0.250 + *(odata+1) * 0.15 + *(odata+2) * 0.10 +
                    *(odata+3) * 0.075 + *(odata+4) * 0.050)/0.625;
        * (ndata+1) += ( *(odata) * 0.150 + *(odata+1) * 0.25 +
                    *(odata+2) * 0.15 + *(odata+3) * 0.100 + *(odata+4) * 0.075+
                    *(data+5) * 0.050)/0.775;
        * (ndata+2) += ( *(odata) * 0.100 + *(odata+1) * 0.15 + *(odata+2) * 0.25 +
                    *(data+3) * 0.150 + *(odata+4) * 0.100 + *(odata+5) * 0.075+
                    *(data+6) * 0.050)/0.875;
        * (ndata+3) += ( *(odata) * 0.075 + *(odata+1) * 0.10 + *(odata+2) * 0.15 +
                    *(data+3) * 0.250 + *(odata+4) * 0.150 + *(odata+5) * 0.100+
                    *(data+6) * 0.075 + *(odata+7) * 0.05)/0.95;
        for (i=4; i<size-4; i++)
                *(ndata+i) += *(odata+i-4) * 0.05 + *(odata+i-3) * 0.075 +
                *(data+i-2) * 0.1 + *(odata+i-1) * 0.15 + *(odata+i) * 0.25 +
                *(odata+i+1) * 0.15 + *(odata+i+2)* 0.1 + *(odata+i+3)* 0.075 +
                *(odata+i+4)* 0.05;
        * (ndata+size-4) += ( *(odata+size-8) * 0.05 + *(odata+size-7) * 0.075 +
                *(odata+size-6) * 0.10 + *(odata+size-5) * 0.15 + *(odata+size-4) *
                0.25 + *(odata+size-3) * 0.150 + *(odata+size-2) * 0.10 +
                *(odata+size-1) * 0.075)/ 0.95;
        * (ndata+size-3) += ( *(odata+size-7) * 0.05 + *(odata+size-6) * 0.075 +
                *(odata+size-5) * 0.10 + *(odata+size-4) * 0.15 + *(odata+size-3) *
                0.25 + *(odata+size-2) * 0.150 + *(odata+size-1) * 0.10)/ 0.875;
                * (ndata+size-2) += ( *(odata+size-6) * 0.05 + *(odata+size-5) * 0.075
                + *(odata+size-4) * 0.10 + *(odata+size-3) * 0.15 + *(odata+size-2) *
                0.25 + *(odata+size-1) * 0.150 )/ 0.775;
        * (ndata+size-1) += ( *(odata+size-5) * 0.05 + *(odata+size-4) * 0.075 +
                *(odata+size-3) * 0.10 + *(odata+size-2) * 0.15 + *(odata+size-1) *
                0.25)/ 0.625;
        }
}
```